

Statistics for Biology and Health

George A. F. Seber
Matthew R. Schofield

Estimating Presence and Abundance of Closed Populations

 Springer

Statistics for Biology and Health

Series Editors

Mitchell H. Gail, Bethesda, MD, USA

Jonathan M. Samet, Department of Environmental & Occupational Health, University of Colorado Denver - Anschutz Medical Campus, Aurora, CO, USA

Statistics for Biology and Health (SBH) includes monographs and advanced textbooks on statistical topics relating to biostatistics, epidemiology, biology, and ecology.

George A. F. Seber • Matthew R. Schofield

Estimating Presence and Abundance of Closed Populations



Springer

George A. F. Seber
Department of Statistics
University of Auckland
Auckland, New Zealand

Matthew R. Schofield
Department of Mathematics and Statistics
University of Otago
Dunedin, New Zealand

ISSN 1431-8776
Statistics for Biology and Health
ISBN 978-3-031-39833-9
<https://doi.org/10.1007/978-3-031-39834-6>

ISSN 2197-5671 (electronic)
ISBN 978-3-031-39834-6 (eBook)

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Paper in this product is recyclable.

Preface

In recent times, particularly with climate change, deforestation, pollution, loss of species, pest control, and the control of some food resources such as in fisheries, there has been a greater interest in protecting our flora and fauna, and especially endangered species. Seber (1982) presented a variety of methods of estimating animal abundance and related parameters for both open and closed populations, which has been updated in two volumes, namely Seber and Schofield (2019) and this one. Here, a closed population is one which is constant in size during the period of investigation, except possibly having the removal of a known number of individuals. An open population is one where there are additional processes taking place such as birth, death, and migration in and out of the population area.

Since Seber (1982), there has been a profusion of research in population methods, so the time came to begin updates (e.g., Seber, 1986, 1992; Schwarz & Seber, 1999; Seber & Schwarz, 2002). A new topic appeared on the scene called adaptive sampling summarized by Seber (2013) that built on the foundational book by Thompson and Seber (1992), and was used mainly for closed populations. Since inference for proportions plays an important role in the subject, this led to the book by Seber (2013). A major part of Seber (1982) was devoted to the use of capture-recapture methods for both open and closed populations, with major developments taking place since then for open populations. This latter research is considered in Seber and Schofield (2019), and is still an expanding subject as models become more sophisticated and complex, requiring specialized computing facilities. It is hoped that that book and this one will provide a useful resource, which can be dipped into.

Seber (1982) also included other methods of population estimation, and this work is extended in this book where we look at methods specifically for closed populations. These can be applied not only to animal populations, but also to their signs, to plants and trees, and to human subpopulations of unknown size, such as the number of people with diabetes or the number of homeless using capture-recapture methods. Here, a wide variety of sampling methods can be brought into play using, for example, points, lines, strips, and plots of various shapes. So-called distance methods, a rapidly growing subject area, together with change in ratio methods and

catch effort methods, are considered along with the assumptions underlying these methods. Bayesian and spatial models are now becoming more popular.

It should be noted that this is not a book on sampling theory as there are several excellent books on the subject at an elementary or more advanced level. However, we do use sampling theory and often simply quote results from standard sampling theory. Our focus is on the type of population investigated and how to sample from it to obtain population estimates. Methods used are often dependent on the species being investigated, so a method which works well with a particular population might not do so well with a different population.

I (GAFS) have found some challenges in writing this book. To begin with, the populations are very diverse, as reflected in the diversity of journals referenced. Also, as with the whole subject of the dynamics of populations, the literature for this book has been almost overwhelming at times, and has been generally “unearthed” in a somewhat random fashion, with some articles occasionally inaccessible by our resources. We are sorry if we have missed any relevant papers. Some topics could fill a book on their own because of the large associated literature, sometimes running into over a thousand articles. We have therefore needed to be selective with an emphasis on research material that focused on new statistical methods and new types of application, but we still have 1400 references. After all, every article has something new to say!

Some papers are very long with detailed mathematics, which has meant some truncation in discussing them, but with a focus on trying to give a general idea of the method(s) used. Very occasionally a general article is not clear which method is being referred to as terms used like “removal” and “point-counts,” for example, have different meanings in different contexts. We generally give alternative names in brackets.

Some earlier models have been supplanted by later more general and more complex ones. However, we have generally tried to begin with simpler models before considering more complex ones to help the reader to ease into the subject, which can be formidable at times. We have also tried to make topics as self-contained as possible, but this is becoming more difficult as topics are overlapping more and more, and different types of data collection methods and models are being combined. Dividing the material into separate chapters has sometimes been difficult, with chapter sizes consequently being somewhat uneven. At times these features have led to occasional repetition to help the reader’s memory, especially with notation, but it has also created some tension in dealing with where to place a particular topic, especially with those comparing different methods.

Another challenge relates to choosing a notation, as notations in research papers vary considerably. We have tried to be consistent throughout, but there are times when it is more appropriate to try and follow the notation used by the research paper(s) being considered. This makes it easier for the reader to follow up on further details as many details needed to be left out of the book because of space restrictions. Computations have become more complicated and more reliant on appropriate computer software. Although there is not space to fill in some of the computational details, there is available an extensive array of many different computing programs

packages used by practitioners, which is complicated by updates. These tend to be just mentioned in passing.

It is hard to define detailed mathematical requirements for the reader. Acquaintance with a wide variety of statistical univariate and multivariate distributions is needed, as well as some basic matrix theory and some multivariable calculus. Capital letters for random variables have generally been avoided because of notational complexities, but occasionally they are appropriate. Knowledge of Bayesian methods is essential. Some additional general background information is given in the opening chapter and in the appendix. We now give a very sketchy description of what is in the chapters.

The book starts in Chap. 1 with some basic ideas involved with model development such as Bayes and frequentist methods, design- and model-based methods, some types of model (e.g., mixture models), model selection, and computational aspects. In Chap. 2, we look at plot sampling, including some design aspects such as number of plots, along with stratification, edge effect, spatial distributions, and adaptive sampling. As some individuals that are present may be missed, basic theory for incomplete detection, which permeates other chapters, is given in Chap. 3.

Estimating occupancy to determine the presence of one or more species is considered in the large Chap. 4 involving such topics as incomplete detection, optimal designs, and use of citizen science. Also mentioned is the use of animal signs, variable habitat, invasive species, disease modeling, the use of DNA, and many other topics. Chapter 5 is all about species, whether single or multiple, with related topics such as species' richness, species' interactions, and biodiversity. Chapter 6 is a short chapter looking at closest distance and nearest neighbor techniques, which are particularly useful for trees and stationary objects. Chapters 7 and 8 look at point and line transect distance methods, allowing for incomplete detection, where distances to observed objects from points or lines are recorded. Chapter 9 is a short chapter dealing with intercept methods, where the length of an intersect by a line with an object is measured to estimate coverage and object numbers, such as with plants.

Chapter 10, which has links with the next two chapters, considers estimating population size by successfully removing individuals either permanently or temporarily from a closed population, including change-in-ratio methods. Chapter 11 considers a similar approach, where the knowledge of effort is included in the modeling. Chapters 12 and 13 cover the subject of the capture-recapture method with its very large number of different models, and is artificially split into two chapters. Chapter 12 deals mainly with frequentist models including discrete and continuous models, and Chap. 13, by Matthew Schofield, deals with Bayesian models. Chapter 14 deals with the ever-growing subject of spatial models, which also occur elsewhere in the book. The book ends with an appendix that highlights some topics.

Auckland, New Zealand
Dunedin, New Zealand
June 2023

George A. F. Seber
Matthew R. Schofield

Contents

| | | |
|----------|-------------------------------------------|-----------|
| 1 | Model Building | 1 |
| 1.1 | Introduction | 1 |
| 1.2 | Model and Design-Based Methods | 2 |
| 1.3 | Classical Frequentist or Bayesian Methods | 3 |
| 1.4 | Hierarchical Modeling | 5 |
| 1.5 | Mixture Models | 5 |
| 1.6 | Regression Models | 6 |
| 1.6.1 | Covariates | 7 |
| 1.7 | Simulation | 7 |
| 1.8 | Model Fitting and Selection | 8 |
| 1.8.1 | Frequency Models | 8 |
| 1.8.2 | Bayesian Models | 9 |
| 1.8.3 | Model Averaging | 10 |
| 1.9 | Computing | 11 |
| 1.9.1 | Model Diagnostics | 13 |
| 1.10 | General Study Design | 13 |
| 1.11 | Where to Now? | 14 |
| 1.12 | Summary | 15 |
| 2 | Plot Sampling | 17 |
| 2.1 | Absolute Density | 17 |
| 2.2 | Design Considerations | 18 |
| 2.2.1 | Plot Shape and Area | 18 |
| 2.2.2 | Proportion of Area Sampled | 19 |
| 2.2.3 | Number of Plots | 20 |
| 2.2.4 | Irregular Areas | 20 |
| 2.2.5 | Trees | 21 |
| 2.2.6 | Primary and Secondary Units | 22 |
| 2.3 | Stratified Sampling | 23 |
| 2.3.1 | Estimation | 24 |
| 2.3.2 | Choosing Number of Plots | 24 |

| | | |
|----------|--------------------------------------------------------|-----------|
| 2.3.3 | Having a Random Distribution | 25 |
| 2.3.4 | Plot Sampling with Animals | 27 |
| 2.4 | Edge Effect | 27 |
| 2.5 | Home Range | 28 |
| 2.6 | Relative Density | 30 |
| 2.6.1 | Index Methods | 30 |
| 2.7 | Population Distribution | 32 |
| 2.7.1 | Types of Distribution | 32 |
| 2.7.2 | Negative Binomial Distribution | 32 |
| 2.7.3 | CMP Distribution | 34 |
| 2.7.4 | Taylor's Power law | 35 |
| 2.7.5 | Fairfield Smith's Variance Law | 35 |
| 2.8 | Widely Dispersed and Clustered Populations | 35 |
| 2.9 | Summary | 37 |
| 3 | Detectability | 39 |
| 3.1 | Detection Probability | 39 |
| 3.1.1 | Constant Known Detection Probability | 39 |
| 3.2 | Unknown Detection | 42 |
| 3.2.1 | Replicate Counts | 43 |
| 3.2.2 | Circular Plots | 45 |
| 3.2.3 | Double Sampling | 47 |
| 3.2.4 | Partial Repeat Sampling | 50 |
| 3.2.5 | Unequal Individual Detection Probabilities | 50 |
| 3.2.6 | Detection Model with Spatially Replicated Counts | 53 |
| 3.2.7 | Detection Methods | 55 |
| 3.3 | Adaptive Sampling with Incomplete Detectability | 56 |
| 3.4 | Conclusion | 56 |
| 4 | Occupancy, Abundance, and Related Topics | 57 |
| 4.1 | An Overview | 57 |
| 4.1.1 | Basic Model | 59 |
| 4.1.2 | Animal Signs | 61 |
| 4.1.3 | Some Reviews | 62 |
| 4.1.4 | Rare Species | 63 |
| 4.1.5 | Repeated Counts | 63 |
| 4.1.6 | Small Scale and Abundance | 66 |
| 4.1.7 | Invasive Species | 67 |
| 4.2 | Detection Probabilities | 68 |
| 4.2.1 | Homogeneous Detection Probabilities | 69 |
| 4.2.2 | Variable Parameters | 69 |
| 4.2.3 | Heterogeneous Detection Probabilities | 70 |
| 4.2.4 | Model Checking | 71 |
| 4.2.5 | Detectability, Covariates, and Detection-Only Data ... | 76 |
| 4.2.6 | Plants and Variable Plot Sizes | 77 |

| | | |
|--------|----------------------------------------------------|-----|
| 4.3 | Some General Models | 77 |
| 4.3.1 | Using the Robust Model | 82 |
| 4.3.2 | Spatial Models | 84 |
| 4.3.3 | Penalized Likelihood | 88 |
| 4.3.4 | Testing for Occupancy Differences | 89 |
| 4.3.5 | Surveys After the First Detection? | 90 |
| 4.4 | Key Assumptions | 92 |
| 4.4.1 | Model Adequacy | 93 |
| 4.4.2 | Closure Assumption | 94 |
| 4.5 | Some Alternative Methods | 98 |
| 4.6 | Choosing Design Parameters | 99 |
| 4.6.1 | Three Designs | 99 |
| 4.6.2 | Bayesian Optimal Design | 101 |
| 4.7 | Adaptive Methods for Rare Species | 103 |
| 4.7.1 | Two-Phase Design | 103 |
| 4.7.2 | Conditional Design | 105 |
| 4.8 | Occupancy and Abundance | 107 |
| 4.8.1 | Population Trend | 109 |
| 4.8.2 | Random or Elusive Aggregated Populations | 111 |
| 4.9 | Observation Errors | 112 |
| 4.9.1 | False Positives | 114 |
| 4.10 | Further Multiple Detection Models | 116 |
| 4.10.1 | Combining Detection Types | 118 |
| 4.10.2 | Multiple Detection Devices | 122 |
| 4.11 | DNA Methods | 123 |
| 4.11.1 | eDNA | 124 |
| 4.12 | Presence-Only Data | 127 |
| 4.12.1 | Pseudo-Absence Problem | 127 |
| 4.12.2 | Poisson Point Process Models | 128 |
| 4.12.3 | Geographic Distribution | 129 |
| 4.12.4 | Using Environmental Conditions | 130 |
| 4.12.5 | Methods of Phillips & Elith, 2013 | 132 |
| 4.12.6 | Combining Presence and Presence/Absence Data | 135 |
| 4.13 | Multi-Scale Models | 137 |
| 4.14 | Multispecies Models | 141 |
| 4.15 | Zero-Inflated Models | 141 |
| 4.16 | Opportunistic and Citizen Surveys | 144 |
| 4.16.1 | Two-Stage Model | 144 |
| 4.16.2 | Pro and Cons | 145 |
| 4.16.3 | Accounting for False Positives | 146 |
| 4.16.4 | Disasters | 147 |
| 4.16.5 | Bird Applications | 147 |
| 4.16.6 | General Comments | 148 |
| 4.16.7 | Citizen Presence-Only Data | 149 |

| | | |
|----------|-----------------------------------------------|------------|
| 4.17 | Multistate Methods | 152 |
| 4.17.1 | Undetected State | 154 |
| 4.17.2 | Dynamic State Models..... | 156 |
| 4.18 | Disease Modeling | 161 |
| 4.18.1 | Underlying Problems | 161 |
| 4.18.2 | Multiple Data Sources | 163 |
| 4.18.3 | Dynamic Model | 164 |
| 4.19 | Multiple Data Methods | 165 |
| 4.20 | Fisheries and Marine Environments | 167 |
| 4.21 | Conclusion | 168 |
| 5 | Species Methods | 171 |
| 5.1 | Introduction | 171 |
| 5.2 | Species Distribution Models (SDMs) | 172 |
| 5.2.1 | Background and Reviews..... | 173 |
| 5.2.2 | Species and the Environment..... | 174 |
| 5.2.3 | Model Building | 174 |
| 5.2.4 | Five Challenges | 175 |
| 5.2.5 | Bayesian Models | 175 |
| 5.2.6 | Community Species | 177 |
| 5.3 | Multivariate Species Models | 179 |
| 5.3.1 | Introduction | 179 |
| 5.3.2 | Various Multivariate Models | 180 |
| 5.3.3 | Occupancy Models | 184 |
| 5.3.4 | Comparing Models | 189 |
| 5.3.5 | Spatial Models | 190 |
| 5.3.6 | Computing | 190 |
| 5.3.7 | Species Correlations | 191 |
| 5.4 | Species Interactions | 192 |
| 5.4.1 | Two Interactive Species | 192 |
| 5.4.2 | Markov Networks | 194 |
| 5.4.3 | Environmental Effects | 195 |
| 5.4.4 | Species' Counts | 197 |
| 5.4.5 | Consumer Resource Dynamical Models..... | 199 |
| 5.4.6 | Using Absences | 200 |
| 5.4.7 | Time-to-Detection Models | 201 |
| 5.5 | Species Richness | 203 |
| 5.5.1 | Replicate Visits | 204 |
| 5.5.2 | Data Augmentation | 207 |
| 5.5.3 | Parametric Methods | 208 |
| 5.5.4 | Nonparametric Estimates: Abundance Data | 212 |
| 5.5.5 | Nonparametric Estimates: Incidence Data | 214 |
| 5.5.6 | Rarefaction and Extrapolation Methods | 216 |
| 5.5.7 | Species Ranking | 220 |
| | Second Sample Adding New Species | 221 |

| | | |
|----------|------------------------------------------------------------|------------|
| 5.5.9 | Choosing the Sample Size | 223 |
| 5.5.10 | Using eDNA | 226 |
| 5.5.11 | Biodiversity | 227 |
| 5.5.12 | Biodiversity Measures | 228 |
| 5.5.13 | Environmental Links | 231 |
| 5.5.14 | Species-Area Relationships (SARS) | 232 |
| 5.6 | Species Misidentification | 234 |
| 5.7 | Opportunistic Surveys | 235 |
| 5.8 | Summary | 236 |
| 6 | Closest Distance and Nearest Neighbor Methods | 237 |
| 6.1 | Description | 237 |
| 6.2 | Closest-Individual Distance | 238 |
| 6.2.1 | Poisson Model | 239 |
| 6.3 | The r th Closest Distance | 240 |
| 6.3.1 | The r th Nearest Neighbor | 242 |
| 6.3.2 | Nonrandom Distributions | 242 |
| 6.4 | Methods for Trees | 244 |
| 6.4.1 | Plot Methods | 244 |
| 6.4.2 | Distance Sampling for Trees | 246 |
| 6.4.3 | Three Problems with Closest Distance Methods | 252 |
| 6.5 | Summary | 252 |
| 7 | Point Counts and Point-Distance Methods | 255 |
| 7.1 | Introduction | 255 |
| 7.2 | Point-Distance Sampling | 256 |
| 7.2.1 | Theory for Point Distances | 257 |
| 7.2.2 | Experimental Design | 260 |
| 7.2.3 | Measurement Errors | 262 |
| 7.2.4 | Detection Functions | 264 |
| 7.2.5 | Point Sampling Along Lines | 265 |
| 7.2.6 | Cluster Sampling | 268 |
| 7.3 | Combined with Removal Method | 268 |
| 7.4 | Trapping Web | 271 |
| 7.5 | Single Trap or Lure Method | 272 |
| 7.6 | Some Model Variations | 274 |
| 7.6.1 | Variable Circular Plot | 274 |
| 7.6.2 | Unconditional Likelihood | 275 |
| 7.6.3 | Repeated Counts Model | 277 |
| 7.6.4 | Spatially Replicated Counts Model | 277 |
| 7.7 | Double Observers | 278 |
| 7.7.1 | Theory | 279 |
| 7.7.2 | Unreconciled Double-Observer Method | 281 |
| 7.7.3 | Aerial Counts | 282 |
| 7.8 | Components of Detectability | 283 |
| 7.8.1 | Validation and Comparison of Point Methods | 288 |

| | | |
|----------|--------------------------------------------------|------------|
| 7.9 | Combined Methods | 289 |
| 7.10 | Summary | 294 |
| 8 | Line Transect and Distance Methods | 295 |
| 8.1 | Line Distances | 295 |
| 8.1.1 | Basic Assumptions | 296 |
| 8.2 | General Distance Theory | 298 |
| 8.2.1 | Variance Estimation | 299 |
| 8.2.2 | Encounter Rate | 301 |
| 8.2.3 | Systematic Sampling | 303 |
| 8.2.4 | Pooling Robustness | 305 |
| 8.2.5 | Detection Function | 306 |
| 8.2.6 | Group (Cluster) Sampling | 311 |
| 8.2.7 | Transects in Plots | 312 |
| 8.3 | Design Considerations | 313 |
| 8.3.1 | Some Designs | 314 |
| 8.3.2 | Choice of Transects | 316 |
| 8.4 | Transects of Random Length | 316 |
| 8.5 | Animal Signs | 318 |
| 8.6 | Line Transect Radial Distances | 319 |
| 8.6.1 | Circular Flushing Region | 320 |
| 8.7 | Plant Sampling | 321 |
| 8.7.1 | Crossed Design | 321 |
| 8.7.2 | Time Spent | 324 |
| 8.8 | Acoustic Methods | 325 |
| 8.9 | Using Presence/Absence Data | 327 |
| 8.9.1 | Log-Linear Model | 330 |
| 8.10 | Point or Line Transects? | 332 |
| 8.11 | Method of Double Observers Using Distances | 334 |
| 8.11.1 | Using Mark-Recapture and Distance | 335 |
| 8.11.2 | Allowing for Heterogeneity | 338 |
| 8.11.3 | Levels of Independence | 340 |
| 8.11.4 | Point Independence and Capture-Recapture | 344 |
| 8.12 | Aerial Censusing | 346 |
| 8.12.1 | Population Size | 348 |
| 8.12.2 | Some Examples | 349 |
| 8.12.3 | Some Model Variations | 350 |
| 8.12.4 | Supplementary Ground Counts | 354 |
| 8.12.5 | Digital Observations Using Cameras | 355 |
| 8.13 | Extension of Model-Based Methods | 356 |
| 8.13.1 | Shipboard Acoustic Surveys | 356 |
| 8.13.2 | Other Marine Surveys | 359 |
| 8.13.3 | Allowing for Movement | 360 |
| 8.13.4 | Full Model-Based Method | 361 |
| 8.13.5 | Spatial Models | 362 |

| | | |
|-----------|--------------------------------------------------|------------|
| 8.14 | Bayesian Methods | 366 |
| 8.15 | Three-Dimensional Line Transects | 368 |
| 8.16 | Summary | 372 |
| 9 | Line Intercept Methods | 373 |
| 9.1 | Line Intercept | 373 |
| 9.1.1 | Comparison of Methods | 374 |
| 9.1.2 | General Theory | 374 |
| 9.1.3 | Estimation of Coverage | 375 |
| 9.1.4 | Estimation of Particle Density | 378 |
| 9.1.5 | Kaiser's Generalization | 379 |
| 9.2 | Transects of Random Length | 382 |
| 9.2.1 | Coverage | 382 |
| 9.2.2 | Particle Density | 382 |
| 9.2.3 | Method of Kaiser | 385 |
| 9.2.4 | Line Point Intercept/Transect | 387 |
| 9.3 | Summary | 387 |
| 10 | Removal and Change-in-Ratio Methods | 389 |
| 10.1 | Basic Models | 389 |
| 10.1.1 | Constant Probability of Removal | 390 |
| 10.1.2 | Trap Reduction | 392 |
| 10.1.3 | Combined with a Mark Release | 393 |
| 10.1.4 | Regression Methods | 395 |
| 10.1.5 | Variable Catchability | 396 |
| 10.2 | Bayesian Methods | 397 |
| 10.2.1 | Random Catchability | 397 |
| 10.2.2 | Use of Effort Information | 399 |
| 10.2.3 | Aerial Sightings | 401 |
| 10.2.4 | Some Electrofishing Applications | 402 |
| 10.3 | Two and Three Removals | 403 |
| 10.3.1 | Three Removals | 404 |
| 10.3.2 | Two Removals | 405 |
| 10.4 | Electrofishing | 406 |
| 10.5 | Design of a Removal Experiment | 409 |
| 10.6 | Removal Methods for Subpopulations | 410 |
| 10.6.1 | Multiple Sites | 414 |
| 10.7 | Removal Methods for Point-Count Surveys | 414 |
| 10.8 | Time-to-Detection Model | 416 |
| 10.9 | Change in Ratio: Two Classes | 418 |
| 10.9.1 | Theory | 419 |
| 10.9.2 | Binomial Model | 420 |
| 10.9.3 | Survival Rates | 422 |
| 10.9.4 | Removals Estimated | 423 |
| 10.9.5 | Optimum Allocation | 424 |

| | | |
|-----------|-----------------------------------------------------|------------|
| 10.9.6 | Exploitation Rate | 425 |
| 10.9.7 | Subsampling | 425 |
| 10.9.8 | Continuous Case | 426 |
| 10.10 | Multiple Removals and Samples | 427 |
| 10.11 | Index-Removal Method | 429 |
| 10.12 | Known Sampling Effort | 429 |
| 10.12.1 | Several Subclasses, Removals, and Samples | 430 |
| 10.13 | Omnibus Removal Methods | 434 |
| 10.13.1 | Combining CIR and IR | 434 |
| 10.13.2 | Combining Three Methods | 436 |
| 10.14 | Summary | 438 |
| 11 | Catch-Effort Models | 439 |
| 11.1 | Types of Model | 439 |
| 11.2 | Homogeneous Models | 440 |
| 11.2.1 | Method of Estimating Functions | 441 |
| 11.3 | Heterogeneous Models | 443 |
| 11.3.1 | Using Coverage and Estimating Functions | 444 |
| 11.4 | Regression Models | 445 |
| 11.5 | Underlying Assumptions | 448 |
| 11.5.1 | Some Variable Catchability Models | 449 |
| 11.5.2 | When Is a Species Extinct? | 450 |
| 11.5.3 | Fisheries | 452 |
| 11.6 | Summary | 453 |
| 12 | Capture-Recapture: Frequentist Methods | 455 |
| 12.1 | Introduction | 455 |
| 12.2 | Two-Sample Capture-Recapture | 456 |
| 12.2.1 | Underlying Assumptions | 457 |
| 12.2.2 | Variable Catchability | 458 |
| 12.2.3 | Two Observers with Time to Detection | 462 |
| 12.2.4 | Epidemiological Population: Two Lists | 463 |
| 12.2.5 | Epidemiological Population: Several Lists | 468 |
| 12.2.6 | Dual Record System | 468 |
| 12.2.7 | Some Individuals Stratified | 470 |
| 12.2.8 | Detection Model | 471 |
| 12.3 | Triple-Catch Method | 472 |
| 12.4 | Several Samples | 475 |
| 12.4.1 | Multinomial Model | 476 |
| 12.4.2 | Conditional Method | 477 |
| 12.5 | Pollock's Model Categories | 478 |
| 12.5.1 | Model M_t | 479 |
| 12.5.2 | Model M_0 | 483 |
| 12.5.3 | Model M_b | 484 |
| 12.5.4 | Model M_{tb} | 485 |
| 12.5.5 | Model M_h | 490 |

| | | |
|---------|--------------------------------------------------|-----|
| 12.5.6 | Model M_{th} | 492 |
| 12.5.7 | Model M_{bh} | 493 |
| 12.5.8 | Model M_{tbh} | 495 |
| 12.5.9 | Choosing the Number of Samples | 495 |
| 12.5.10 | Model Selection | 495 |
| 12.5.11 | Huggins' Models | 496 |
| 12.5.12 | Test of Model Assumptions | 497 |
| 12.6 | Mixture Models | 497 |
| 12.6.1 | Models of Norris and Pollock | 501 |
| 12.6.2 | Models of Pollock and Otto | 501 |
| 12.6.3 | Identifiability | 504 |
| 12.6.4 | Model of Dorazio and Royle | 504 |
| 12.6.5 | Model of Tounkar and Rivest | 505 |
| 12.7 | Log-Linear Poisson Models | 506 |
| 12.7.1 | Using the Conditional Model | 507 |
| 12.7.2 | Pollock's Models in Log-Linear Form | 507 |
| 12.8 | Coverage Models | 509 |
| 12.8.1 | Heterogeneity | 509 |
| 12.8.2 | Coverage and Pollock's Models | 511 |
| 12.8.3 | Coverage and Dual Record System | 513 |
| 12.9 | Martingales and Estimating Functions | 515 |
| 12.9.1 | Combined with Coverage | 516 |
| 12.10 | Time-to-Detection Method | 518 |
| 12.11 | Prior Detection Effects | 519 |
| 12.12 | Using DNA Tags | 520 |
| 12.13 | Misidentification | 521 |
| 12.14 | Sampling One at a Time | 523 |
| 12.15 | Continuous-Time Models | 525 |
| 12.15.1 | Markov Model | 526 |
| 12.15.2 | General Frailty Models | 526 |
| 12.15.3 | Some Model Variations | 528 |
| 12.15.4 | Models of Schofield et al. | 529 |
| 12.15.5 | Random Removals | 534 |
| 12.15.6 | Self-Exciting (Self-Correcting) Processes | 535 |
| 12.15.7 | Single-Catch Spatial Model | 536 |
| 12.16 | Conclusion | 540 |
| 13 | Capture-Recapture: Bayesian Methods | 543 |
| 13.1 | Overview | 543 |
| 13.2 | Bayesian Statistics | 543 |
| 13.3 | Two-Sample Capture-Recapture | 545 |
| 13.3.1 | Roberts (1967) | 545 |
| 13.3.2 | Other Approaches | 547 |
| 13.4 | Model M_t | 548 |
| 13.4.1 | Poisson Model | 549 |

| | | |
|---------|---------------------------------------------------------------|-----|
| 13.4.2 | Sequential Updating of Posterior for N | 551 |
| 13.5 | Bayesian Development of Species Distribution Models | 551 |
| 13.6 | Model Fitting with Markov Chain Monte Carlo..... | 552 |
| 13.7 | List Dependence | 553 |
| 13.8 | Behavior | 555 |
| 13.9 | Heterogeneity | 556 |
| 13.9.1 | Rasch Models | 556 |
| 13.9.2 | Mixture Model | 558 |
| 13.9.3 | Latent Class Model | 561 |
| 13.9.4 | Non-parametric Approaches | 562 |
| 13.10 | Estimating Abundance | 563 |
| 13.10.1 | Marginalize over p | 565 |
| 13.10.2 | Joint Update | 566 |
| 13.10.3 | Numerical Integration | 567 |
| 13.10.4 | Semi-complete Data Likelihood | 569 |
| 13.10.5 | Laplace Approximation..... | 569 |
| 13.10.6 | Bounding Abundance | 570 |
| 13.10.7 | Data Augmentation | 571 |
| 13.10.8 | Comparing Approaches That Bound Abundance | 573 |
| 13.10.9 | Reversible Jump Markov Chain Monte Carlo | 575 |
| 13.11 | Priors | 576 |
| 13.11.1 | Non-informative Priors for Capture-Recapture | 578 |
| 13.11.2 | Bayes Ancillarity | 579 |
| 13.11.3 | Informative Priors | 584 |
| 13.11.4 | Weakly Informative Priors | 585 |
| 13.11.5 | Hierarchical Models/Priors | 585 |
| 13.12 | Bayes in the Twenty-First Century | 587 |
| 13.13 | Conclusion | 589 |
| 14 | Spatial and Camera Methods | 591 |
| 14.1 | Spatial Recapture Models | 591 |
| 14.1.1 | Introduction | 591 |
| 14.1.2 | Description | 592 |
| 14.1.3 | A Basic Model | 593 |
| 14.2 | Some Examples | 594 |
| 14.2.1 | Multi-catch Traps | 594 |
| 14.2.2 | Passive Detector Array | 595 |
| 14.2.3 | Spatial Passive Acoustical Surveys | 597 |
| 14.2.4 | Some Miscellaneous Applications | 599 |
| 14.2.5 | Using Telemetry | 600 |
| 14.2.6 | Modeling Interactions | 601 |
| 14.2.7 | Occupancy and Spatial Combined | 603 |
| 14.3 | Bayesian Models | 603 |
| 14.3.1 | Binomial Point Process | 604 |
| 14.3.2 | Connecting Bayes and Frequency Methods of Estimation | 607 |

| | | |
|-------------------|-------------------------------------------------------|-----|
| 14.4 | Further Extensions | 609 |
| 14.4.1 | Stratified Populations | 609 |
| 14.4.2 | Presence-Absence Data Only | 611 |
| 14.4.3 | Viability for Introduced Species | 612 |
| 14.4.4 | SECR and Distance Sampling (DS)..... | 613 |
| 14.5 | Spatial Mark-Resight Models | 614 |
| 14.5.1 | Different Sightings for Marked and Unmarked | 616 |
| 14.5.2 | Lack of Individual Recognition | 617 |
| 14.6 | Camera Methods | 619 |
| 14.6.1 | Some Reviews | 620 |
| 14.6.2 | Camera Advantages..... | 621 |
| 14.6.3 | Design Considerations | 622 |
| 14.6.4 | Partial Identity | 624 |
| 14.6.5 | Some Applications | 626 |
| 14.6.6 | Community Method | 627 |
| 14.6.7 | Spatial Point Process Model | 629 |
| 14.6.8 | Modeling Identification Errors | 630 |
| 14.7 | Adaptive Cluster Sampling | 632 |
| 14.8 | DNA Methods | 634 |
| 14.8.1 | Comparing DNA and Camera Methods | 635 |
| 14.8.2 | Combining DNA and Camera Methods | 635 |
| 14.9 | Summary | 636 |
| A | Some General Results | 639 |
| A.1 | Poisson Process | 639 |
| A.1.1 | One-Dimensional | 639 |
| A.1.2 | Two-Dimensional | 640 |
| A.2 | N-Mixture Models | 640 |
| A.3 | Multinomial Distribution | 641 |
| A.3.1 | Some Properties | 641 |
| A.3.2 | Conditional Multinomial Distribution | 642 |
| A.4 | Delta Method | 643 |
| A.5 | Parameter-Expanded Data Augmentation | 645 |
| A.6 | Conditional Expectations | 645 |
| A.7 | Profile Likelihood Intervals | 646 |
| A.8 | Large Sample Hypothesis Tests | 647 |
| A.8.1 | Diagnostic Residuals for Frequency Models | 648 |
| A.9 | Bayesian Methods | 649 |
| A.9.1 | Sequential Updating | 650 |
| A.9.2 | Bayes Factors and Posterior Model Probabilities | 651 |
| References | | 653 |
| Index | | 717 |

Chapter 1

Model Building



1.1 Introduction

We live in an age when every organization seems to be collecting data. For example, if we buy something from a shop or receive some kind of service on the Internet, we often get an email asking us to rate the quality of the service we received. We also live at a time when our environment is under threat, which can affect endangered animal species, food resources, animal populations generally, and even human populations. For example, pollution in our rivers and seas is a growing concern, and, to give just one example, the loss of pollinating bees would be catastrophic. These developments have led to extensive research into studying population changes. We collect various kinds of data and then try and fit a mathematical model to enable us to estimate various population parameters such as population size or population density.

Populations can be categorized as either closed or open, where a closed population is one that is unchanged or at least unchanged during the time of investigating the population. An open population is one that can change due to various processes taking place such as birth, mortality, migration, and so on. The latter probabilistic models described, for example, by Seber and Schofield (2019) can be very complicated mathematically and are not for the faint-hearted! Training people to work in this subject area proposes a challenge. In this book we confine ourselves to closed populations and consider not just animals but also plants and trees and humans. An elementary introduction to closed populations is given by Borchers et al. (2002), with technical material signposted.

In the development of both closed and open models, many models can have a very large number of parameters to estimate, which can lead to estimates with large standard errors. In the past the emphasis was to obtain closed form expressions for estimates, but with increasing model complexity this is becoming impossible, and numerical methods are needed to calculate the estimates and their standard errors.

This has led to a proliferation of computer packages to carry out the computations, so that an ecologist may need to be acquainted with the various packages available.

Complex models can be developed in two basic ways. We can either start with a simple model and then extend it in a number of different directions, a kind of “forward” selection, as in dealing with regression models or else we can start with the most complex model and then try and reduce its size by trying to remove some parameters: a “backward” selection approach. For example, we can put some parameters equal to each other such as assuming a survival rate is constant over time or assuming mortality is the same for males and females. With complex models and special cases, we have the problem of model selection. As mentioned below, the use of suitable criteria for model selection is paramount, and a detailed discussion of this is given by Seber and Schofield (2019, chapter 14), as well as being mentioned briefly in this chapter.

Ultimately we want to find a model that seems to fit the data adequately, but with the minimum number of unknown parameters: the most “parsimonious” model. Clearly in fitting a model, it is important to test the validity of the assumptions underlying the model. In regression, for example, considerable use is made of residual plots, while for multinomial types of distribution used in this book, residuals and goodness-of-fit procedures are commonly used along with model-fitting techniques.

In this chapter we introduce some basic ideas, which are enlarged on in later chapters.

1.2 Model and Design-Based Methods

In the literature a distinction is made between model-based methods and design-based methods. In the former we endeavor to fit some known distribution with unknown parameters to observed data. The normal distribution has always been a useful distribution as it is found, for example, that some natural measurements seem to follow such a distribution. The reason for this is related to the existence of the central limit theorem, where a sum of small changes can lead to approximate normality. Also, under certain general conditions, maximum likelihood estimates are asymptotically normally distributed.

In this book, a model-based example is the use of the negative binomial distribution as a possible model for the number of animals on a sample plot. In another example, the probability of detecting an individual present at some distance y has some density function $g(y)$, and various candidates are considered for this role. With model-based methods, the choice of model can be very critical as a wrong model can lead to considerable bias in the resulting estimators.

The design approach uses a particular design to avoid some bias problems. To demonstrate the difference between the two methods, we can assume for a model-based approach that animals are randomly distributed, which leads to the use of the binomial and Poisson models that may lead to biased estimates if randomness does

not hold. However, if we use simple random sampling, namely, random sampling without replacement, (or even with replacement on occasion) using plots, points, or lines (transects), we can frequently obtain unbiased estimates depending on the method used. When it comes to variance estimation, a good model fit with a suitable distribution can lead to estimates with smaller standard errors than those from design-based models, though again we have to have confidence in our model.

One problem that can arise is that a random sample may lead to an unequal coverage of the population being sampled such as random points being unequally spread throughout a population, or being too close to a boundary leading to what is known as “edge effect.” Also, random transects may turn out to be very close together. For this reason systematic sampling may be used with a random starting point or line, which will give better coverage and cost less. Since a systematic sample can be regarded as a random selection of one cluster, it is not possible to give an unbiased, or even consistent, estimator of the design variance. Wolter (1984) discussed estimators for this scenario for a single or replicated systematic samples, and a more general introduction to variance estimation was given by Wolter (2007). This problem is considered by Fewster (2011), and discussed in a later chapter.

Sometimes both model and design approaches are combined, as, for example, in distance methods where the location of the reference point is designed based, but a model is used for distance measurements. The estimation of the detection function and its variance is model-based, but estimation of the encounter probability and its variance is design-based.

1.3 Classical Frequentist or Bayesian Methods

In classical methods, parameters are regarded as fixed quantities that are typically estimated by the method of moments, maximum likelihood, or simulation. Also, so-called parameter-free methods are sometimes used for estimation. One aspect of the frequentist approach is that we usually ignore any information that we might know about a parameter such as its possible range, e.g., population size is between 100 and 500. Furthermore, large sample properties of estimators are usually invoked such as approximate normality to obtain confidence intervals, though profile likelihood intervals are being used more, even though they are also asymptotic and more computer intensive. With frequentist methods we may end up with a number of different models that perform equally well, but have different estimates, so the problem is to choose one. There is no such thing as a true model; only some are better than others.

Bayesian methods have been used for analyzing ecological data for some time, though there are still some topics where there is little or no use of the methods. Here a parameter is assumed to have some probability distribution or so-called prior distribution. Although there is some arbitrariness here in the choice of such a distribution, priors can be very general, providing little constraint on a parameter

and often provide estimates with good properties. In any case, several priors can be tried. Posterior distributions are then needed, which are then sampled.

One advantage of Bayesian inference is that it is not constrained by asymptotic assumptions and allows for unbiased and precise estimates even with small sample sizes and low detection rates. A second advantage is that a Bayesian formulation using, for example, data augmentation, is readily applied in widely available software (e.g., WinBUGS, JAGS, OpenBUGS), and is more versatile in terms of the model specification. A third advantage is in model fitting. In addition to having a prior distribution for a parameter, we can also have prior distributions for the various alternative models and compute posterior model probabilities, that is, the probability that a model is the “correct” one. We can then use these probabilities to rank the models or else provide a weighted average of the models, where the weights are determined by the model probabilities. Posterior distributions are based on the simple Bayes rule.

$$f(\theta | y) = \frac{f(y | \theta) f(\theta)}{f(y)}.$$

Bayesian inference is based on this posterior distribution using its mean, median, or mode, and constructing Bayesian confidence intervals, known as “credible” intervals, using quantiles of the posterior distribution. For example, a central 95% credible interval would be the range from the 2.5 percentile to the 97.5 percentile. We could also use the so-called highest posterior interval obtained by taking the shortest interval that contains a prescribed fraction of the probability. See also Appendix [A.9](#).

Hooten and Hobbs (2015) have some useful comments about Bayesian model selection for ecologists. They mentioned that model averaging was a tool along with out-of-sample model validation, cross-validation, and “statistical regularization” that links Bayesian and non-Bayesian approaches to model selection. They mentioned the methods of imposing a penalty function or a so-called Lasso. A guide to Bayesian model checking is given by Conn et al. (2018). General Bayesian methods in ecology are given by Link and Barker (2010), Kéry (2010), and Held and Sabanés Bové (2020): chapters 6–9.

Empirical Bayes are frequentist methods that use frequency data to obtain empirical prior distributions, p -values, and confidence intervals. However, because we have the empirical priors, we can also use some of the probabilistic ideas from Bayesian analysis.

Chapter 15 of Seber and Schofield (2019) have further details about such topics as prior probabilities and distributions, simulation methods including Markov chain Monte Carlo sampling, model averaging, general computational techniques, and some computer software available at the time, as well as some related topics.

1.4 Hierarchical Modeling

In a different context, one description of a hierarchical model is a nested sequence (hierarchy) of observed and unobserved random variables. This is a particularly flexible and transparent way to model complex dependencies in observed data and can enter at different stages, where these stages describe conceptual but unobservable latent processes. Cressie et al. (2009) made some useful general comments about hierarchical statistical modeling. Usually, the first level of the hierarchy is a data model, the second is a process model, and the third (optional) level is a parameter model. For example, in this book we can have the probability parameter that an animal is detected, given the animal is present, and then we have the probability that the animal is actually present, which is part of an unobserved (latent) process. More generally we can have

$$\begin{aligned} \text{Pr}(\text{data, process} \mid \text{parameters}) &= \text{Pr}(\text{data} \mid \text{process, parameters}) \\ &\quad \times \text{Pr}(\text{process} \mid \text{parameters}). \end{aligned}$$

The hierarchical model involves conditional thinking of the form A conditional on B , which is conditional on C , and so on. Cressie et al. (2009) provided some helpful practical examples. Kéry and Royle (2016, 2020) discussed hierarchical models, providing useful computer code and emphasizing simulation methods. N-mixture models, described next, are an example.

Hierarchical models can be analyzed using both frequency and Bayesian methods, and either method may be best in some situations. An associated method is hierarchical centering, which is a reparameterization method used for random effects models and can facilitate a reversible jump MCMC algorithm (Oedekoven et al., 2016).

1.5 Mixture Models

Mixture models play an increasingly important role in the ecology of populations. N-mixture models can be described as a class of state-space models in which the true state of the system (abundance) is observed imperfectly. For example, N-mixture binomial (Binmix) models are frequently used in this book, where we end up with binomial distribution that is inflated with zeros. Using only replicated counts and assuming population closure, these models combine a Poisson or similar distribution for spatial variation of latent abundance (N) with a binomial distribution, conditional on N , for the counts. Dennis et al. (2015) showed that the Binmix model with Poisson mixture can sometimes yield estimates of abundance and detection probability p that are infinite and zero, respectively. They found this especially for small abundance, small detection probability, and small number of repeat visits. A generalization of this is a multinomial N-mixture (Multimix)

model, where a multinomial observation model is mixed over a Poisson or similar distribution. However, care is needed in using such models as they can be affected by a number of problems. For example, Knape et al. (2018) discussed the problem of overdispersion and considered the following examples.

Suppose we have R sites with N_i , individuals at site i and the N_i are distributed according to some discrete distribution with probability function g , namely,

$$N_i \sim g(N; \lambda_i, \boldsymbol{\theta}).$$

where the draw for each site is independent. Here λ_i is the expected abundance in site i , and $\boldsymbol{\theta}$ is an optional vector parameter for overdispersion in the abundance distribution. In most applications, g is either a Poisson, a zero-inflated Poisson (ZIP), a binomial, or a negative binomial distribution. The ZIP model is a mixture model with excess zeros, namely,

$$N_i = 0, \text{ with probability } \psi,$$

$$N_i \sim \text{Poisson}(\lambda_i), \text{ with probability } 1 - \psi, N_i \neq 0,$$

where ψ is the probability of an excess zero. For the negative binomial distribution, we can use the parameterization

$$N_i \sim \text{Neg.Bin}(\lambda_i, \theta),$$

where $\text{var}[N_i] = \lambda_i + \theta\lambda_i^2$. The authors emphasized the importance of testing the underlying assumptions and suggested three types of residuals that can be used graphically to check the goodness of fit of the model instead of the computationally intensive bootstrap procedure. Barker et al. (2018) demonstrated that the Binmix model with constant probability of capture can lead to identifiability problems. Kéry (2018) also considered the problem and showed that some Binmix models had satisfactory identifiability.

1.6 Regression Models

Regression methods have become very popular in ecology. For example, we may have the loglinear model

$$\log(y_i) = \alpha_0 + \boldsymbol{\alpha}' \mathbf{x}_i + u_i,$$

where y_i is a measure on plot i and \mathbf{x}_i is a vector of covariate (predictor) observations. The parameters α_0 and $\boldsymbol{\beta}$ are unknown parameters, which may be constants or have prior distributions, and u_i where, if present, may be modeled as a random effect. This, in turn, might have a normal distribution with prior distributions

for the mean and variance. Another popular regression model, used particularly for probabilities (say p_i), is the logistic model

$$\text{logit}(p_i) = \beta_0 + \boldsymbol{\beta}' \mathbf{z}_i + v_i,$$

with comments about the loglinear model also applying here. With the logistic model, we have $\text{logit}(p_i) = \log(p_i/(1 - p_i))$. So-called “link” functions other than the logistic, such as the probit, have found to be useful. Here covariates can be continuous (e.g., water content, tree girth, etc.), discrete, or categorical (e.g., 1 for presence and 0 for absence, types of habitat, 1, 2, 3). With more complex models, covariates can be both spatial and temporal. Regression trees have also been used. In this book individual covariates are extensively used and are becoming more popular

There are standard methods for evaluating linear regression models by examining the residuals (e.g., Seber and Lee, 2003, chapter 10). Variance heterogeneity is a common problem in ecological data, and Cleasby and Nakagawa (2011) showed by surveying the literature that it is a neglected topic. Guisan et al. (2002), in the context of a conference, reviewed the use of generalized linear (GLM) and generalized additive (GAM) models in studies of species distributions. Guisan et al. (2006) considered further models such as the vector extensions (VGLM, VGAM), multiple additive regression splines (MARS), and generalized linear and additive mixed models (GLMM/GAMM). They also mentioned other packages now available to ecologists. Guisan and Rahbek (2011) introduced a package called SESAM. Computing and diagnostics are discussed further in Sect. 9.1.1.

1.6.1 Covariates

Individual covariates can apply at two levels. First, covariates that apply at the sampling occasion and are applicable to all animals are relatively easy to use with current software, and the methodology is well developed. Individual based covariates, however, are more problematical. As we see in this book, such models have been developed for closed populations, and these generally condition upon the covariates observed in the study. Unconditional inference is more difficult as the distribution of covariates in the unobserved population must also be modelled. A major problem is how to model the covariates for the unobserved animals prior to their initial observation or capture.

1.7 Simulation

When introducing a new method of population analysis, which may require the input of new information, there may be a paucity of available real data examples. It has become popular to simulate data using a range of different parameter

values to see how close parameter estimates are to the true values. This approach has strengths and weakness and should be used with caution. On the positive side, it can show how prior distributions can affect the analysis if the data are analyzed using different priors to the ones used for the simulation, or the use of different estimation procedures. The robustness of a method to some departures from underlying assumptions can be checked, particularly with goodness of fit and other tests. Fortunately every article published generally has at least one real application. Computer programs have been emerging to readily facilitate the use of simulation methods.

1.8 Model Fitting and Selection

In general, there is no clear consensus on which criteria to use for model selection, though there are favorites that are commonly used and found to be useful. Some methods are described by Seber and Schofield (2019, 498–507) including some criteria that are also adjusted for over-dispersion. Hooten and Hobbs (2015) discussed the criteria AIC, BIC, DIC, WAIC, and some modifications. For convenience, we consider frequency models and Bayesian models separately, though there can be overlap such as with an empirical Bayes approach and cross-validation. In either case, rather than choose just the “best” model, model (weighted) averaging is often used, where better fitting models according to some criterion are given more weight in the averaging process, as mentioned briefly below. Estimates for different models can be treated the same way. For a good coverage of modeling average

1.8.1 Frequency Models

These are based on frequency data, where maximum likelihood methods predominate. Confidence intervals for the unknown parameters can be based on assuming estimates are asymptotically normally distributed. Alternatively, one can use profile likelihood intervals, which seem to work better, although more computationally intensive. Commonly used selection criteria are AIC and AIC_c, where

$$\text{AIC} = -2 \log L(\hat{\theta}) + 2p.$$

Here the likelihood L is the full likelihood function of a particular model, including constants, and is dependent on an unknown parameter vector θ . This has maximum likelihood estimator $\hat{\theta}$, and p the penalty term is the number of independent estimable parameters. The first term on the righthand side of the above equation is the so-called deviance, and p is a bias adjustment. The aim is to choose models with the lowest AIC values. The reason for the penalty is to discourage overfitting because increasing the number of parameters in the model almost always improves

the calculated goodness of fit. It is assumed that θ is common to all the models, even if some of the θ_i are zero, or we restrict ourselves to the set of models for which this is true. Galipaud et al. (2014) assessed the reliability of the summing of weights (SW) used with AIC. Caution is needed in using SW values as they can take a wide range of possible values, even for predictor variables unrelated to the response. They concluded that “it seems hazardous to define a threshold for SW above which a variable is considered as having a statistical effect on the response and SW is not a measure of effect size.”

Since AIC is based on an asymptotic result, a “finite” version of AIC is

$$\text{AIC}_c = \text{AIC} + \frac{2p(p+1)}{n-p-1},$$

where n denotes the “sample” size. In applications where this measure is applied, there is a problem sometimes of how to choose n . The criterion BIC is given by

$$\text{BIC} = -2 \log L(\hat{\theta}) + p \log n,$$

for large n .

Model selection was also considered for both frequency and Bayesian models by Held and Sabanés Bové (2020).

1.8.2 Bayesian Models

These models are based on using prior distributions for unknown parameters which, combined with data, lead to posterior distributions that can be used to construct so-called credible intervals. Hooten and Hobbs (2015), in discussing Bayesian model selection for ecologists, emphasized the need for checking model assumptions first before engaging in model selection. They commented that neither AIC nor BIC are appropriate for Bayesian model averaging in all situations, as they were designed to be used with maximum likelihood estimates and make fairly strong assumptions about a priori model probabilities. Here AIC excels at finding good predictive models, while BIC was developed mainly for model averaging purposes and is good for small sets of well-justified models. They noted that another criterion DIC is not ideal for all classes of model, and it does not have a theoretical justification for use in model selection. These measures are not easy to apply, or work poorly for complex hierarchical models. A Bayesian information criterion is WAIC, the Watanabe-Akaike information criterion (Watanabe, 2013), which they say has some problems; it needs further research.

Conn et al. (2018) noted that ecologists increasingly use hierarchical Bayesian statistical models in their research, but are less focused on checking model assumptions and assessing potential lack of fit. They say that if we accept the

premise that Bayesian models should be routinely checked for compatibility with data, a logical next question is how best to conduct such checks. Unfortunately there is no single best answer, and they provided a useful flow chart for the whole process. Checks can include “Bayesian P values” based on the intuitive notion that data simulated from the posterior distribution should be similar to the data being analyzed. However, these P values can be considerably conservative (biased toward 0.5) because the data is used twice: once to fit the model and once to test the model. They do detect extreme model inadequacy, but often do not detect more subtle cases of lack of fit.

Alternatively, approaches to model checking such as prior and posterior predictive checks, cross-validation tests, pivotal discrepancy measures, residual tests, and graphical techniques are also considered by Conn et al. (2018). They may produce more accurate characterizations of goodness of fit, but may not be as well-known to ecologists. The authors stated that one third of all of the articles they examined did a commendable job in reporting convergence diagnostics to support their contention that Markov Monte Carlo chains had reached their stationary distribution. However, convergence to a stationary distribution, combined with a lack of prior sensitivity, checks the algorithm but not the model. They also considered providing checks on spatial regression models, assumptions of the N-mixture model, and hidden Markov models.

O’Hara and Sillanpää (2009) provided a helpful earlier review of Bayesian variable selection methods. Broms et al. (2016) considered some methods of model selection and assessment for multi-species occupancy models. They included diagnostic plots, plotted the deviance residuals for each species and site, and then plotted the residuals against the covariates to highlight outliers and areas of lack of fit. WAIC was used for model comparison. Millar (2009) considered the criterion DIC for model selection with respect to overdispersion where the negative binomial and Poisson-lognormal tended to be used. The applicability of DIC was found to vary depending on how it was used, and in some cases was not appropriate.

Variable or model selection (e.g., model averaging) described below) for hierarchical models is complex, computationally challenging, and originally no consensus seemed to have emerged in the literature on a single approach (Link & Barker, 2006). Comparing non-nested models such as a comparison of the fit of Poisson and negative binomial models to count data can be challenging. Approaches to variable and model selection are implemented in BUGS language.

1.8.3 **Model Averaging**

Model averaging amounts to using a weighted mean of estimates from different models with the weights reflecting a measure of the potential value of that model for estimation. The weights used depend on whether a frequentist or Bayesian approach is used. The theory of model averaging is discussed in detail in the book by Fletcher (2018) who described a large variety of applications and was summarized

briefly in Seber and Schofield (2019, appendix A.11.4). Link and Barker (2006) commented that Bayesian model averaging (BMA) is the proper way to obtain multimodel inference under the Bayesian methodology because it gives a valid probability-based mechanism for considering multiple models. AIC, for example, has problems. Link and Barker provided a very helpful discussion about BMA and some philosophical aspects of it (see also Link and Barker, 2010). Dormann et al. (2018) gave an extensive review of model averaging, looking at the mathematical foundations of model averaging along with the diversity of approaches available in ecology, as a large range of different model averaging methods exists. They mention that, for a given set of weights, the prediction error of model-averaged predictions depends on three things: the bias of the model average, as emerging from the bias of the individual models, the prediction variances of the individual models, and the covariance of those predictions. Also, estimating weights can thwart the benefit of model averaging. They give a number of recommendations and conclude that: “While our general results and outlook on this field are positive, in the sense that model averaging is often useful, the complexity of the topic prevents us from providing final answers about the best approach for ecologists”.

Dey et al. (2019) discussed Bayesian model selection for spatial capture-recapture (SCR) models and compared some of the recommended Bayesian model selection tools: (1) Bayes Factor, using (a) Gelfand and Dey (1994) and (b) harmonic mean methods, (2) Deviance Information Criterion (DIC), (3) Watanabe-Akaike’s Information Criterion (WAIC) and (4) posterior predictive loss criterion. It all depends on the level of information available, as when it is low no particular model selection tool can be recommended to help realize, simultaneously, both the goals of model selection and parameter estimation. However, in general, when both the objectives are considered together, the authors recommended the use of their application of the Bayes Factor (Gelfand-Dey with MAP approximation) for Bayesian SCR models. Their study highlighted the point that although new model selection tools are emerging (e.g., WAIC) in the applied statistics literature, previous tools based on sound theory, even under approximation, may still perform much better.

Tenan et al. (2014) considered Bayesian model selection, and explained two well-known procedures, the product space method and the Gibbs variable selection. They clarified theoretical aspects and practical guidelines through applied examples on the comparison of non-nested models and on the selection of variables in a generalized linear model problem.

1.9 Computing

As models have become more complex, we find that there has been a plethora of new programs to carry out the computations. Fortunately many of these have been linked to the package **R**, which enhances availability. As it is not possible to bring them all together, we have generally just mentioned them in a chapter as they arise

in conjunction with a research paper. Just a selection of programs and packages, in random order, follows to give an indication as to the type of package available:

- (1) Vieilledent et al. (2019). hSDM: hierarchical Bayesian species distribution models. R package.
- (2) Sólymos et al. (2020). *detect: Analyzing wildlife data with detection error. R package version 0.4-4*. Available at <https://cran.r-project.org/package=detect>.
- (3) MacKenzie and Hines (2015). RPresence: R Interface for Program PRESENCE. R package version 1.1.1. 2.
- (4) Conn (2012). hierarchicalDS: functions for performing hierarchical analysis of distance sampling data. R package version 2.01.
- (5) Johnson et al. (2014). DSpat:spatial modelling for distance sampling data. R package version 0.1.6.
- (6) Johnson (2012). stocc: fit a spatial occupancy model via Gibbs sampling. R package version 1.0-7.
- (7) McDonald (2010). mra: Analysis of mark-recapture data. R package version 2.7. Available online at <http://CRAN.R-project.org/package=mra>.
- (8) McDonald et al. (2012). Rdistance: distance sampling analyses. R package version 1.1.
- (9) Miller (2017). Distance: distance sampling detection function and abundance estimation. R package version 0.9.7.
- (10) Miller, Rexstad et al. (2014). dsm: density surface modelling of distance sampling data. Rpackage version 2.2.3.
See also Bravington et al. (2021))
- (11) Miller et al. (2019). Distance Sampling in R. Journal of Statistical Software, 89(1), 1–28.
- (12) Efford (2013). SECR: spatially explicit capture-recapture models. R package version 2.6.1. (<http://CRAN.R-project.org/package=secr>).
- (13) Gopalaswamy et al. (2012). Program SPACECAP: Software for estimating animal density using spatially explicit capture-recapture models.
- (14) Phillips and Dudík (2008). Modeling of species distributions with MAXENT: New extensions and a comprehensive evaluation. (<https://rawgit.com/johnbaums/rmaxent/master/README.html>).
- (15) Fiske and Chandler (2011). Unmarked: an R package for fitting hierarchical models of wildlife occurrence and abundance.
- (16) Chao and Shen (2010). *User's guide for program SPADE (Species Prediction And Diversity Estimation)*. Program and User's Guide published at <http://chao.stat.nthu.edu.tw/wordpress/software/download>.

Clearly there is a need here for further unifying packages. This is made difficult due to the different speeds of the methods and the rate of software development. At present, the computational method most commonly used for fitting Bayesian models is Markov chain Monte Carlo (MCMC). Variable selection methods are therefore needed that can be implemented easily in the MCMC framework. For

example, having these models implemented in the BUGS language (e.g., in the older WinBUGS, OpenBUGS, and JAGS) means that the methods can be easily slotted into different models. Some practical details are also given, along with syntax, in the BUGS book by Lunn et al. (2013). More recently Denwood (2016) introduced an R program called **runjags**, which deals with MCMC in JAGS.. Turek et al. (2021) and earlier papers have provided software that gives dramatic improvements in the implementation of MCMC using an R package called **nimble** that “include removing unnecessary computations, integrating out latent states, vectorizing declarations, and restricting calculations to the locality of individuals.” The difficulty with software is that it is continually being updated.

Another computational method that is very popular for estimating biases and variances is bootstrapping. There are two types of bootstrap, parametric in which we use a statistical model to do the resampling, and nonparametric in which the observations are resampled with replacement. For a brief introduction to bootstrap methods in biology, see Manly (1997), and for more general introductions, see Davison and Hinkley (1997) and Efron and Tibshirani (1993). Some useful practical examples are given by Borchers et al. (2002).

1.9.1 Model Diagnostics

Diagnostics to determine model adequacy are often given minimal attention when fitting population models. Various diagnostics involving residuals are available for linear and log-linear models (e.g. Lee & Seber, 2001), and we need similar tools for our models and for studying covariate relationships. Chi-squared goodness of fit tests can have notoriously low power, particularly when a large number of parameters are fitted. A closer look at the data may tell us a bit more about what is going on. Cormack (1989) made some suggestions about appropriate residual plots, but the key problem in capture-recapture models is the sparseness of the individual capture histories. Some sort of automated collapsing of histories to view slices of the data will be important, such as plotting the residual versus the number of capture occasions.

Some tests for closure exist, though no omnibus test exists as existing tests may indicate nonclosure even when a population is closed if other effects are present such as for example trap response, temporary migration (Otis et al. (1978: 120), and heterogeneity (Stanley & Burnham, 1999). Tests for various models are mentioned in passing throughout our book.

1.10 General Study Design

Selecting a method for carrying out a population survey will depend very much on the species being investigated and its mobility, if any. Therefore deciding carefully on the population and any subpopulations to be considered is the first step. Choosing

what parameters are to be estimated will impact on how much data is needed in order to get sufficiently precise estimators of the parameters. The habitat and access to the population area will be important factors on deciding the survey method, e.g., use of roadside surveys or animal trails. The next step is to choose the method of sampling, e.g., plot sampling, point or transect sampling, which involves the choice of the sample unit and the number and size of the units. This will depend on the size and scale of the study area. Methods involving cameras or acoustic arrays will require some thought such as how close together the units should be, as this will depend on the size of the home range or territory of the species being investigated. DNA methods need to consider the number of phenotypes used and account for possible misclassification. Capture-recapture has some special requirements such as choice of type of mark or tag and possible tag loss. Radio telemetry can be used.

The timing of the survey, such as time of the day or hunting season and environmental conditions are important considerations, especially if closed population methods are going to be used with a mobile population. Most methods involve repeat samples so that the number and timing is important, and the question of detectability invariably arises. There needs to be a trade-off between the number of visits and time spent at each visit, as, for example, in occupancy studies. If more than one species is to be considered, then some consideration needs to be given to rare and elusive species so as to obtain enough data. After deciding what data is going to be collected, the question arises of which computer package is to be used. In any method used, there will be specific design questions, e.g., how many observers in aerial or shipboard censusing, and observer(s) location.

1.11 Where to Now?

For both closed and open population, the research keeps expanding because of the various population problems that arise and is partly driven by the development of specialized computer packages that make complicated data sets more accessible. Technological advances have provided better automatic digital information (e.g., DNA analysis, camera trapping, drones, telemetry, acoustical and spatial methods generally) that have generally reduced the labor and the cost of obtaining vital information. This has encouraged the development of noninvasive procedures (e.g., DNA and cameras) with an appropriate protocol, which have less effect on the animal (Zemanova, 2020). Citizen science has also become popular as a cheap method of obtaining data.

A helpful process that is taking place and needs to continue at a greater pace is the integration and updating of computer packages, like **R**, to include the many very specialized packages that are available and referred to above. There is a growing interest in combining various types of data that may or may not be independent. For example, capture-recapture methods have joined forces with distance, removal, and spatial methods. Adaptive sampling, where further sampling is determined by what has turned up already, will begin to invade other standard methods like distance

sampling. For open populations there is the development of more comprehensive integrated populations models that combine different sorts of data (e.g., reviews by Zipkin & Saunders, 2018; Miller et al., 2019, species distributions).

As models become more complicated, there is a greater need for models to be tested as to how well they fit the data, a topic that has been underemphasized.

1.12 Summary

This introductory chapter is about fundamental ideas involved in model selection such as between a model-based and design-based approach, or a combination of both, and between a frequentist and Bayesian-type model. Hierarchical and mixture models are becoming more commonly used along with logistic and loglinear regression models involving covariates. Model selection is always present in studies, with model averaging becoming more popular. This inevitably leads to the development of specialized computational programs, of which there are many; a selection of 17 is listed. Diagnostics for models are considered, including the use of residuals. Finally there are some comments about where we are heading in the subject development.

Chapter 2

Plot Sampling



2.1 Absolute Density

Suppose we have a population of unknown size N of objects spaced over an area A and we want to estimate N and the absolute density $D = N/A$. With some animal populations, for example, insects, population intensity such as number per unit of habitat (e.g., per leaf, per shoot, per plant, per host) is also of interest, particularly if related to plant or host damage. Ignoring practical problems for the moment, suppose that the population area is rectangular and is divided into S square plots (quadrats) each of area a with x_i the number of objects on the i th plot. Then $A = Sa$ and $\sum_{i=1}^S x_i = N$. We now take a simple random sample (without replacement) of s plots from the population and count the number of individuals in each sample plot. This will give us the proportion of the area A sampled as $p = sa/A = s/S$ and the sampling fraction as $f = s/S$. Let $n = \sum_{i=1}^s x_i$ be the total count on the s plots, and $\bar{x} = n/s$ be the mean of the individuals on the sample plots. This is an unbiased estimate of the population mean N/S . Dividing the estimate by a to get the mean sample density and then multiplying up to area A we get from standard sampling theory for \bar{x} that

$$\hat{N} = \bar{x}S \quad (2.1)$$

is an unbiased estimate of N with variance

$$V_N = \frac{S^2}{S-1} \left[\sum_{i=1}^S (x_i - \frac{N}{S})^2 \right] \frac{1-f}{s}, \quad (2.2)$$

where the definition of variance has $S - 1$ instead of S . An unbiased estimate of V_N is given by

$$v_N = S^2 v \frac{1 - f}{s}, \quad (2.3)$$

where

$$v = \frac{1}{s - 1} \sum_{i=1}^s (x_i - \bar{x})^2. \quad (2.4)$$

An approximate two-sided $100(1 - \alpha)$ percent confidence interval for N is given by $\hat{N} \pm t_{s-1}[\alpha/2]\sqrt{v_N}$, where t_{s-1} represents the t distribution with $s - 1$ degrees of freedom and the probability that $t_k \geq t_k[\alpha/2]$ is $\alpha/2$.

A useful measure of the precision of an estimate is the coefficient of variation, the standard deviation of the estimator divided by its mean. In the case of \hat{N} this is $C = \sqrt{V_N}/\hat{N}$, which can be estimated by $\hat{C} = \sqrt{\hat{v}_n}/\hat{N}$. An estimate of population density is then $\hat{D} = \hat{N}/A$ with the confidence interval divided through by A .

2.2 Design Considerations

In sampling some populations, there are practical issues relating to the extraction of samples from a particular environment such as soil or water (Henderson, 2021). In order to use the theory of the previous section, a number of design considerations need to be taken into account. These include dealing with regular- and irregular-shaped population areas, choosing the plot shape, and its area. We also need to decide what proportion of the population area we should sample (in terms of some precision criterion), which leads to determining plot area and number of sample plots. These requirements are now discussed individually below. In practice it is a good idea to measure the amount of effort put into carrying out a study for future use and for costing future proposals.

2.2.1 Plot Shape and Area

Our first question is plot shape. For a given area, we find that a rectangular plot has the shortest total boundary when it is a square. This is useful if there is some edge effect with an individual overlapping two plots as with some plants, for example, or with animals (e.g., small mammals) moving in and out of the population area. This boundary reduces as the plot tends to a circle of the same area, but circles or plots with curved sides are much more difficult to sample as they do not fit together. However, for sampling a large population, rectangles (strip transects) may be more

efficient to use in some situations. However, circular plots of small area are useful for sampling soil animals where we do not want to dig up too much of the population area. In general, square plots are most convenient to sample, but it should be born in mind that the edge effect discussed in Sect. 2.4 increases as the area decreases. It is a good idea to have a convention for edge overlaps such as including those individuals in the top lefthand corner and left side, for example.

The next question is to determine what the plot area a should be. For a given proportion p of the population area sampled, should we have a large number of smaller plots, which will give estimates with smaller variance in some situations (e.g., with individuals randomly dispersed) or fewer larger plots? The latter may be easier to set up and with comparatively less boundary effects. Normally we would like both S and s to be as large as possible so that large sample theory can be used. However, a should be not too small, and a rough rule (Greig-Smith, 1964) would be to have no more empty quadrats than those with just one individual. We begin by determining what size area we should sample. We do not want too many empty quadrats. This kind of situation with few sparse highly clustered objects is discussed briefly in Sect. 2.8 mentioning adaptive sampling. It is often of value to take at least two different-sized sampling units.

2.2.2 Proportion of Area Sampled

We wish to choose p , the proportion of the population area to be sampled. This will be determined by what sort of precision is required, as determined, say, by the coefficient of variation C above. Beginning with a simpler case in which we assume the N objects are randomly distributed and therefore independent of one another, then n has the binomial distribution with probability function

$$f(n) = \binom{N}{n} p^n q^{N-n} \quad (q = 1 - p).$$

This gives rise to the unbiased estimator

$$\widehat{N} = n/p \tag{2.5}$$

with variance Nq/p and coefficient of variation $C = \sqrt{q/(Np)}$, or

$$p = 1/(1 + NC^2). \tag{2.6}$$

Therefore, given a rough lower bound for N and a given C , we can calculate a conservative p . For example, if we follow the usual recommendation of p in the range 5–10%, say $p = 0.1$, and choose $C \leq 0.2$, then we would have to have $N \geq 225$. This means we would need to sample a very high proportion of the population to obtain relatively precise estimates for small populations. If $p < 0.2$, so that $\sqrt{q} >$

$0.94 \approx 1$, and since $E[n] = Np$ and $\text{var}[n] = Npq$, we can estimate C by $\hat{C} = \sqrt{q/n}$, which is approximately $1/\sqrt{n}$. This means that C depends approximately on the number of individuals seen when less than 20% of the population area is sampled. Once we have determined p and a we then have $s = pA/a$, the number of plots to be sampled. We now remove the assumption of random distribution.

2.2.3 Number of Plots

Unfortunately most populations are not randomly distributed as, for example, plants tend to scatter seeds locally so that there is clustering effect. This can also happen with small mammals as they are social animals using a similar habitat to shelter and are therefore not independent of one another. This means that the variance of n is greater than Npq and the above p in (2.6) will be underestimated. An alternative approach is to choose a so that S is large and carry out a pilot survey using, say, s_0 plots. Calculating \bar{x} and v from this sample, a little algebra then shows that we need to sample $s - s_0$ more quadrats, where

$$\frac{1}{s} \leq \frac{1}{S} + \frac{C^2 \bar{x}^2}{v}, \quad (2.7)$$

to give a prescribed C .

Hou et al. (2015) were particularly interested in plot number and in a particular method of correcting for edge effect (see Sect. 2.4). Using simulation they concluded that with regard to plot size, and a given total proportion p of the area sampled, the smaller the plot, the better the sampling precision. Their paper is referred to for further general conclusions.

Sampling Method

To carry out simple random sampling, one needs to locate and number the plots from $1, 2, \dots, S$, and then randomly without replacement generate a subset of s of the S numbers. This could be done, for example, by using a GPS system to locate the plots and then using random number tables or appropriate software.

2.2.4 Irregular Areas

Life is not always simple and often population areas are irregular in shape so that if a grid of quadrats is placed on the population area, some areas on the boundary will not be covered. Generally these boundary areas can just be ignored if the population area is large. Sometimes the population area may already be divided up

into T plots of varying shapes and sizes by natural or human-made barriers such as watercourses, fences, paths or roads, and so on. However, if a simple random sample of s plots is carried out, some of the above theory still applies, but with T instead of S . For example, we still have $E[\bar{x}] = N/T$, the population mean per plot, so that $\hat{N} = \bar{x}T$, and V_N and v_N are the same as above but with T instead of S . If the population is randomly distributed, then Eq. (2.1) and the theory that follows still apply. The estimate of population density is still \hat{N}/A . However, the problem with unequal-sized plots is to determine the required sample size s . If the plots are small and \bar{a} is the average plot area (if available), then we could perhaps use s , where $s\bar{x} = pA$. If the population is not randomly distributed, we can still use a pilot survey and Eq. (2.7).

2.2.5 Trees

We digress briefly to look at forests, as sampling trees has its own set of differences from the usual plot sampling. To begin with, the centers of the sample plots are usually systematically spaced on line transects and systematic sampling is used with equally spaced line transects, commonly referred to as “timber cruising.” The percentage of the area sampled is referred to as the “cruise percent.” The number of trees is counted on each sample plot, and various measurements are made on each tree, e.g., girth.

As noted by Hou et al. (2015), the attributes of a tree are rarely independent of those of its neighbors, so that forest attributes are often spatially autocorrelated, and spatial patterns tie in closely with spatial autocorrelation in that the former are indicated by the statistics of the latter. In the case of circular plot sampling, trees within a given distance from the sample plot location constitute a sample. If the sample is collected using a technical device located at the sampling point, e.g., the increasingly popular terrestrial laser scanner, all trees of the sample plot cannot be observed because they can hide behind each other.

Kansanen et al. (2021) proposed a Horvitz-Thompson-like estimator with distance-based detection probabilities derived from stochastic geometry for the estimation of population totals such as stem density and basal area in such situations. They focused on the density (the number of tree stems per unit area) and basal area (the sum of the cross-sectional areas of the stems at breast height per unit area).

Another variation in tree sampling is when there is interest in trees outside forest (TOF) over large areas. A TOF unit is intended as a small group of trees within rural and urbanized areas, and not classified as forest. They involve certain sampling problems discussed by Fattorini et al. (2016). They checked the performance of point and plot sampling using aerial photoimagery.

2.2.6 Primary and Secondary Units

Spatial patterns may be categorized as regular, random, or aggregated, although combinations of these are commonly seen in forestry. Spatially systematic sampling has been popular for forest inventories partly because of its convenience in field-work. It has also been found to have superior efficiency with spatially autocorrelated populations except in the rare case where the spatial variation period of an attribute coincides with that of the systematic design. Each transect can be regarded as a primary unit, and the plots on the transect as secondary units. We now briefly summarize some of the theory (e.g., Thompson, 2012, chapter 12).

Suppose we have S primary units in the population with a total of N objects, say trees, in a population area A , and a simple random sample (without replacement) of s primary units is selected. Let M_i be the number of secondary units in the i th primary unit of the sample, let $M = \sum_{i=1}^S M_i$, the total number of secondary units, and let y_{ij} be the value of the variable of interest of the j th secondary unit in the i th primary unit of the sample. The population mean per primary unit is $\mu_1 = N/S$, and population mean per secondary unit is $\mu = N/M$. Let $y_i = \sum_{j=1}^{M_i} y_{ij}$ be the sum for the i th primary unit. Considering just primary units, an unbiased estimator of N is

$$\widehat{N} = \frac{S}{s} \sum_{i=1}^s y_i = S\bar{y},$$

where $\bar{y} = \sum_{i=1}^s y_i/s$ is the sample mean of the primary unit totals. Then,

$$\text{var}[\widehat{N}] = S(S-s) \frac{\sigma^2}{s}, \quad \text{where } \sigma^2 = \frac{1}{S-1} \sum_{i=1}^S (y_i - \mu_1)^2,$$

the finite-population variance of the primary unit totals. An unbiased estimate of the above variance is

$$\widehat{\text{var}}[\widehat{N}] = S(S-s) \frac{s^2}{n}, \quad \text{where } s^2 = \frac{1}{s-1} \sum_{i=1}^s (y_i - \bar{y})^2.$$

The corresponding unbiased estimator of the population mean per secondary unit (μ) is therefore

$$\widehat{\mu} = \frac{\widehat{N}}{M}.$$

It is not uncommon to use a single primary unit ($s = 1$) with M_1 secondary units with forestry, for example, to reduce costs and simplify forestry operations. If the primary unit is chosen at random, then the first secondary unit is essentially

chosen at random with the rest of the secondary units obtained systematically in a linear fashion. However, we then have a problem of obtaining a variance and its estimator. If we treat the M_1 units as a simple random sample from the M secondary units, there will be a tendency with natural populations to grossly overestimate the variance of \hat{N} because of positive correlation between neighboring units. Some further comments on this problem are given by Murthy and Rao (1988).

If the primary unit total y_i is highly correlated with primary unit size M_i , then the following ratio estimator may be preferred, namely, (Thompson, 2012: 160)

$$\hat{N}_r = rM, \text{ where } r = \frac{\sum_{i=1}^s y_i}{\sum_{i=1}^s M_i},$$

which will have small bias with large samples. An approximate formula for the variance (actually mean-square error) is

$$\text{var}[\hat{N}_r] \approx \frac{S(S-s)}{s(s-1)} \sum_{i=1}^s (y_i - M_i \mu)^2,$$

with estimator

$$\widehat{\text{var}}[\hat{N}_r] = \frac{S(S-s)}{s(s-1)} \sum_{i=1}^s (y_i - M_i r)^2.$$

Thompson (2012, chapter 7) gave several variance estimators including the “adjusted” estimator

$$\widetilde{\text{var}}[\hat{N}_r] = \left(\frac{sM}{S \sum_{i=1}^s M_i} \right)^2 \widehat{\text{var}}[\hat{N}_r].$$

2.3 Stratified Sampling

If some parts of the population are very different with regard to the distribution of population, it may be appropriate to divide the population into domains or strata and sample each stratum separately. Let J be the number of strata, and let A_j ($j = 1, 2, \dots, J$) be the area of the j th stratum. Assuming, for completeness, the general case where the plots in strata are of different sizes, let T_j be the number of plots in stratum j .

2.3.1 Estimation

Following Thompson (2012, 143), suppose that s_j plots are chosen at random (without replacement) from the T_j plots in stratum j , and let x_{ij} be the count on the i th sample plot ($i = 1, 2, \dots, s_j$; $\sum_j s_j = s$, $\sum T_j = T$). If

$$\bar{x}_j = \frac{1}{s_j} \sum_{i=1}^{s_j} x_{ij}$$

is the average sample count for stratum j , then from Eq. (2.1) in Sect. 2.1, $N = \sum_j N_j$ is estimated by

$$\tilde{N} = \sum_{j=1}^J \hat{N}_j = \sum_{j=1}^J \bar{x}_j T_j,$$

and an unbiased estimate of the variance $\text{var}[\tilde{N}]$ is given by

$$\tilde{\sigma}_N^2 = \sum_{j=1}^J \tilde{\sigma}_{N_j}^2 = \sum_{j=1}^J T_j(T_j - s_j) \frac{v_j}{s_j},$$

where $v_j = \sum_i (x_{ij} - \bar{x}_j)^2 / (s_j - 1)$. A $100(1 - \alpha)$ percent confidence interval for N is given by $\tilde{N} \pm t_k[\alpha/2] \tilde{\sigma}_N$ for small samples, where k , the degrees of freedom is given by Satterthwaite's (1946) correction

$$k = \frac{\tilde{\sigma}_N^4}{\sum_j \left(\frac{\tilde{\sigma}_{N_j}^4}{s_j - 1} \right)}.$$

We emphasize that we do not have to assume that the individuals are randomly distributed in each stratum.

2.3.2 Choosing Number of Plots

If estimates \tilde{N} and the v_j are available from a pilot survey or past data, then using these estimates for the true parameters for a given s , the optimum value of s_j that

minimizes $\text{var}[\tilde{N}]$ is, by Thompson (2012, 147),

$$s_j = \frac{s T_j \sqrt{v_j}}{\sum_j T_j \sqrt{v_j}} = s r_j, \quad \text{say.}$$

With this choice of s_j , we find that for a given coefficient of variation C , we use (Seber, 1982, 27)

$$s = \frac{\sum_j T_j^2 v_j r_j^{-1}}{\tilde{N}^2 C^2 + \sum_j T_j v_j}.$$

If, however, proportional sampling is used so that $s_j/T_j = s/T$ for all j , then we simply set $r_j = T_j/T$ in the above equation.

In some practical situations, it may not be possible or practicable to use optimum sampling so that we have to use the pilot survey or past data to obtain \tilde{N} with a coefficient of variation no greater than C . Calculating \bar{x} , v_j , and \hat{N}_j from the pilot sample, we choose s_j so that the coefficient of variation of the estimate of N_j will be no greater than C . Then, from (2.7) in Sect. 2.2.3

$$\frac{1}{s_j} \leq \frac{1}{T_j} + \frac{C^2 \bar{x}_j^2}{v_j}.$$

Defining

$$b_j = \left\{ T_j \frac{v_j}{s_j} (T_j - s_j) \right\}^{1/2}$$

we have $b_j/\hat{N}_j \leq C$, and

$$\widehat{C}[\tilde{N}] = \frac{(\sum b_j^2)^{1/2}}{\sum \hat{N}_j} < \frac{\sum b_j}{\sum \hat{N}_j} \leq \max_j \left\{ \frac{b_j}{\hat{N}_j} \right\} \leq C.$$

2.3.3 Having a Random Distribution

Let p_j be the proportion of the area A_j sampled, and let $n_j = \sum_i x_{ij}$ be the total number of objects counted in stratum j . If the objects are randomly distributed within each stratum, then \tilde{N} can now be expressed in the form $\sum_j (n_j/p_j)$. The variance of \tilde{N} is, from the binomial variance,

$$\text{var}[\tilde{N}] = \sum_j (N q_j / p_j).$$

To determine the optimum allocation of sampling effort, we minimize $\text{var}[\tilde{N}]$ subject to

$$\sum_j p_j A_j = \text{constant} (= pA, \text{say})$$

using a Lagrange multiplier, and obtain (Seber, 1982, 26)

$$p_j = K\sqrt{N_j/A_j} = K\sqrt{D_j},$$

where $K = pA / \sum_j \sqrt{D_j}$. If these values of p_j are used, we find that

$$\text{var}[\tilde{N}] = \frac{[\sum_j \sqrt{N_j A_j}]^2}{pA} - N,$$

and for a given coefficient of variation C , we must have

$$p = \frac{[\sum_j \sqrt{N_j A_j}]^2}{AN(1 + NC^2)}.$$

If rough estimates of the N_j are available, then for a given C , we can determine p and hence the p_j from the above equations.

We see that for optimum allocation, we must choose p_j proportional to the square root of the population density D_j in the j th stratum. This means we can use even relative density estimates (discussed in Sect. 2.6) for determining the p_j . For example, if $p_j = K_1 \sqrt{R_j}$, where R_j is an estimate of the relative density (the same units being used in each stratum), and p is determined in advance (by the resources available for sampling), then

$$K_1 = pA / \sum_j (A_j \sqrt{R_j}),$$

and the p_j can be calculated. If proportional allocation of area is used, so that $p_j = p$, then

$$\tilde{N} = \sum_j n_j / p$$

and p can be determined from (cf. (2.6)),

$$p = 1/(1 + NC^2).$$

2.3.4 Plot Sampling with Animals

In an animal population, the time spent on searching a plot, which is related to the size of the plot, is critical because of possible disturbance to the population and consequent undetected animal movement in and out of the plot. In this case distance methods discussed in later chapters may be more useful. Another problem is that it is sometimes hard to determine whether an animal near the edge of a plot is in or out of the plot because of movement. This leads to the problem of so-called edge effect, which we consider in the next section. There is one further problem which occurs with both animals and plants, and that is the problem of incomplete detection and estimating the probability of detection. This is considered in Chap. 3.

2.4 Edge Effect

In sampling plants and trees, for instance, there is a problem when a sample plot overlaps the boundary of the population of area A . Such plots will have a portion of their area outside the boundary so that individuals in such a plot will have a smaller probability of being sampled than those in interior plots. The species composition and size of trees near the edge of a forest are often markedly different from those in the interior, because of the difference in light and water availability near the edge. As noted by Gregoire and Scott (2003), sampling practitioners will often establish a periphery zone inside the boundary and prevent any sample plot or line from being located within that zone. For example, with circular plots of radius R , a zone as wide as R can be used to ensure containment. This means moving a randomly chosen plot orthogonally to the boundary and within the zone if it overlaps the boundary, referred to as the “pullback method.” Here any plot within a distance R from the tract boundary is pulled back into the interior of the tract until its center is exactly R from the edge so that a circular plot will be tangent to the tract boundary.

With a randomly chosen center point at a distance $2R$ from the boundary, the circular plot of area $a = \pi R^2$ is in the inclusion area, and the probability of selection will be $p = a/A$. Whereas if in fact an overlapping plot has area a' inside the boundary, then the probability of the plot selection is $p' = a'/A$ rather than p if the pullback method is used. Clearly this will lead to estimation bias. If the distance of the center is d from the boundary, then Gregoire and Scott (2003) derived different inclusion probabilities depending on d , and, in certain conditions on d , an extra rectangle of area has to be added. They found that with the pullback method, those closer than approximately $0.8R$ from the edge will have smaller than nominal selection probability, that is, $p' < a/A$, whereas those between $0.8R$ and $2R$ from the edge will have greater than the nominal selection probability.

The authors concluded that when the species composition near the edge differs from that found more to the interior, “species occurring close to the edge would be underestimated whereas those occurring greater than $0.8R$ to the interior would be

overestimated.” Also, “The use of the pullback procedure is inadvisable: not only does it fail to eliminate so-called boundary bias, it may actually exacerbate bias and mean square error of estimation of individual species characteristics.” One method that endeavors to allow for the edge effect is to use the so-called mirage point, first proposed by Schmid-Haas (1969) and described in Sect. 2.4, along with some related methods.

Our focus is on estimating population numbers, but population density can be more useful for comparing different populations. However, apart from a fixed population such as with plants, we can have difficulty with defining a population area, especially with small mammals that move around. If grid trapping is used for them, then simply dividing the total count N by the area of the experiment A will generally lead to a major overestimation of the density $D = N/A$. This problem is also referred to as the edge effect where the traps on the boundary of the trapping area tend to catch more animals than the inner traps leading to an overestimation of population size. This edge effect is due to immigrants and to animals living outside the trapping area attracted by the bait, or animals have home ranges (foraging grounds) overlapping the population area so that the effective trapping area is greater than A . Another term used in the literature is “center of activity”, which is a kind of “central point” of the home range.

2.5 Home Range

The idea of a home range was apparently introduced by W. H. Burt in 1943 using maps and refers to the area that an animal generally occupies during the day. This area will depend on many factors such as species, size of the animal, environment with weather and location having an effect, population density where there might be competition for food, and social behavior of the animals. A number of methods that attempt to quantify the home range of an animal based on capture-recapture data have been used in the past (e.g., Seber, 1982: section 12.1).

A variety of shapes have been considered for a home range such as circles, ellipses, or long thin areas and using bivariate distributions to model the probability of capture from the “center” of the home range. They have been used for a variety of species including even fish (Palmer et al. (2011)). Unfortunately, for some animals the trap-revealed range can change an animal’s behavior. Even if the actual home range stays constant, there will be a tendency for the trap-revealed range to grow as the number of recaptures increases with a consequent overestimation of the home range. However, this trend tends to level off after a certain number of captures, though this number varies in the literature.

Following Dice (1938), a common method in the past to determine the “effective trapping area” was to add a strip round the population area of width $W = R/2$, where R is some sort of linear measurement of the home range, for estimating population density (cf. Seber, 1982, 51, 446–449). Other measures that have been used include the mean distance moved by individuals between successive captures

and the mean maximum distance moved between recaptures for animals caught at least twice (see also Tioli et al., 2009, who used radio-tracking). Dice's method has little theoretical justification, and there is doubt on its utility for wide-ranging carnivores (cf. Thompson et al., 2012: 864, for some references).

In the past various grid-trapping schemes were used for dealing with animals involving trapping grids to estimate W , or W and the density D at the same time (e.g., Seber, 1982: 51, 449). An interesting method using geometrical ideas, which was found to be helpful, was given by Link and Barker (1994). Unfortunately, the reliability of past methods is questioned, especially with removal trapping, as depletion of the population will tend to affect home ranges. However, with improving technology, we can avoid multiple trapping and simply have a single trapping to attach radio transmitters that can be combined with global positioning systems (GPS) technology to pinpoint locations at regular time intervals. This technology can be used to establish both home ranges and utilization distributions, the latter being represented by isopleths demarcating regions in space with different probabilities or rates of usage by individuals (e.g., Steiniger & Hunter, 2012). The technology can, for example, be used even for fish (Roux et al., 2018, Tiger fish). Once we have a set of location points, there are a number of methods for determining the home range, which we mention just briefly as our main concern is with the edge effect.

The simplest method is to join up the location points and construct the convex hull, the minimum convex polygon, containing them. Although still popular, it has a few problems and can end up overestimating the home range (Burgman & Fox, 2003). For example, this and other methods can be very sensitive to outlying points and may well reflect exploratory animal movements rather than those necessary for survival and perhaps reproduction. Getz and Wilmers (2004) suggested taking the union of the minimum convex polygons associated with the $k - 1$ nearest neighbors of each location point in the data and obtained better results than those associated with kernel and α -hull methods. Getz et al. (2007) extended this further to develop a method called the local convex hull, which also uses a non-parametric kernel method.

Other methods such as parametric kernel methods are also used (Worton, 1995) that utilize various bivariate distributions. Horne et al. (2007) developed a Brownian bridge movement model for estimating the expected movement path of an animal, using discrete location data obtained at relatively short time intervals. Downs et al. (2011) presented time-geographic density estimation as a new technique of animal home range analysis. It can be applied to animal tracking data to create a density surface, so that home ranges and core areas can be delineated using specified contours of relative intensity (e.g., 95% or 50%). Long and Nelson (2012) have a method called Potential Path Area home range. An interesting issue is raised by Powell and Mitchell (2012) who suggested there are two sides to the question "What is a home range?" In addition to trying to measure it, they suggested shifting the focus towards how animals cognitively view their environment.

Clearly software plays an important role in implementing various methods such as those referred to above that have been developed for LoCoH, Brownian bridge,

line-based kernel, GeoEllipse, and line buffer techniques. Available packages online include LoCoH, Animove, and particularly OpenJUMP HoRAE and AdeHabitat for **R**.

2.6 Relative Density

Although the above discussion focuses on individuals, we can use the methods there for counting animal signs such as bird calls, animal droppings, burrows, visits to scent stations, catch per unit effort, and tracks in the snow for large animals. Roadside counts are also used extensively as we shall see later in this book. Clearly the methods of population estimation in this chapter can be applied to a wide variety of signs, even something as small as ghost shrimp burrow openings (e.g., Butler & Bird, 2007).

The density estimates for signs now refer to relative density giving us an “index” of population density and are useful for determining population changes. These indices are based on the idea that the number of signs or searching effort is approximately proportional to the density and that the rate of proportionality is (hopefully) constant. If the index doubles, we can then assume that the population has roughly doubled, even if we do not know the actual size of the population. Similar indices are used in fisheries where catch per unit effort (CPUE) is used as a measure of relative abundance, considered in Chap. 11.

Sometimes an index can be converted to a population density if we have a reliable and stable “correction factor” of the average number of signs per animal. For example, Plhal et al. (2014) used pellet groups from transects to estimate the population density of wild boar, a native animal species of the Czech Republic that has significantly increased its population density in recent years due to the decline in the intensity of hunting pressure. Both the quality and quantity of food accessible to free-living wild boars vary markedly throughout the year so that the defecation rate can also fluctuate. Censusing during a winter period reduced the decomposition rate of the pellets, and supplementary feed helped to maintain a closed population. The daily defecation rate per individual was determined from animals in an enclosure with different rates for adults and piglets, and the total count was converted to a density. The distribution on transects was uneven and varied with habitat, and the number of zero counts was high.

2.6.1 Index Methods

We now consider further examples and some theory for index methods. Many surveys of vocal anurans consist of calling surveys in which an observer records a calling index such as the “Wisconsin index” used by the North American Amphibian Monitoring Program (NAAMP). It takes on values 0–3, where zero is equivalent

to not detected; one represents discrete, no overlapping calls; two corresponds to discrete, overlapping calls; and three indicates a full chorus (Weir & Mossman, 2005). It is generally difficult to count individuals so that most anuran monitoring efforts collect ordinal data, with levels representing the intensity of vocalizations, which are used as an index to population size. Royle and Link (2005) considered the following multinomial mixture model where s sites are sampled at times $t = 1, \dots, T$. Let N be the number of possible index classes determined by the calling-index definition. We shall use the NAAMP as an example where $N = 4$, so that N takes the values 0, 1, 2, and 3.

Let N_i be the latent abundance class for site i , which may or may not be observed during animal activity and the probability distribution of the N_i (among sites) as the latent abundance class distribution. The object of inference is the probability distribution of N given the calling index histories for a sample of sites, i.e., the probabilities $\psi_k = \Pr(N = k); k = 0, 1, 2, 3$. Let y_{it} be the index value for site i at time t . For example, $y = (020)$ represents indices for three replicate samples at a single site.

Given $N_i = k$, y_{it} has a multinomial distribution with four cells and cell probabilities $\pi_{kj} = \Pr(y = j \mid N_i = k)$ with $\pi_{kj} = 0$ for $j > k$, and $\pi_{k0} = 1 - \sum_{j=1}^3 \pi_{kj}$. This means that π_{kk} is the probability that the true abundance state is observed and $\pi_{jk}, j < k$, are the misclassification probabilities. Alternative parameterizations are considered, and Royle and Link (2005) should be consulted for further details.

Index methods have been widely used for birds, as indicated by the review paper of Rosenstock et al. (2002) relating to landbird surveys. They suggested distance sampling, discussed in Chaps. 7 and 8 chapter, as an alternative. Acoustic methods are described further in Sect. 8.8.

Johnson (2008) discussed some of the advantages of using indices (with a focus on bird surveys), particularly as related to the problem of incomplete detectability, where some of the standard methods have some weaknesses (e.g., point-count methods, cf. Chap. 7). He provided a good survey of the methods at the time and the problems associated with each one, where underlying assumptions may not be fulfilled. Johnson stated: “I argue here that, although the criticisms of indices are largely valid, the cures proposed for the disease may be little or no better than the original ailment.”

Bart et al. (2004) provided support for the use of population indices. They noted that some researchers have expressed the view that index methods should virtually never be used, or at least not on their own. (cf. Anderson, 2001). Others have responded by defending index methods and questioning whether better alternatives exist. Although index methods can be cost-effective, they recommend that double-sampling (Eberhardt & Simmons, 1987), or another procedure that corrects for bias or establishes bounds on bias is essential.

As indices can be used for looking for trends in population density estimation, Bart et al. (2004) commented that a constant detection detection rate is not needed, but rather that there is no trend in the rate (e.g., no change more than about 15-20%

during a survey), which is “probably approximately met by many well-designed index surveys.” In considering bird populations, they discussed in detail four sources of bias: coverage (how completely the sampling frame covers the population of interest), closure (how many of the target birds are in the study area during the study period), surplus birds (e.g., migrant birds), and detection rates. Care is needed in balancing bias and precision (measured by mean-square error) as a biased estimate may be more precise.

2.7 Population Distribution

2.7.1 Types of Distribution

In practice, there are three types of population distribution: regular, random, and aggregated. We saw above that if individuals are randomly distributed, the distribution of x individuals in a sample plot of area a is $\text{Binomial}(N, p)$ where $p = a/A = 1/S$. Often S , the number of plots, is large so that p is small, and N is large such that Np is moderate. If we formally let $N \rightarrow \infty$ and $p \rightarrow 0$ such that Np is a constant μ , this leads to the Poisson distribution $\text{Poisson}(\mu)$, where μ is the mean of x and also the variance of x . Alternatively, we can simply assume that the number of individuals in a sample plot is from a Poisson process with mean Da , where D is the population density. If x_1, x_2, \dots, x_s are the numbers in the s sample plots, then a test of randomness is equivalent to testing the hypothesis that these observations come from a Poisson distribution. For doing this a standard chi-squared goodness-of-fit test can be used or the Poisson Dispersion test

$$T = \sum_{i=1}^s (x_i - \bar{x})^2 / \bar{x} = \sum_x f_x(x - \bar{x})^2 / \bar{x},$$

which is asymptotically distributed as χ_{s-1}^2 when the hypothesis is true. For further details, see, for example, Seber (1982, section 1.3.7).

Before looking at other distributions, we note that for some populations (e.g., crops like wheat), seeds are planted in rows, but plant spacing in a row can vary. Density can be estimated if the number of plants in a row can be estimated if plant distance can be fitted to a distribution. For example, Liu, Baret et al. (2017a) used a gamma distribution.

2.7.2 Negative Binomial Distribution

As noted above, most populations are not randomly distributed. This means that there will be more empty plots and more plots with many individuals than expected,

so that the variance of x will be greater than the mean. When the Poisson distribution is not applicable, a number of alternative distributions, usually referred to as contagious distributions, have been put forward to account for the spatial patterns observed. The most popular, because of its flexibility, is the negative binomial distribution (NBD)

$$f(x) = \frac{k(k+1)\cdots(k+x-1)}{x!} P^x Q^{-x-k}, \quad x = 0, 1, \dots,$$

where $f(0) = Q^{-k}$, $Q = 1 + P$ and $k > 0$. It has mean kP and variance $kP(1+P)$. From Sect. 2.1,

$$N = E[\hat{N}] = SE[\bar{x}] = SkP$$

so that

$$\text{var}[\hat{N}] = \frac{S^2}{s} \text{var}[x] = \frac{S^2}{s} kP(1 + P).$$

We see then that $kP = N/S$, the population density per plot. Also the square of the coefficient of variation of \hat{N} is

$$(C[\hat{N}])^2 = \text{var}[\hat{N}]/N^2 = \frac{1}{s} \left(\frac{S}{N} + \frac{1}{k} \right).$$

Rearranging this last equation yields

$$s = \frac{1}{C^2} \left(\frac{S}{N} + \frac{1}{k} \right),$$

so that if approximate estimates of N and k are available, s can be determined to provide a prescribed C . If a pilot sample is used, then $N/S = kP = d$ can be estimated by \bar{x} from the survey, and k can be estimated by various computer methods that are available.

From above we have that d is the population density of the plot. If x , the number on a sample plot, has a negative binomial distribution, then using $\pi_1 = \Pr(x > 0)$,

$$1 - \pi_1 = \Pr(x = 0) = (1 + P)^{-k} = (1 + d/k)^{-k}.$$

From the above equation, we have $-k(\log(1 + P)) = \log(1 - \pi_1)$ giving us

$$d = k[(1 - \pi_1)^k - 1] \quad \text{or} \quad d = Pk = -P \log(1 - \pi_1)/\log(1 + P).$$

Then a quick estimate of d can be found if π_1 is estimated by the number of nonempty plots and a prior estimate of k or P is available. An alternative version of

the NBD using as parameters the mean and variance of x is given by Linden et al. (2011).

We note that the NBD can be fitted to a wide range of distributions and different scenarios beginning with the Poisson that has the dispersion parameter $k = \infty$, though a value of 8 or more suggests the distribution is approaching a Poisson distribution (Southwood, 1978, 40). The smaller the value of k , the more the clumping. For example, insect populations have often been found to have a value of about 2. As $k \rightarrow 0$ the distribution tends to the logarithmic series distribution. Unfortunately the value of k is often influenced by the size of the sampling unit, so that comparisons can then only be made using the same-sized unit.

Care is need in using contagious models as a good fit of a model does not mean that the model is ecologically useful. It is helpful if any parameters can be given a biological interpretation (e.g., Binns, 1986). For more details concerning the negative binomial distribution, see Hilbe (2011). Some other distributions were described by Seber (1982, section 12.6.1 (2)). Parameter estimation and model selection for Neyman-Scott point processes was considered by Tanaka et al. (2008).

Most ecological data is skewed and is often transformed using, for example, the log or square root transformations. These have the property of reducing the spread of the larger values thus pulling in the tail of the distribution and making it more bell-shaped looking. They can also be used to reduce the dependence of the variance on the mean. However, such transformations can cause some problems and should be used with care. With the development of generalized linear models in which all additive effects and the relationship between mean and variance are specified separately, the need for variance-stabilizing transformations has been made unnecessary.

2.7.3 **CMP Distribution**

Anan et al. (2007) discussed the Conway-Maxwell-Poisson discrete distribution that includes the Bernoulli, the Geometric, and the Poisson distributions as special cases and allows for both overdispersion and underdispersion. It is defined to be

$$f(x) = \frac{\lambda^x}{(x!)^\nu} \frac{1}{z(\lambda, \nu)}, \quad x = 0, 1, \dots; \lambda > 0; \nu \geq 0,$$

where the normalizing constant is

$$z(\lambda, \nu) = \sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^\nu}.$$

When $v = 1$, $z(\lambda, v) = e^\lambda$ and the distribution reduces to the Poisson distribution with parameter λ . When $v = 0$ and $0 < \lambda < 1$, the CMP distribution reduces to the geometric distribution $f(x) = \lambda^x(1 - \lambda)$.

2.7.4 Taylor's Power law

Taylor (1961) found that the empirical “power law”, $v = a\bar{x}^b$, where $v = \sum_{i=1}^s (x_i - \bar{x})^2/(s - 1)$, which relates the sample variance to the sample mean, was satisfied by a number of animal species when an animal population is divided among a number of distinct territorial units. The constant a depends on the density, but b is often found to be independent of density and has been used as a species index. How the model has arisen is hotly debated. The appropriateness of the relationship can be checked by seeing whether the plot of $\log v$ ($= \log a + b \log \bar{x}$) versus $\log \bar{x}$ is reasonably linear. For a recent discussion on this relationship with regard to insect populations, see Taylor (2018, 2021, sampling efficiency). Without going into details, the method has had wide applications, even including human demography (e.g., Bohk et al., 2015: mortality; Oedekoven et al., 2016; and Naccarato and Benassi, 2018).

2.7.5 Fairfield Smith's Variance Law

Smith (1938) found that some plants satisfied the following relation between variances and plot size in appropriate units, namely, $V_x = V_1 x^{-b}$, where V_x is the variance of yield per unit area among plots of size x units, V_1 is the variance among plots of size unity, and b the heterogeneity coefficient. The value of b lies between zero and one. Parameters can be estimated by using

$$\log(V_x) = \log(V_1) - b \log(x),$$

with V_x estimated by the sample variance. Some model variations are given by Narayana Reddy and Ramanatha Chetty (1982), and Lynch (2017) used the method to obtain the optimal sample size and plot size or point sampling factor based on cost-plus-loss.

2.8 Widely Dispersed and Clustered Populations

This type of population can arise when sampling rare or elusive populations such as bats and Mexican spotted owls (Thompson, 2004). Freshwater mussels can be (1) rare, (2) clustered, and (3) elusive as a proportion of the population is buried in

substrate and not easily detected. McDonald (2004), in sampling such a population, discussed definitions of a rare population and noted that all three previous possible characteristics of a population can be involved. From a statistical point of view, a rare population can be defined as one that is hard to detect. McDonald gave some interesting examples of successes and failures of trying to deal with such a population.

One method of handling a sparse population is to use an occupancy model such as described in Chap. 4. Another method is to use double sampling (Sect. 3.2.3), and stratification can also be helpful if the population area is well-known. DNA methods are helpful in dealing with rare populations, and photographic sampling (Karanth et al., 2004) and animal tracks in the snow (Becker et al., 2004) are useful methods in dealing with elusive animals. These methods are discussed in later chapters.

As previously noted, a population may be widely dispersed but highly clustered. In this case if a plot is chosen at random and the plot is comparatively small compared with the spacing between clusters, then there is a high probability that it will be empty or contain few individual if it falls on the edge of a cluster. Clearly random sampling without replacement may not yield many individuals because of the amount of empty space.

An alternative approach that still uses a random sample is to utilize the idea that if a plot is not empty, it may be part of a cluster, so that it would make sense to sample a neighborhood of the plot. An example of a neighborhood is to define it as a cross of plots with the one sampled in the middle. If an object is found in the middle plot, then we sample the other four surrounding plots. If individuals are found in another of these neighborhood plots, then the neighborhood of that plot is also surveyed. This process continues until no more individuals are encountered and we end up with the whole cluster, but with empty plots surrounding the cluster.

This type of cluster sampling, referred to as adaptive sampling, has been extensively developed along with many variations. It was initially introduced in this format by Thompson (1990), and the foundational theory and examples were given by Thompson and Seber (1996). An earlier brief survey of the topic was given by Smith et al. (2004) along with two-phase stratified versions by Manly (2004). Since then there have been a number of extensions and further practical developments described, for example, by Seber and Salehi (2013), and by Yang et al. (2016) who applied the method to trees using circular plots. Prentius and Grafström (2022) also applied adaptive sampling to spaced circular plots with or without a restriction on distance for forming neighborhoods. Clearly adaptive sampling could be applied to a number of methods considered in this book, in addition to plot and transect sampling.

2.9 Summary

The theory of sampling has been well described in the literature, but with less emphasis on density estimation. The topics covered are simple random sampling and associated design implications such as plot shape, proportion of area sampled, and number of plots to be used. Irregular shaped areas can be included in the theory, as well as the use of primary and secondary plots. Stratified sampling and associated design questions are discussed. Side issues considered are edge effects in using plots and the related issue of the home range of animals. Various spatial distributions are described for the number of individuals on a plot, which lead to the problem of sampling widely dispersed but clustered populations where the method of adaptive sampling can be used.

Chapter 3

Detectability



3.1 Detection Probability

When sampling wildlife populations, it is not uncommon to find that an individual in a survey may not be detected or only detected by the sounds it makes such as with bird or anuran surveys. For example, insects on plants, birds in the bush, large mammals in an aerial survey, and mammals in the sea may have varying probabilities of being detected. Amphibian and reptile species can be difficult to detect as many species are nocturnal, have cryptic color patterns, live underground or in murky waters, and may be conspicuous only during certain times of the year or under particular weather conditions (Mazerolle et al., 2007; also a good survey of some detection methods). Along with detectability there is also the problem of extra zero counts, sometimes referred to as “zero inflation,” which can occur because individuals are actually missing due to a variety of population processes or are present but undetected (Dénes et al., 2015). This chapter is closely linked with the next chapter on occupancy.

We now develop some theory, initially with a focus on plot sampling, to introduce the topic. However, the problem of detection arises with other sampling methods discussed in later chapters, where it is developed further. If it is not possible to correct population estimates for under-detection, we may end up with measuring just the relative density in the form of a density index. This will only be useful if similar populations can be compared or general population trends can be followed.

3.1.1 Constant Known Detection Probability

Some earlier background to this topic was given by Thompson and Seber (1994, and 1996: chapter 9). We begin by developing a detection model with known detection

probabilities in preparation for the unknown case with the following theory from Thompson (2012) chapter 16 and particularly his section 16.7, but with some notational changes. The population number of plots S replaces N , the latter we have used for population size, and s replaces n the sample size. Initially, let P_i be the probability of detection in plot i ($i = 1, \dots, S$), given the animal is available for detection. Let y_i denote the number of objects observed in plot i , and let x_i be the actual number. The population total is $N = \sum_{i=1}^S x_i$, the population mean is $\mu_x = N/S$, and the finite population variance is $\sigma_x^2 = \sum_{i=1}^S (x_i - \mu_x)^2/(S-1)$. We now assume that $P_i = P$ for all i , with P known, and a simple random sample of s plots is taken. Then y_i is Binomial(x_i, P) with mean $x_i P$ and variance $x_i P(1-P)$. An estimate of x_i is $\hat{x}_i = y_i/P$ with mean x_i and variance $x_i(1-P)/P$. An estimate of N is then

$$\hat{N}_P = S\bar{y}/P,$$

where \bar{y} is the sample mean. Given the sample σ of s units (and the corresponding unknown x_i), we have from Sect. 1.1 and the usual sampling theory that

$$E[\hat{N}_P | \sigma] = E\left[\frac{S}{sP} \sum_{i=1}^s y_i | \sigma\right] = \frac{S}{Ps} \sum_{i=1}^s Px_i = S\bar{x}.$$

Unconditionally, $E[S\bar{x}] = N$ from sampling theory so that

$$E[\hat{N}_P] = E_\sigma[E(\hat{N}_P | \sigma)] = N,$$

and \hat{N}_P is unbiased for N ($= S\mu_x$).

Variance

To obtain the variance, we first have

$$\text{var}[\hat{N}_P | \sigma] = \frac{S^2}{P^2 s^2} \sum_{i=1}^s \text{var}[y_i] = \frac{S^2}{P^2 s^2} \sum_{i=1}^s x_i P(1-P) = \frac{S^2}{s} \left(\frac{1-P}{P}\right) \bar{x},$$

since, conditional on σ , the y_i are independent. Unconditionally, and from simple random sampling, we have

$$E[\text{var}(\hat{N}_P | \sigma)] = \frac{S^2(1-P)}{Ps} E[\bar{x}] = \frac{S^2(1-P)}{Ps} \mu_x,$$

and

$$\text{var}[E(\hat{N}_P | \sigma)] = \text{var}[S\bar{x}] = S^2 \left(\frac{S-s}{Ss}\right) \sigma_x^2.$$

Finally, using a standard result about conditional means and variances (Appendix A.6),

$$\begin{aligned}\text{var}[\widehat{N}_P] &= \text{var}[\mathbb{E}(\widehat{N}_P \mid \sigma)] + \mathbb{E}[\text{var}(\widehat{N}_P \mid \sigma)] \\ &= S^2 \left[\left(\frac{S-s}{S} \right) \frac{\sigma_x^2}{s} + \left(\frac{1-P}{P} \right) \frac{\mu_x}{s} \right].\end{aligned}$$

Here the first term relates to the sampling process, while the second term relates to the detection process. In the latter case, the variance goes down as P goes up. Note that $\text{var}[\bar{y}] = (P^2/S^2)\text{var}[\widehat{N}_P]$.

Variance Estimation

To obtain an estimate of the variance, we first let $\mathbf{y} = (y_1, \dots, y_N)'$ with population mean and variance, *conditional* on \mathbf{y} , of μ_y and σ_y^2 , respectively, where

$$\sigma_y^2 = \sum_{i=1}^S (y_i - \mu_y)^2 / (S-1).$$

Then, with simple random sampling,

$$\mathbb{E}[\bar{y} \mid \mathbf{y}] = \mu_y,$$

and

$$\text{var}[\bar{y} \mid \mathbf{y}] = \frac{S-s}{Ss} \sigma_y^2.$$

Also, with $v_y = \sum_{i=1}^n (y_i - \mu_y)^2 / (s-1)$,

$$\text{var}[\bar{y} \mid \mathbf{y}] = \mathbb{E} \left[\left(\frac{S-s}{Ss} \right) v_y \mid \mathbf{y} \right].$$

Since $\mathbb{E}[y_i] = Px_i$ we have, unconditionally

$$\mathbb{E}[\bar{y}] = \mathbb{E}[\mu_y] = P\mu_x,$$

and since the y_i have independent binomial distributions,

$$\text{var}[\mathbb{E}(\bar{y})] = \text{var} \left[\frac{1}{S} \sum_{i=1}^S y_i \right] = \frac{P(1-P)}{S^2} \sum_{i=1}^S x_i = \frac{P(1-P)\mu_x}{S}.$$

This leads to the following two ingredients for our unconditional variance,

$$\mathbb{E}\left[\frac{(1-P)\bar{y}}{S}\right] = \text{var}[\mathbb{E}(\bar{y} | \mathbf{y})]$$

and

$$\mathbb{E}\left[\left(\frac{S-s}{Ss}\right)v_y\right] = \mathbb{E}[\text{var}(\bar{y} | \mathbf{y})].$$

Since

$$\text{var}[\bar{y}] = \text{var}[\mathbb{E}(\bar{y} | \mathbf{y})] + \mathbb{E}[\text{var}(\bar{y} | \mathbf{y})],$$

an unbiased estimate of $\text{var}[\bar{y}]$ is

$$\widehat{\text{var}}[\bar{y}] = \left(\frac{S-s}{Ss}\right)v_y + \frac{(1-P)\bar{y}}{S}.$$

Finally, since $\widehat{N}_P = S\bar{y}/P$, an unbiased estimator of $\text{var}[\widehat{N}_P]$ is

$$\widehat{\text{var}}[\widehat{N}_P] = (S^2/P^2)\widehat{\text{var}}[\bar{y}],$$

an expression given by Thompson (2012, 220). The problem remains of how to estimate P , the subject of the next section. A number of methods are considered in the next chapter. If we have an estimate \widehat{P} and v_p is an estimate of its variance, then we have the estimate $\widehat{N} = S\bar{y}/\widehat{P}$ and, from Thompson (2012, 221), a variance estimate is

$$\widehat{\text{var}}[\widehat{N}] = \frac{S^2}{\widehat{P}^2} \left[\left(\frac{S-s}{S}\right) \frac{v_y}{s} + \left(\frac{1-\widehat{P}}{S}\right) \bar{y} + \frac{\bar{y}^2}{\widehat{P}^2} v_P \right].$$

We note that we can write $P = \psi p$, where p is the probability of observing an individual given it is present and ψ is the probability of being present.

3.2 Unknown Detection

Unknown detectability that is not accounted for can have an effect on producing any overdispersion or underdispersion. To see a broad-brush effect of this, we follow Etterson et al. (2009), who assume $N \sim \text{Poisson}(\lambda)$ and that P , the proportion detected, is a random variable. If $C = NP$ is the total number seen and N and P

are independent random variables, then $E[C] = E[P]\lambda$. From Goodman (1960)

$$\begin{aligned}\text{var}[C] &= \text{var}[P](E[N])^2 + (E[P])^2\text{var}[N] + \text{var}[P]\text{var}[N] \\ &= \text{var}[P]\lambda(1 + \lambda) + \lambda(E[P])^2.\end{aligned}$$

In general, this will not be equal to $E[C]$. For example, if P is constant, $\text{var}[C] = \lambda P^2 < \lambda P = E[C]$, and counts will be underdispersed. More generally, counts will tend to be overdispersed when the righthand side of the above equation is greater than $E[C] = \lambda E[P]$; that is when

$$\text{var}[P] > \frac{E[P](1 - E[P])}{(1 + \lambda)}.$$

Hence this simplistic approach indicates that heterogeneity is more likely to cause overdispersion in counts of abundant species (large λ) with either high or low mean detection probability.

If $N_i \sim \text{Poisson}(\lambda_i)$, where λ_i is now a random variable, then

$$E[N_i] = E[\lambda_i],$$

and

$$\begin{aligned}\text{var}[N_i] &= E[\text{var}(N_i | \lambda_i)] + \text{var}[E(N_i | \lambda_i)] \\ &= E[\lambda_i] + \text{var}[\lambda_i].\end{aligned}$$

This means that Poisson mixtures are overdispersed by the amount $\text{var}[\lambda_i]$ so that environmental heterogeneity will always result in overdispersion if subpopulations are conditionally Poisson distributed and the environmental effects are not incorporated into covariate analysis. Clearly more data information is required, which is accomplished by obtaining replicate counts. We now consider this briefly.

3.2.1 Replicate Counts

When observations are carried out at a number of sites, there is the problem of how to estimate the detection probabilities. One way around this is to have replicate surveys at each site. Royle (2004a) used a metapopulation approach for this situation, assuming a constant probability of detection, and assuming the population sizes of the sites followed prior distributions such as the $\text{Poisson}(\lambda_i)$ distribution for site i . Here $\log(\lambda_i)$ can be expressed as a linear model in covariates.

Various methods for handling such data, discussed in later chapters, were given by Royle (2004b). In particular, he mentioned the so-called removal method where once an individual is first observed in a particular replicate, it is considered as being

“removed,” i.e., future observations of it are ignored. This method is also referred to as the “time to detection” or “time to removal” method. It was discussed extensively by Sólymos et al. (2018) who noted that removal model estimates of availability may be improved by accounting for variation in singing rates using covariates for time of day and day of year. Behavioral heterogeneity in singing can be incorporated by subdividing the sampled population of a species at a given point into a finite mixture of birds with low and high singing rates, which requires the additional estimation of the proportion of birds in the sampled population with low singing rates.

They also noted that several studies have recommended finite mixture models over conventional removal models to reduce bias in abundance estimates. However, large sample sizes are needed to fit the more complex models. The authors considered sample size requirements for time-invariant removal models, as well as adjusting for a number of variations such as variation in count duration, which can have a substantial effect on average counts. They also evaluated alternative formulations of the time-removal model (a conventional removal model and a finite-mixture removal model), with and without covariates. The study demonstrated the importance of accounting for temporal heterogeneity in bird availability when analyzing avian point data, especially when combining data from diverse studies.

Spatial replication with repeat visits is considered in Chap. 4 to overcome some estimation problems. However, there can be problems with multiple surveys such as the analysis of historical datasets where replicate surveys are unavailable; the significant amounts of time, energy, and money involved in several surveys; lack of independence of replicate surveys; and the closure assumption can be violated, even on very short-time scales because of within territory movement and the time taken to carry out the repeated surveys. Many historical data sets exist that were collected using only a single visit, and this combined with logistical/cost considerations has prevented many current research programs from collecting multiple visit data.

The previous problems have led to an interest in single visit surveys, which has led to some controversy. Lele et al. (2012) and Sólymos et al. (2012) showed that provided the covariates for the probability of occupancy and the probability of detection are not exactly the same, the conditional likelihood (Appendix 3.2) can be used to estimate the regression parameters of a binomial-zero-inflated Poisson (ZIP) mixture model and correct for detection error. However, Knape and Korner-Nievergelt (2016) argued that there is essentially no information about absolute abundance contained in single-visit count data and additional more direct data on detection probabilities is required to robustly estimate absolute abundances.

Resource Selection

A related topic is that of resource selection. This was described in detail by Manly et al. (2002) and is used when resources are used disproportionately to their availability. Understanding the differential selection of resources by animals is an essential component of conservation biology, wildlife management, and applied ecology. Lele et al. (2013) provided some clarification with regard to this topic and underlying concepts. Common tools used for gathering such understanding are the

resource selection probability function (RSPF) and the resource selection function (RSF).

The RSPF is a function that gives the probability that a particular resource, as characterized by a combination of environmental variables, will be used by an individual animal. Sólymos and Lele (2016) linked this topic in with the single-visit (SV) approach method described in the previous paragraph and clarified the conditions under which the SV methodology is valid. They argued that the situations under which the RSPF and SV methodologies fail are rare. Also, they gave a distance sampling extension of the binomial-ZIP model and extended the conditional likelihood method to the multinomial-ZIP model. Various candidates have been put forward for the RSPF (Lele & Keim, 2006; Lele, 2009).

3.2.2 Circular Plots

A common situation is when we have circular plots. These arise naturally when a point is chosen at random at a site (e.g., on a quadrat) or on a randomly placed line transect and the number of individuals detected in a circular plot centered on the point and out to a radial distance r are counted, called “point counts.” In order to estimate detection probabilities, repeated counts at specific time intervals are made on each circular plot. Royle and Nichols (2003) discussed a number of methods using such data (see also Haines, 2016). This approach with replicate observations is used in conjunction with distance measurements in Chap. 7. Two protocols that need to be decided are the length of the time interval and the radius length, both of which can affect the outcome. Also the choice of any covariates incorporated in the model is important.

Simons et al. (2009) have a helpful overview of using point counts, with an emphasis on avian counts, though it can apply to other species as well. They give a range of factors that can affect detection probabilities on auditory counts. We mention three, namely, ambient noise, misclassification errors due to the difficulty of accurately identifying and localizing sounds during a count, and measurement error due to the difficulty of estimating the distance to a sound source when visual cues are lacking.

McClintock et al. (2010) also discussed observation error and commented that false-positive errors (wrongly recording an individual present when not, as with a species incorrectly identified) and detection probability heterogeneity may be much more prevalent than expected. Even low levels of false-positive errors, constituting as little as 1% of all detections, can cause severe overestimation of site occupancy, colonization, and local extinction probabilities. Methods of dealing with such errors are discussed in Chap. 4.

Kansanen et al. (2021) used circular plot sampling for trees. If the sample is collected using a technical device located at the sampling point, such as a terrestrial laser scanner (TLS), all the trees of the sample plot cannot be observed because they can hide behind each other. For example, the tree stems closer to the scanner

produce nonvisible areas behind them so that trees further away from the scanner located in the nonvisible areas cannot be seen. Also, undergrowth, low branches, or other objects can block the submitted laser pulse.

TLS is a method of close-range remote sensing where a LiDAR scanner is operated on ground level to produce a three-dimensional point cloud of the surroundings. Typically, the scanner remains stationary on a tripod and rotates 360 degrees in the horizontal plane, while the laser scans in the vertical direction in some device-dependent opening angle. The authors proposed a Horvitz-Thompson-like estimator with distance-based detection probabilities derived from stochastic geometry for the estimation of population totals such as stem density and basal area in such situation, as follows.

Consider the forest as a realization of a marked point process $M = \{(x_i, d_i, m_i)\}$, where x_i are the locations of the stem centers generated by some spatial point process; d_i are the diameters at breast height (DBH) of the trees, governed by some distribution; and m_i is a mark of interest related to tree i . For example, if we are interested in a stem density estimate $m_i = 1$, while if we interested total basal area, then m_i is the basal area of tree i . Other marks, such as volume or biomass, are also possible. With these detection probabilities, a Horvitz-Thompson-like estimator can be obtained for a population total of interest. Tree stems are modeled as discs centered at x_i with radius d_i , and the tress are ordered based on their distances from the origin $r_i - d_i/2$, where r_i is the distance of the stem center from the origin. In other words, the tress are ordered based on the shortest distance to their outer bark. Thus tree $i = 1$ is closest to the origin and $i = 2$ is the second closest, and so on. If α is a tuning parameter to be determined before estimation, the probability of detection takes the form $p_\alpha(r_i, d_i)$ or p_i for short. With these detection probabilities, a Horvitz-Thompson-like estimator is available for a population total of interest. For example, if $\tau = \sum_{i=1}^N m_i$, we have

$$\hat{\tau} = \sum_{i=1}^N \frac{m_i z_i}{\hat{p}_\alpha(r_i, d_i)},$$

where $z_i = 1$ if tree i is detected and 0 otherwise. The above estimator was shown to be unbiased for a homogeneous Poisson process along with an unbiased variance estimator

$$\widehat{\text{var}}[\hat{\tau}] = \sum_{i=1}^n \left(\frac{1}{\hat{p}_i^2} - \frac{1}{\hat{p}_i} \right) m_i^2.$$

The authors found that the above method compared favorably with other methods

Johnson (2008) listed two issues with bird surveys, which will also apply to other species. Firstly, there is heterogeneity in detectability among individual birds. Capture-recapture and mixture models can also be used here. Secondly, care is needed in choosing the length of count period. Longer count periods increase the numbers of birds seen, especially birds with low availability or detectability.

Conversely, longer periods may lead to biased estimates because birds not initially in the count area may enter during the count period. The longer the count period and the greater the speed of birds, the greater is the potential for a biased estimate of density. Two protocols then that are needed to be decided are the length of the time interval and the radius length, both of which can affect the outcome. Also the choice of any covariates incorporated in the model is important.

3.2.3 Double Sampling

One general method that can be used to estimate unknown detectability is to use double sampling. This refers to designs in which a sample of units is first selected for obtaining basic or auxiliary information only (a “rapid” method), and then a second sample is selected in which the variable of interest is observed in addition to the auxiliary information. In the case of detectability, the variable of interest is the “accurate” measurement x_i (the true count on plot i), while the less accurate but easier to obtain y_i (the observed count on plot i) serves as an auxiliary variable. The first sample is used to obtain raw count data and the second sample used to get detectability information and thus estimate detectability. For example, in aerial censusing, extensive counts on transects can be made from the air, where there may be visibility bias. Another sample or subsample of the transects can then be used for ground counts, where it is assumed that there is complete detectability (cf. Sect. 3.1). The method is now considered as follows.

In addition to the initial sample of s plots, we consider two cases. For case 1 we assume that an independent sample of s_1 plots is taken for which both x_i and y_i are known for each plot through careful searching. For case 2 the s_1 plots are a subsample of the s plots.

Case 1: Independent Second Sample

If the probability of detection in the i th plot is P_i , then $E[y_i] = P_i x_i$, and a moment estimate of P_i is y_i/x_i . If we set $P_i = P$, we have several estimates of P which can be combined for the s_1 plots to give the ratio estimate

$$r_P = \frac{\sum_{i=1}^{s_1} y_i}{\sum_{i=1}^{s_1} x_i} = \frac{\bar{y}}{\bar{x}}. \quad (3.1)$$

Conditional on the x_i and the sample, this estimate is unbiased.

Another method is to use a regression approach with the model $y_i = \beta x_i + \epsilon_i$, ($i = 1, \dots, s_1$), where $\beta = P$, $E[\epsilon_i] = 0$, and $\text{var}[\epsilon_i] = P(1 - P)x_i$, and obtain a (weighted) least squares estimate of β with weights $w_i = x_i^{-1}$. This leads to the estimator $\hat{\beta} = \sum_{i=1}^{s_1} w_i y_i x_i / \sum_{i=1}^{s_1} w_i x_i^2$, which is the same as as r_P . An unbiased estimate of the variance is then $\sum_{i=1}^{s_1} (y_i - r_P x_i)^2 / (s_1 - 1)$.

At this stage let \hat{P} be some estimator of P that is approximately unbiased and uncorrelated with \bar{y} , and then an estimator of N is $\hat{N} = S\bar{y}/\hat{P}$, which is given in

Sect. 3.1.1 above, but with P estimated. We now use a well known result which follows by using the so-called delta method (Seber & Schofield, 2019, 274). If $Y = X_1/X_2$, then $\mu_Y \approx \mu_1/\mu_2$, where $\mu_i = E[X_i]$ and

$$\text{var}\left[\frac{X_1}{X_2}\right] \approx \frac{\mu_1^2}{\mu_2^2} \left(\frac{\text{var}[X_1]}{\mu_1^2} + \frac{\text{var}[X_2]}{\mu_2^2} - 2 \frac{\text{cov}[X_1, X_2]}{\mu_1 \mu_2} \right).$$

Also

$$\text{var}[X_1 X_2] \approx \mu_2^2 \text{var}[X_1] + \mu_1^2 \text{var}[X_2] + 2\mu_1 \mu_2 \text{cov}[X_1, X_2],$$

which we use later.

Then setting $X_1 = S\bar{y}$ and $X_2 = \hat{P}$ so that $\hat{N} = X_1/X_2$ we find that, with zero covariance (Thompson, 2012, 221),

$$\text{var}[\hat{N}] \approx S^2 \left[\left(\frac{S-s}{S} \right) \frac{\sigma_x^2}{s} + \left(\frac{1-P}{P} \right) \frac{\mu_x}{s} + \frac{\mu_x^2}{P^2} \text{var}[\hat{P}] \right],$$

with an estimate

$$\widehat{\text{var}}[\hat{N}] = \frac{S^2}{\hat{P}^2} \left[\left(\frac{S-s}{S} \right) \frac{v_y}{s} + \left(\frac{1-\hat{P}}{S} \right) \bar{y} + \frac{\bar{y}^2}{\hat{P}^2} \widehat{\text{var}}[\hat{P}] \right].$$

In some applications, it is not P that is estimated but its reciprocal, namely, $\alpha = 1/P$. If $\hat{\alpha}$ is approximately unbiased and independent of \bar{y} , we have from Thompson (2012, 221),

$$\text{var}[\hat{N}] \approx S^2 \left[\left(\frac{S-s}{S} \right) \frac{\sigma_x^2}{s} + \frac{(\alpha-1)\mu_x}{s} + \frac{\mu_x^2}{\alpha^2} \text{var}[\hat{\alpha}] \right],$$

with estimate

$$\widehat{\text{var}}[\hat{N}] = S^2 \left[\frac{\hat{\alpha}^2(S-s)}{S} \frac{v_y}{s} + \frac{\hat{\alpha}(\hat{\alpha}-1)\bar{y}}{S} + \frac{\bar{y}^2}{S} \widehat{\text{var}}[\hat{\alpha}] \right].$$

Case 2: Second Sample is a Subsample

Bart and Earnst (2002) applied the double sampling method to estimating the mean number of shorebirds per plot on the North Slope of Alaska (with x and y interchanged and n_1 plots a subsample from the n plots sampled, so that the latter can be treated as a population of size n). Let y'_i be the number observed in plot i of the n plots, and let \bar{y}' be their mean number observed per plot, where r_P is given by (3.1) in Sect. 3.2.3. Define $g_i = x_i - y'_i/r_P$ ($i = 1, \dots, s_1$). The authors give the

estimate of d , the average number per unit plot as

$$\hat{d} = \bar{y}' / r_P$$

with a variance estimate of

$$\widehat{\text{var}}[\hat{d}] = \frac{v(x_i)}{n} + \left(\frac{1}{n_1} - \frac{1}{n} \right) v(g_i),$$

where

$$v(x_i) = \sum_{i=1}^{s_1} (x_i - \bar{x})^2 / (s_1 - 1) \quad \text{and} \quad v(g_i) = \sum_{i=1}^{n_1} (g_i - \bar{g})^2 / (s_1 - 1).$$

Also, since r_P is also of interest, a variance estimate of it is

$$\widehat{\text{var}}[r_P] = v(h_i) / (s_1 \bar{x}^2),$$

where $h_i = y_i - r_P x_i$ and

$$v(h_i) = \sum_{i=1}^{s_1} (h_i - \bar{h})^2 / (s_1 - 1).$$

When several species are counted, the authors considered the case where the number of a particular species in the subsample is zero so that \bar{x} is zero. They also considered an important question of how to allocate resources between the rapid sample of size n and the more intensive subsample of size n_1 . If c is the cost per plot of the rapid sample and c_1 for the subsample, then the total cost is $C = c_1 n_1 + cn$. Bart and Earnst (2002) stated that the standard error of \hat{d} is minimized when $c_1/c = \alpha$, where

$$\alpha^2 = \frac{n}{n_1} \left(\frac{v(g_i)}{v(x_i) - v(g_i)} \right).$$

Then, for a given C ,

$$n = \frac{C}{c + c_1 \alpha} \quad \text{and} \quad n_1 = \frac{C - cn}{c_1}.$$

If estimates of the above terms are available from a pilot study, then the effort can be appropriately allocated. Advantages with using double sampling were also discussed. Farnsworth et al. (2005) felt that the method is extremely effort intensive and hence is not a comparable alternative method for counting birds in most habitats in which point counts are conducted.

3.2.4 Partial Repeat Sampling

Loehle (2010) introduced a method of capture-recapture in which a subset of the area is sampled twice for counts to get an estimate of detectability and then used it to correct a larger population where there are just single counts. In the subset, the individuals are marked in the first survey. This area is surveyed again quickly enough so that birth, death, immigration, and emigration can be ignored. Let m_{10} being the number of individuals observed during the first survey but not the second, m_{01} the opposite, m_{11} the number observed during both surveys, and $r = m_{11} + m_{01} + m_{10}$. Let p_i be the probability of being detected in sample i ($i = 1, 2$), then $P_2 = p_1 + p_2 - p_1 p_2$ is the probability that an individual is seen at least once, i.e., the detectability over the two samples. Using the Petersen method and estimate (Sect. 12.2), and the fact that the maximum likelihood estimates are also the moment estimates, we get

$$\widehat{P}_2 = \frac{r m_{11}}{(m_{11} + m_{10})(m_{11} + m_{01})},$$

and an estimate of the subarea population size based on the two samples is

$$\widehat{N}_2 = r/\widehat{P}_2.$$

For the area sampled once, the individual probability of detection P_1 is estimated by

$$\widehat{P}_1 = m_{11}/(m_{11} + m_{10}) \quad \text{and} \quad \widehat{N}_1 = y/\widehat{P}_1,$$

where y is the number of counts for the larger population of size N_1 . Advantages of the method over the usual two sample capture-recapture experiment are as follows: less overall sampling is required, more spatial variation in density can be captured, and there is higher detectability.

3.2.5 Unequal Individual Detection Probabilities

We consider a special case of Thompson (2012, 224). Let x_{ij} be an observation associated with the j th object in the i th plot (the ij th object), and let P_{ij} be its probability of detection. We begin by assuming that the latter is known. If instead we are considering just groups, then x_{ij} is the number in the ij th group. As we are primarily interested in just abundance, we can set the $x_{ij} = 1$ and let x_i be the number of objects in the i th plot, of which y_i are detected. Define the indicator variable z_{ij} for the ij th individual, taking the value 1 with probability P_{ij} if detected, and 0 with probability $1 - P_{ij}$ if not detected. We can assume that the z_{ij} are independent Bernoulli random variables so $E[z_{ij}] = P_{ij}$ and $\text{var}[z_{ij}] = P_{ij}(1-P_{ij})$. Let \mathbf{z} be the vector representing the set of z_{ij} . We note that $y_i = \sum_{j=1}^{x_i} z_{ij}$ is the number observed on plot i . Let π_i be the probability that plot i is included in the s

sample plots and let $\pi_{ii'}$ be the probability that both plot i and plot i' are included. Let I_i be an indicator variable that takes the value 1 if plot i is included and 0 otherwise. It is assumed that I_i and the z_{ij} are independent, which will generally be the case. Also I has mean and variance π_i and $\pi_i(1 - \pi_i)$ respectively, and $E[I_i z_{ij}] = \pi_i P_{ij}$.

Consider the estimator

$$\widehat{N} = \sum_{i=1}^S \frac{I_i}{\pi_i} \sum_{j=1}^{x_i} \frac{z_{ij}}{P_{ij}}.$$

Then

$$E[\widehat{N} | \mathbf{z}] = \sum_{i=1}^S \sum_{j=1}^{x_i} \frac{z_{ij}}{P_{ij}},$$

and

$$E[\widehat{N}] = \sum_{i=1}^S \sum_{j=1}^{x_i} 1 = \sum_{i=1}^S x_i = N.$$

Conditional on \mathbf{z} , \widehat{N} is a Horvitz-Thompson estimator of $\sum_{i=1}^S \sum_{j=1}^{x_i} (z_{ij}/P_{ij})$ with conditional unbiased variance estimator of $\text{var}[\widehat{N} | \mathbf{z}]$, namely,

$$v_1 = \sum_{i=1}^S \left(\frac{1 - \pi_i}{\pi_i^2} \right) \widehat{x}_i^2 + \sum_{i=1}^S \sum_{i' \neq i} \left(\frac{\pi_{ii'} - \pi_i \pi_{i'}'}{\pi_{ii'} \pi_i \pi_{i'}} \right) \widehat{x}_i \widehat{x}_{i'},$$

where $\widehat{x}_i = \sum_{j=1}^{y_i} 1/P_{ij}$. Hence

$$E[v_1 | \mathbf{z}] = \text{var}[\widehat{N} | \mathbf{z}] \quad \text{and} \quad E[v_1] = E[\text{var}(\widehat{N}) | \mathbf{z}].$$

We now need to find an unbiased estimator of $\text{var}\left[E(\widehat{N} | \mathbf{z})\right]$, where

$$\begin{aligned} \text{var}[E(\widehat{N} | \mathbf{z})] &= \text{var}\left[\sum_{i=1}^S \sum_{j=1}^{x_i} \frac{z_{ij}}{P_{ij}}\right] \\ &= \sum_{i=1}^S \sum_{j=1}^{x_i} \frac{1}{P_{ij}^2} \text{var}[z_{ij}] \\ &= \sum_{i=1}^S \sum_{j=1}^{x_i} \frac{1 - P_{ij}}{P_{ij}}. \end{aligned}$$

This is the population total of $w_{ij} = (1 - P_{ij})/P_{ij}$ so that by the Horvitz-Thompson method, an unbiased estimate of this total is

$$\begin{aligned} v_2 &= \sum_{i=1}^S I_i \frac{1}{\pi_i} \sum_{j=1}^{x_i} \left(\frac{w_{ij} z_{ij}}{P_{ij}} \right) \\ &= \sum_{i=1}^s \frac{1}{\pi_i} \sum_{j=1}^{y_i} \left(\frac{1 - P_{ij}}{P_{ij}^2} \right). \end{aligned}$$

Finally $v_1 + v_2$ is an unbiased estimate of

$$\text{var}[\widehat{N}] = E[\text{var}(\widehat{N} \mid \mathbf{z})] + \text{var}[E(\widehat{N}) \mid \mathbf{z}].$$

We have deliberately not specified the design used, and just assumed the existence of a design unbiased estimator of N . The above theory can therefore be applied generally to a number of conventional designs as well as adaptive sampling referred to briefly in Sect. 2.8. In the special case of simple random sampling, we have $\pi_i = s/S$ and $\pi_{ii'} = s(s-1)/S(S-1)$.

The next problem is how to estimate the P_{ij} . Clearly this cannot be done without restrictions on the P_{ij} as we have too many unknowns. Two possible restrictions are $P_{ij} = P$ and $P_{ij} = P_i$. We have already discussed the first case in detail above, where each plot of a separate sample of size s_1 plots is thoroughly searched. This means that the x_i , the number in plot i , are known for each of those plots. In the case of $P_{ij} = P_i$ we have the estimate $\widehat{P}_i = y_i/x_i$ (with mean P_i and variance $P_i(1-P_i)$) from the sample of size s_1 . Unfortunately this tells us nothing about the P_i for the members of the s sample unless we can link up the two sets of P_i using for example a regression method.

Taking a simple approach, if we have a suitable covariate u_i , we can try setting $P_i = \alpha + \beta u_i$. This leads to the model $\widehat{P}_i = \alpha + \beta u_i + \epsilon_i$ ($i = 1, \dots, s_1$), which can be looked at graphically. If the model seems a reasonable fit, we can then obtain regression estimates $\widehat{\alpha}$ and $\widehat{\beta}$ along with variances and variance estimates. For the s sample we can now obtain the predictions $\widehat{P}_i = \widehat{\alpha} + \widehat{\beta} u_i$, $i = 1, \dots, s$. Because P_i lies in $(0, 1)$ it is possible to get out of range estimates, which is usually overcome by using a link function, typically the logistic function where we set $\log(P_i/(1-P_i)) = \alpha + \beta u_i$ ($i = 1, \dots, s$). Using estimates of α and β from the s_1 samples, we can get predicted values of $\log(P_i/(1-P_i))$ by using the u_i values for the s sample. Then, using these predicted \widehat{P}_i , we can consider the following estimator

$$\widehat{N} = \sum_{i=1}^S \frac{I_i}{\pi_i} \frac{y_i}{\widehat{P}_i} = \sum_{i=1}^s \frac{1}{\pi_i} \frac{y_i}{\widehat{P}_i}.$$

For simple random sampling $\pi_i = s/S$. If $E[1/\widehat{P}_i] \approx 1/P_i$, and noting that π_i , y_i , and \widehat{P}_i are mutually independent because of different samples,

$$E[\widehat{N}] \approx \sum_{i=1}^s \frac{S}{s} \frac{E[x_i] P_i}{P_i} = SE[\bar{x}] = N.$$

Then

$$\text{var}[\widehat{N}] = \frac{S^2}{s^2} \sum_{i=1}^s \text{var}\left[\frac{y_i}{\widehat{P}_i}\right],$$

which involves individual variances of the ratios of two independent variables.

Clearly further research is needed to find a suitable variance estimate of \widehat{N} as well as an extension to the case with more than one covariate. It might be possible to extend this approach to the case of different P_{ij} using the variable z_{ij} and a regression model with covariates. If an approximately unbiased estimate of \widehat{P}_{ij} can be found for the sample of size s_1 , then we can consider the estimate

$$N^* = \sum_{i=1}^s \frac{1}{\pi_i} \sum_{j=1}^{x_i} \frac{z_{ij}}{\widehat{P}_{ij}} = \sum_{i=1}^s \frac{S}{s} \sum_{i=1}^{y_i} \frac{1}{P_{ij}}.$$

Since $E[z_{ij}] = P_{ij}$ we find that N^* is an approximately unbiased estimate of N .

3.2.6 Detection Model with Spatially Replicated Counts

We now consider a model developed by Royle (2004a) that uses spatial and temporal data. Let n_{it} be the number of distinct individuals counted on plot (site) i at time t ($t = 1, \dots, T$). It is assumed that $n_{it} \sim \text{Binomial}(N_i, p)$, where N_i is the population size on site i at time t ($t = 1, \dots, T$, and p is the detection (or capture) probability. The likelihood is then

$$L(N_i, p | n_{i1}, \dots, n_{iT}) = \prod_{t=1}^T \binom{N_i}{n_{it}} p^{n_{it}} (1-p)^{N_i - n_{it}}. \quad (3.2)$$

The joint likelihood for the s sites is now

$$L(\{N_i\}, p | \{n_{it}\}) = \prod_{i=1}^s \left[\prod_{t=1}^T \binom{N_i}{n_{it}} p^{n_{it}} (1-p)^{N_i - n_{it}} \right]. \quad (3.3)$$

Royle (2004a) noted that there are some problems in obtaining estimates for the model such as numerical instability, often sparse data, and s large with some N_i possibly zero. One approach is to use a random effects model and view the N_i as being independent random variables from a density function $f(N; \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is some (possibly vector valued) parameter, and integrate (3.3) over this prior distribution. The integrated likelihood is then

$$L(p, \boldsymbol{\theta} | \{n_{it}\}) = \prod_{i=1}^s \left\{ \sum_{N_i=\max_t n_{it}} \left(\prod_{t=1}^T \text{Binomial}(n_{it}; N_i, p) \right) f(N_i; \boldsymbol{\theta}) \right\}. \quad (3.4)$$

This expression can be maximized numerically with the summation restricted to a finite but large bound K , say. This can be done by successively increasing K until there is little change in the estimates. For example, one can use the Poisson distribution

$$f(N; \lambda) = e^{-\lambda} \lambda^N / N!,$$

with $\boldsymbol{\theta} = \lambda$, the average density per site unit. In this case an estimate of the total abundance for the *sampled* area is $\widehat{N} = s\widehat{\lambda}$. Covariates can also be introduced such as

$$\log(\lambda_i) = \mathbf{x}'_i \boldsymbol{\beta},$$

with covariate values \mathbf{x}_i .

Alternatively one can use a negative-binomial distribution which results from a Gamma prior distribution on site-specific intensity parameters λ_i , namely

$$f(N; \alpha, r) = \frac{\Gamma(N + \alpha)}{\Gamma(\alpha)N!} r^\alpha (1 - r)^N.$$

This is frequently parameterized in terms of the mean, $\mu = \alpha(1 - r)/r$, or the Gamma parameters (α, β) with $r = \beta/(1 + \beta)$.

An estimate of $\boldsymbol{\theta}$ can be used in conjunction with Bayes theorem to yield an estimate of the conditional (on $\boldsymbol{\theta}$) posterior distribution of any particular N_i , namely

$$\Pr(N = k | n_1, n_2, \dots, n_T, \boldsymbol{\theta}, p) = \frac{\Pr(\{n_i\} | N = k, p) \Pr(N = k; \boldsymbol{\theta})}{\sum_{k=0}^{\infty} \Pr(\{n_i\} | N = k, p) \Pr(N = k; \boldsymbol{\theta})},$$

using estimates $\widehat{\boldsymbol{\theta}}$ and \widehat{p} in place of $\boldsymbol{\theta}$ and p . Thus given the counts n_{i1}, \dots, n_{iT} we have a posterior distribution of N_i , which can be used to make inferences about particular values of N_i .

Carroll and Lombard (1985) used an alternative approach to the above theory by integrating (3.2) over a prior distribution for p , say a Beta distribution with fixed

parameters, though Royle (2004a) noted a few problems with this method. He also gave an example of a large-scale study involving 6 species of bird using repeated point counts on 4000 routes, where each route had 50 sample sites located one-half mile apart. Dail and Madsen (2011) extended Royle's model to open populations using a generalized mixture model and then obtained a test of closure for Royle's model. An asymptotic version of this model that alleviated some computational difficulties was given by Brintz et al. (2018).

3.2.7 *Detection Methods*

Thompson (2002) and McCallum (2005) summarized some of the population methods involved with estimating detection probabilities; these are discussed in appropriate sections later in the book. For instance, looking ahead we have the following examples:

- Double sampling. This has already been used above.
- Variable circular plots.
- Distance sampling with points and lines.
- Removal sampling.
- Capture-recapture methods.
- Double observer methods.

McCallum (2005) mentions four features that the probability of detection depends on: (1) count duration (the more time the more individuals observed); (2) the probability has two independent components, namely, the availability of cues and the detectability of cues; (3) detectability is a function of both the conspicuousness of cues (e.g., vocalizations, conspicuous movements) and the abundance of cues; and (4) discontinuous production of cues lowers availability. He mentions that double observer methods and distance sampling are best suited to estimating detectability, while double and removal sampling are better suited to estimating availability. Software for using these methods is available such as **R** (Mazerolle, 2015).

As already inferred, the probability of detection can incorporate two components, and in fact three components can be used. For example, if P is the probability of detection, we can have

$$P = P_p P_{a|p} P_{d|a,p}.$$

Here P_p is the probability of being present for the sample, $P_{a|p}$ is the probability of being available for sampling, given present, and $P_{d|a,p}$, is the probability of being detected given present and available. Frequently $P_{a|p} = 1$ and $P_{d|a,p} = P_{d|p}$, which is the case in much of the future chapters, where the availability component is not mentioned, or else the presence and availability are combined. In addition to plot sampling, this idea arises in Chap. 4 where occupancy models are considered (e.g.,

Sect. 4.8). As Chap. 4 is already a large chapter, we defer a fuller discussion of components of detectability until Chap. 7, Sect. 7.8.

3.3 Adaptive Sampling with Incomplete Detectability

Smith et al. (2010) considered the effect of imperfect detectability on adaptive sampling in studying freshwater mussels as they are often rare, spatially clustered, and imperfectly detected. Adaptive sampling, alluded to in Sect. 2.8, is essentially a method of cluster (network) sampling, and the way the network grows depends on individuals present in a neighboring plot being added if the plot overlaps the network. If the individuals are present but not detected, the plot is not added to the network. This is a problem because the unbiased estimation of abundance using adaptive sampling relies on accurate intersection probabilities, which are typically not obtained when detection is imperfect. Using simulation it was found that the efficiency of adaptive sampling, relative to that of conventional sampling under perfect detectability, decreased with detectability. These decreases largely arose from increases in bias induced by detection error. The bias was similar for conventional and adaptive designs, even though inclusion probabilities in adaptive cluster sampling were inaccurate due to detectability. Relative to conventional designs, adaptive sampling designs outperformed conventional designs on efficiency criterion only at the lowest density and only when detection was imperfect. Further adaptive methods are described for line transects in Sect. 8.13.6.

3.4 Conclusion

In much of the early research, little attention was paid to the possibility of incomplete detection, which is now built into most models. Because this involves further unknown parameters, further data is needed, and a further reduction in the number of parameters can be made using regression-type modeling with a link function and covariates. To aid in the theory development, a constant and known probability of detection is assumed to begin with, which then leads to considering a constant unknown probability of detection. Estimates and variance estimates are derived. The theory is extended to the case of constant but unknown detection probability and then to variable unknown detection probabilities. Further data can be obtained using time replicate counts or spatial replication, double sampling, and partial repeat sampling. In some populations, counts out to a certain distance are made from a point, referred to as “point” counts, involving circular plots. Resource selection is considered briefly, and other detection methods arising in later chapters are listed. Finally, some mention is made of the possibilities of using adaptive sampling for sparse and clustered populations.

Chapter 4

Occupancy, Abundance, and Related Topics



4.1 An Overview

In this book, our focus is on estimating closed population sizes. However, an associated problem arises when we are interested in whether a sample site contains a certain species of individual or not, and then there is the question of how to proceed when some individuals go undetected. For example, what proportion of sites are occupied by one or more individuals of a given species? The term given to this scenario is “site occupancy,” and clearly incomplete detection is an important factor that needs to be considered (Kéry et al., 2010). Sometimes, the focus is on just the estimation of the probability of occupancy or “presence,” or it can be combined with abundance estimation.

We shall use the term site here rather than plot that was used in previous chapters since the site may be a naturally occurring sampling unit such as a discrete pond or patch of vegetation, an investigator-defined monitoring station, a quadrat chosen from a predefined area of interest, a point at the center of a circle of prescribed or nonprescribed radius (Chap. 7), or a line transect with selected points of observation or other sampling methods such as cluster sampling (Chap. 8). This is a very large chapter and can occupy a whole book (e.g., MacKenzie et al., 2018) as it overlaps with the next three chapters depending on the method of sampling, so there will be some repetition of notation for the reader, as well as cross-referencing. Occupancy methods can also be used for open populations when occupancy data is combined with count data (e.g., Zipkin et al., 2017). It has also been applied to paleontology and the investigation of fossils (e.g., Liow, 2013).

There are two aspects of occupancy underlying the models discussed below. We can have a false positive where we have an object not present but detected as present (e.g., wrong species identified) or a false negative where we have an object that is present but not detected, the more common situation.

In the past, there has been a tendency to ignore imperfect detection in ecology (Kellner & Swihart, 2014), which can have a major effect on estimation (Gu & Swihart, 2004), as we saw in Chap. 3. However, models below incorporate detectability. With technical advances, we are beginning to see the use of autonomous recording units (Pérez-Granados et al., 2019 and the review by Sugai et al., 2019 for some references). Acoustic methods are considered in Sects. 8.8 and 8.13.1.

Detectability can depend on a lot of factors, which often can be built into a model using covariates such as weather, environment, and habitat, as well as spatial and temporal factors. Tanadini and Schmidt (2011) showed that detectability for six amphibian species can also depend on population size, which can have consequences for small and rare populations, and also affect long-term monitoring programs. Sites were visited three times so that a detection history (110) at a site has probability $\psi p_1 p_2 (1 - p_3)$, where ψ is the probability that a species is present and p_i the probability of being detected at visit i , given present.

In addition to determining site occupancy, those sites found to be occupied can be used to estimate the area of occupation by the species or community. A related problem is considered by Etterson et al. (2021) who use a spatially explicit model for estimating risks of pesticide exposure to bird populations. As mentioned above, the topic of site occupancy involves jointly modeling two probabilities, the probability of occupancy (ecological factor) and the conditional probability of detection given an individual animal or plant is present (observation factor). The product of these two probabilities is the unconditional probability of detection, which is what has been used in Chap. 3. Also, covariates are often needed in the model, for example, the type of habitat could be built into the probability of detection (Kéry et al., 2006). Just because a certain species of plant is not detected, it does not mean it is extinct from the site.

In the past, there has been a focus primarily on stationary objects such as plants, but studies on the degree of site occupancy have also been used for insects, fish, birds, and various other species of animals. They have also been used for (1) multiple species investigations and their interactions (or independence), (2) endangered species where numbers are low and there is a problem of distinguishing between low numbers and low probability of detection, (3) large-scale monitoring programs, and (4) so-called metapopulation dynamics. Gálvez et al. (2016) considered cost-efficient effort allocation for camera traps and examined camera trap survey costs as a function of the number of sample units (SU), survey duration, and camera traps per SU, linking costs to precision in occupancy estimation.

Camera methods have been used to examine occupancy (see the extensive review by Burton et al. (2015) and Sect. 10.6), though they are more suited to mobile populations, but movement can cause problems (Neilson et al., 2018). Spatial models are becoming more popular for investigating occupancy including using camera methods. Sutherland et al. (2014) considered patch occupancy models. Occupancy has been extended to multiple species, which are discussed in Sect. 5.3.3.

4.1.1 Basic Model

An introduction to a basic site occupancy model is as follows. We have a population with a selected number of sites and a number of visits to each site to observe the presence or absence of a given species. For site s ($s = 1, \dots, S$) and visit t ($t = 1, \dots, T$), let $z_s = 1$ with probability ψ_s if site s is occupied and 0 otherwise, and let $y_{st} = 1$ with probability p_{st} , the probability of observation/encounter at site s and visit t given present, and 0 otherwise. Then, if $\mathbf{y}_s = (y_{s1}, \dots, y_{sT})'$,

$$\Pr(\mathbf{y}_s | z_s = 1) = \prod_{t=1}^T p_{st}^{y_{st}} (1 - p_{st})^{1-y_{st}}, \quad s = 1, \dots, S.$$

The likelihood over all the sites is then

$$L(\{\psi_s\}, \{p_{st}\}) = \prod_{s=1}^S \{\Pr(\mathbf{y}_s | z_s = 1)\psi_s + I(\mathbf{y}_s)(1 - \psi_s)\},$$

where I is an indicator variable with $I(\mathbf{y}_s = \mathbf{0}) = 1$ and $I(\mathbf{y}_s \neq \mathbf{0}) = 0$, otherwise. The second part of the above equation refers to the case when $z_s = 0$, i.e., absent, and not $\mathbf{y}_s = \mathbf{0}$ when present. It is usually assumed that $\text{logit}(\psi_s)$ is linear in covariates and maximum likelihood estimation can be carried out.

MacKenzie and Royle (2005) referred to three general sampling schemes that have been considered: (1) S sites are revisited T times (the standard method); (2) S sites are revisited T times, and S_1 of the sites are initially selected to be used to determine which sites have repeated surveys (a form of double sampling); and (3) a removal design where S sites are revisited up to T times, but surveying stops at a site once an individual is detected. They found that the second design may not be a good design, in most circumstances.

Various Bayesian models can be used. For example, p_{st} can be expressed as logistic linear regression in covariates, say,

$$\text{logit}(p_{st}) = \beta_0 + \beta_1 u_s + \varepsilon_t,$$

where u_s is a covariate, ε_t is a random effect having some distribution with prior distributions for its parameters, and the coefficient β_1 has a suitable prior. We also have z_{st} , which is estimated by applying Markov chain Monte Carlo (MCMC) to posterior distributions.

The basic model is often expressed in the form

$$z_s | \psi_s \sim \text{Bernoulli}(\psi_s), \quad y | z_s \sim \text{Bernoulli}(z_s p_s),$$

so that $y_s \sim \text{Binomial}(T, p_s z_s)$. As already mentioned, covariates are usually added giving the model

$$\text{logit}(\psi_s) = \alpha + \boldsymbol{\beta}' \mathbf{x}_s,$$

with \mathbf{x}_s measurements on the covariates (e.g., environmental variables) for site s . A logarithmic or logistic model with covariates can also be used for p_s . Parameters can be then estimated using maximum likelihood estimation.

Welsh et al. (2013) considered such a model and found that it is not always straightforward to fit as the estimating equations can sometimes have multiple solutions, including boundary estimates which produce fitted probabilities of zero or one. They also found that the estimates are unstable when the data are sparse and raised other issues relating to allowing or not allowing for incomplete detection.

As already mentioned, a Bayesian method can be applied to occupancy using prior distributions for the unknown parameters such as α and $\boldsymbol{\beta}$ mentioned above. Typically, a parameter like α , for example, is given a normal prior distribution $N(0, \sigma^2)$, where σ^2 also has a prior distribution. Using the data, posterior distributions are then obtained for the parameters for estimation purposes. A key question is then how much the prior distributions affect estimation. In an endeavor to minimize this effect, so-called “vague” and “uninformative” priors are frequently used. However, it should be realized that the priors are applied in the context of a logistic model. Northrup and Gerber (2018) demonstrated that such priors can easily be unintentionally highly informative using normal priors. Clearly, the choice of prior is important. Bayesian methods are sprinkled throughout this chapter, along with maximum likelihood methods.

Clark and Altweig (2019) used a Bayesian approach model along with logit link functions, similar to those used above. Using the above notation, they used the logistic regressions $\text{logit}(\psi_s) = \mathbf{x}'_s \boldsymbol{\beta}$ and $\text{logit}(p_{st}) = \mathbf{w}'_{st} \boldsymbol{\alpha}$, where \mathbf{x}_s relate to covariates for occupancy and \mathbf{w}_{st} relate to covariates for detection. Let $\pi_1(\boldsymbol{\alpha})$ and $\pi_2(\boldsymbol{\beta})$ be the prior distributions of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, respectively. Using $\mathbf{z} = \{z_s\}$ and $\mathbf{Y} = (y_{st})$, the joint posterior distribution of the model parameters is

$$f(\mathbf{z}, \boldsymbol{\alpha}, \boldsymbol{\beta} | \mathbf{Y}) \propto \pi_1(\boldsymbol{\alpha}) \pi_2(\boldsymbol{\beta}) \left(\prod_{s=1}^S \psi_s^{z_s} (1 - \psi_s)^{1-z_s} \right) \prod_{s:z_s=1} \prod_t p_{st}^{y_{st}} (1 - p_{st})^{1-y_{st}}.$$

A Gibbs sampling algorithm is used to sample from the marginal posterior distributions of the parameters. The method was compared with previous Bayesian methods.

4.1.2 Animal Signs

In some situations, signs can be used such as active burrows of the small mammal, Chinese pangolin (*Manis pentadactyla*), as considered by Sharma et al. (2020). We can also use automatic acoustic recordings for birds and anurans (Chambert et al., 2018) and bats using multiple automated echolocation detectors (Gorresen et al., 2013). Wrege et al. (2017) gave a broad picture of aspects of acoustic monitoring especially with regard to tropical forests. Acoustic methods are discussed further in Sect. 8.8 and have also been used for marine populations (Sect. 8.13.1) as well as for land populations. In applying such models, it is generally assumed that a population index I is proportional to abundance N , namely, $I = \beta N$, where β is a function of N . Stanley and Royle (2005) provided a method using this method as follows.

Given S detection stations (sites), let $w_{st} = 1$ if station s ($s = 1, \dots, S$) with N_s individuals was visited between occasions $t - 1$ and t ($t = 1, \dots, T_s$) during time interval τ_{st} (say days) and $w_{st} = 0$ otherwise. (Here, $t = 0$ is defined to be the occasion on which stations are established.) Let p_{st} be the unconditional detection probability for the s th detection station on the t th sampling occasion. If r is the detection probability of an individual over a unit time interval (conditional on presence), then

$$p_{st} = \Pr(w_{st} = 1 \mid N_s, r) = 1 - (1 - r)^{\tau_{st} N_s},$$

and the likelihood for the s th station is

$$L(r \mid \{w_{st}\}, N_s) = \prod_{t=1}^{T_s} p_{st}^{w_{st}} (1 - p_{st})^{1-w_{st}}.$$

Since the N_s are unknown, we need to sum them out, given they are realizations of a random variable with discrete probability distribution $f(N; \lambda)$, where λ is a parameter vector and $\Pr(N = k) = f_k$ ($k = 0, 1, \dots$). Then, assuming independence of data from the S stations,

$$L(r, \lambda \mid \{w_{st}\}) = \prod_{s=1}^S \left[\sum_{k=0}^K \left\{ \prod_{t=1}^{T_s} p_{st}^{w_{st}} (1 - p_{st})^{1-w_{st}} \right\} f_k \right],$$

where $p_{st} = 1 - (1 - r)^{\tau_{st} k}$ and K is the number of support points (Royle & Nichols, 2003). Here, $f(\cdot)$ could be a Poisson or negative binomial distribution. Linear models for $\log(\lambda_s)$ and $\text{logit}(r)$ using covariates can also be fitted. Examples were given for tracking tubes for small mammals and scent stations. Stanley and Royle (2005) extended this approach.

One problem that can arise with signs is the possible deterioration of a sign. Rhodes et al. (2011) assumed that an individual present will leave signs, say y per day, and $\Pr(T > t \mid v)$ is the probability that a sign deposited on day v survives

until at least day t . If ϕ_u is the probability of surviving to day u , then $\Pr(T > t | v) = \prod_{u=v}^t \phi_u$, and the probability that all signs at a site have decayed by time t (i.e., the probability no signs are detected, given the site is occupied) is

$$\Pr(x = 0 | \text{occupied site}) = \prod_{v=t-d}^t [(1 - \Pr(T > t | v))]^y,$$

where x is the number of signs found during the survey and d is a period of time long enough that $\Pr(T > t | t-d)$ is very small, approximately zero. Further, the expected age of signs found at the time of a survey (i.e., the time interval prior to the survey that signs relate) is

$$E[t - v] = \frac{y \sum_{v=t-d}^t \Pr((T > t | v)(t-v))}{y \sum_{v=t-d}^t \Pr((T > t | v))},$$

where $t - v$ is the age of the sign. This sets out the basic model. Using sample pellets, koala pellet decay trials were conducted in two study areas in New South Wales, Australia, to estimate ϕ_{ijt} the daily pellet survival probability for day t in sampling interval j in pellet group i . Then logit(ϕ_{ijt}) was expressed as a linear model in terms of covariates and a random effect. Further details of the model are given by the authors.

4.1.3 Some Reviews

A helpful review of occupancy is given by Bailey et al. (2014), and a broad coverage of the subject is given by MacKenzie et al. (2018), which we will sometimes use in this chapter. Further details are given in Kéry and Schaub (2012, Chapter 13), where practical examples are given along with computer code, as well as extensions to multistate site occupancy models (see also Sect. 4.17). Another comprehensive source is Royle and Dorazio (2009: Chapters 3 and 4), which includes metapopulations as occupancy varies over time (their Chap. 9) and analyzes many examples using computer code. Occupancy trends were investigated by Dennis et al. (2019), who included an extensive bibliography.

One other variation involves decomposing the probability of detection into components allowing for the fact that an individual may be present but not available (Sect. 7.8). For practical issues like choosing the number of sites versus the number of surveys and the type of design used, the reader is referred to MacKenzie et al. (2006, Chapter 6) and MacKenzie et al. (2018, Chapters 11 and 12). Also, the authors considered multi-season models (MacKenzie & Nichols, 2004), modeling dynamic processes, multiple species models, and community-level studies. Occupancy modeling has been compared with radio telemetry by Duquette et al. (2014).

4.1.4 Rare Species

A problem that arises when allowing for incomplete detection is that rare species or species with a low detection rate will have less precise estimates and wide confidence intervals because of less data and the need to estimate more parameters. Banks-Leite et al. (2014) in their review noted that: “Paradoxically … models that adjust for imperfect detection are most effective when a species or individual is commonly detected, yet in this situation there is less need to adjust for imperfect detection because the raw data already indicate if or when the species/individual is present.”

With rare or poorly detected species, their fitted probabilities are often found to be equal to one (Welsh et al., 2013), a result that would generally suggest that the species is in effect widespread. The latter authors questioned the usefulness of models that estimate species occupancy while accounting for detectability. This was responded to by Guillera-Arroita et al. (2014) who illustrated the dangers of disregarding imperfect detection. When imperfect detection is ignored, occupancy and detection are confounded. However, “hierarchical occupancy models separate occupancy and detection, and imprecise estimates simply indicate that more data are required for robust inference about the system in question.” It is clear that the reader needs to be aware of the issues raised in applying the methods in this chapter. Species methods are considered in Chap. 5.

4.1.5 Repeated Counts

In the above occupancy models, the occupancy probability and detection probability are separately estimable if sites are visited and sampled more than once, though a single visit model can be useful in some circumstances. Rather than record all occasions when an individual is encountered, it is sometimes preferable to record just the first occasion when an individual is first sighted. This is essentially a removal model, discussed in Chap. 10, as an individual is “removed” from the sampling protocol after its first encounter (e.g., Rota et al., 2011).

As noted by Halstead et al. (2018), both double-observer and metapopulation occupancy designs are costly because they require multiple observers or multiple visits to sites. They suggested using the removal approach, which means that more sites can be sampled, thus reducing time and effort. Also, the times to/of detection provide useful information, such as for designing future studies.

The RN Model

Royle and Nichols (2003) introduced a model in an extensive paper called the RN model, often referred to in the literature and similar to the basic model in Sect. 4.1.1, which considered the estimation of abundance from repeated presence-absence data and allowed for heterogeneity in the detection probabilities. Let N_s be the number of animals at site s and y_s the number observed there ($s = 1, \dots, S$). It is assumed

that each site has probability ψ of being occupied. Let $y_{st} = 1$ if detection occurs at site s in repeat observation t ($t = 1, \dots, T$) and then $y_s = \sum_t y_{st}$. Let p_s be the probability of observation at site s , conditional on presence. If $p_s = p$ for all s , then

$$f(\{y_s\}) = \left\{ \prod_{s:y_s > 0} \binom{T}{y_s} p^{y_s} (1-p)^{T-y_s} \psi \right\} \times \left\{ \prod_{s:y_s=0} [\psi(1-p)^T + (1-\psi)] \right\}.$$

Heterogeneity is modeled on the assumption that the detection probability varies by site location only according to the number N_s of animals available to be detected, namely, $P_s = 1 - (1-r)^{N_s}$, which replaces p in the above equation. Here, r is a binomial sampling probability that a particular individual is detected. Prior distributions on the parameters and N_s are used, and a variety of models were considered. Haines (2016) referred to an integrated likelihood for a Poisson mixing distribution used for the RN model and showed that it can be expressed as a finite rather than an infinite sum of terms. We now consider the question of what is the optimal number of visits to achieve some criterion.

Number of Visits

The number of visits needed to obtain a desired precision will depend on the occupancy and detection probabilities (Guillera-Arroita et al., 2010). Generally, species with high occupancy and low detection probabilities (i.e., species that are widespread but hard to find where they occur) need the most repeat visits, whereas species with high detection probabilities (easy to find) need fewest. It must be born in mind that, in general, the population must closed, which may not be true with an increasing number of visits to a site. However, this assumption may be relaxed, provided that any changes occur completely at random (MacKenzie et al., 2006).

If Q is the probability of not seeing a species after T visits, with average detection probability p , then $Q = (1-p)^T$. To be 95% confident that a species is absent, then $Q = 0.05$, and T_{\min} , the minimum number of visits necessary, is $T_{\min} = \log(0.05)/\log(1-p)$. Alternatively, taking a Bayesian approach (Wintle et al., 2012, 2005), let p' be the prior probability (before data collection) that a species is detected in any single visit at single site, and let ψ' be the corresponding prior probability of presence. If T visits are made to a site, then the probability of T non-detections is $(1-p')^T$. By Bayes' theorem, the posterior probability that the site is occupied is

$$\psi = \frac{\psi'(1-p')^T}{(\psi'(1-p')^T + (1-\psi'))}$$

and $1 - \psi$ is the posterior probability of being absent from the site. The model ignores the possibility of false presences arising from the misidentification of species, although it may be extended to do so. Rearranging the above equation, and solving for T , the number of sequential non-detections (T^*) to achieve a particular

posterior probability of absence from the site ($1 - \psi$) is given by

$$T^* \geq \frac{\log(\psi/(1 - \psi)) - \log(\psi'/(1 - \psi'))}{\log(1 - p')}.$$

For example, we could consider $\psi = 0.05$ to be 95% certain of absence. Wintle et al. (2012) then used the basic model given in Sect. 4.1.1 using a Bayesian method.

Bailey et al. (2007) noted that the need for temporal replication at sampled sites to estimate detection probability creates a trade-off between spatial replication (number of sample sites distributed within the area of interest/inference) and temporal replication (number of repeated surveys at each site). MacKenzie and Royle (2005) presented the first investigation of these trade-offs, and their findings provided some needed guidance for efficient design of occupancy studies, though based on a simple model and large sample theory. Three types of design were considered: the “standard design,” in which each site is surveyed the same number of times; the “double sampling design,” in which a subset of sites is surveyed multiple times and the remaining sites are sampled only once; and the “removal design,” in which sites are surveyed multiple times until the target species is detected. They described software (program GENPRES) developed to allow investigators to easily explore design trade-offs and sampling limitations. This program is available online.

Dependent Replicates

Hines et al. (2010) considered occupancy modeling using a sampling design based on spatial replicates that are not selected randomly and that are expected to exhibit Markovian dependence. Focusing on tiger trails and picking up signs such as tracks, scats, and scent, the population area consisted of large grid cells (sites) based on an expected home range of approximately 200 km². Within each cell, the sample s was a trail. It was assumed that the survey began at one end of the trail and then walked by three trained surveyors looking for signs, with the distance covered subdivided into segments of equal length (e.g., 1 km) in order along the K total segments of the route. These were then treated as geographic replicates in the occupancy analysis, but as individual animals will walk for distances exceeding segment length, counts on segments will not be independent, indicating a form of cluster sampling. Here, not sampling with replacement was not regarded as critical, as the species was wide ranging and highly mobile. For sample unit s ($s = 1, \dots, S$), let

$$p_s = \Pr(\text{detection at a segment} \mid s \text{ occupied, species present on segment})$$

$$\psi_s = \Pr(\text{sample unit } s \text{ occupied})$$

$$\theta_s = \Pr(\text{species present on segment} \mid s \text{ occupied, species not present on previous segment})$$

$$\phi_s = \Pr(\text{species present on segment} \mid s \text{ occupied, species present on previous segment})$$

As usual, we score “1” for a detection of the species and “0” otherwise. For example, for cell s , we can have $h_s = (01011)$ for five segments. Then

$$\begin{aligned}\Pr(h_s = 01011) \\ = \psi_s[(1 - \theta_s)\theta_s + \theta_s(1 - p_s)\phi_s] \times p_s(1 - \phi_s)\theta_s + \phi_s(1 - p_s)\phi_s]p_s\phi_s p_s.\end{aligned}$$

Another example is given but with just two segments, namely,

$$\Pr(h_s = 00) = (1 - \psi_s) + \psi_s[\theta_s(1 - p_s)(1 - \phi_s p_s) + (1 - \theta_s)(1 - \theta_s p_s)].$$

The likelihood function is then

$$L(\{\psi_s, \theta_s, \phi_s, p_s\} | \{h_s\}) = \prod_{s=1}^S \Pr(h_s).$$

Two new occupancy models for data collected under this sort of design were introduced, one based on an underlying Markov model for spatial dependence and the other based on a trap response model with Markovian detections. Computational details are given in the appendix of Hines et al. (2010).

Karanth et al. (2011) used the above model, which explicitly accounted for such potential dependency using a first-order Markov process and the variant that assumes that surveys can begin on any randomly chosen replicate. Various covariates were considered. Jathanna et al. (2015) also used the model to determine the habitat occupancy of Asian elephants in India using fresh dung samples. The landscape included 14 wildlife reserves that cover an area of 5,500 km² and lie in a human-dominated matrix populated by more than 10.2 million people. The elephant habitats are subject to threats from illegal hunting, livestock grazing, and forest biomass extraction, as well as pressures from developmental projects and industrial growth. Because of low elephant densities, low sighting probabilities, and logistical factors, rather than sightings, the authors chose to survey and detect elephant presence via signs, primarily fresh dung. A dynamic model for elephants is discussed below in Sect. 4.17.2.

The model of Hines et al. (2010) was extended by Hines et al. (2014) to several seasons by adding an additional time component that also permits inference about local probabilities of extinction, colonization, and occupancy for sampling.

4.1.6 Small Scale and Abundance

There are situations where, because of the small spatial scale used, occupancy and abundance are essentially the same, as the site chosen is so small that at most one individual or pair can occupy it (e.g., some spotted owl data sets and the habitat selection of kittiwakes in breeding cliffs with regard to potential nest sites).

Occupancy is therefore sometimes used as surrogate for abundance (e.g., Casner et al., 2014, butterflies), like a measure of relative density, as one would expect it to be related to population density if certain underlying conditions remain constant. In some sparse populations, it is difficult to use capture-recapture and other methods to estimate population size and density. Instead, presence-absence (detection or non-detection) data can be readily collected and used to estimate size and density. We consider this relationship further below.

As we shall see in this chapter, occupancy studies can be used to examine colonization and local extinction (e.g., Heard et al., 2013), the efficacy of wildlife monitoring programs, species habitat relationships, and general changes in occupancy over time (Altweig et al., 2008). Kéry (2004) demonstrated that detectability needs to be included in considering extinction rates for plants and recommended the use of capture-recapture models even for plants. We note that the term occupancy is sometimes referred to as the proportion of area occupied by a species (MacKenzie & Nichols, 2004).

4.1.7 *Invasive Species*

Invasive species is an important topic as it represents a huge worldwide problem and will occur elsewhere in this book (e.g., Chap. 5), as well as scattered throughout this chapter below. An important application is where the presence of a disease needs to be detected and to derive estimates of population prevalence and test sensitivity under imperfect pathogen detection (Lachish et al., 2012). Also, estimating the degree of site occupancy is particularly important in dealing with invasive species where it is essential to get in early to identify a recently established species. Unfortunately, we are not always aware of an invasion until it is well established. This means it is important to detect low-density so-called “founder” populations, which can help to slow or even stop the spread of a harmful invasion (Fitzpatrick et al., 2009).

With regard to invasive species, it is important to get some idea of the big picture. Pyšek et al. (2012) presented the first global overview of frequencies of significant and non-significant ecological impacts and their directions on 15 outcomes related to the responses of resident populations, species, communities, and ecosystems. They referred to 287 publications with 1551 individual cases that addressed the impact of 167 invasive plant species belonging to 49 families.

Yackulic et al. (2012) looked at the expansion of the barred owl (*Strix varia*) in Western Oregon, USA, over the period 1990–2010. The basic parameters are (1) the probability of extinction, ε_{st} , which is the probability that site s is not occupied at time $t + 1$, given the site is occupied at time t ; (2) the probability of colonization γ_{st} , which is the probability that site s is occupied at time $t + 1$, given the site is not occupied at t ; and ψ_{st} , the probability that site s is occupied by focal species at t .

We then have the basic equation

$$\psi_{s,t+1} = \psi_{st}(1 - \varepsilon_{st}) + (1 - \psi_{st})\gamma_{st}.$$

Under a so-called autologistic modeling approach, probabilities of local extinction and colonization are influenced by the occupancy status of neighboring patches, where “neighbor” may be defined in a variety of ways, depending on the ecology of the modeled system. For species with very limited movement, the neighborhood of a focal site may be restricted to the set of other sites whose borders touch those of the focal site. For wide-ranging species, the neighborhood may include sites that are not in direct contact with the focal site, but lie within some specified distance. We can use either the total number of occupied sites in a neighborhood θ_{is} or the average probability of occupancy for sites in a neighborhood. For the former, we have the estimator

$$\widehat{\theta}_{st}^{(n_s)} = \sum_{j \in \{n_s\}} \widehat{\psi}_{jt},$$

where n_s is the set of sites constituting the neighborhood of focal site s . For the latter, we have the estimator

$$\widehat{\psi}_t^{(n_s)} = \frac{1}{\ell_s} \sum_{j \in \{n_s\}} \widehat{\psi}_{jt},$$

where ℓ_s is the number of sites located in the neighborhood of focal site s . We can then use

$$\text{logit}(\gamma_{st}) = \beta_{0t} + \beta_1 \widehat{\psi}_t^{(n_s)} + \beta_2 x_{st},$$

where x_{st} is the value of a habitat covariate for site s at time t . Further details are given by the authors Yackulic et al. (2012).

4.2 Detection Probabilities

We now focus on the detection probabilities introduced in the above basic model and begin with the special case of ψ the probability of presence and p the probability of detection, given presence, both constant across all site measurements.

4.2.1 Homogeneous Detection Probabilities

Let S be the number of sites sampled, and suppose each site is surveyed T times. Then the probability of observing y_s detections at site s ($s = 1, \dots, S$) with T surveys is

$$\Pr(y_s | \psi, p) = \psi \binom{T}{y_s} p^{y_s} (1-p)^{T-y_s} + (1-\psi) I(y_s = 0; \psi, p),$$

where $I(\cdot)$ is the usual indicator function taking the value 1 when $y_s = 0$ and 0 otherwise, ψ is the presence (occupancy) probability, and p is the probability of detection given presence. Here,

$$\Pr(y = y_0 = 0 | \psi, p) = \psi(1-p)^T + (1-\psi),$$

so that the (mixture) distribution used here is the (conditional on presence) binomial distribution, but is inflated at zero. Then the likelihood function is

$$L(\psi, p; \{y_s\}) = \prod_{i=0}^S \Pr(y_s | \psi, p).$$

This likelihood can be used to obtain maximum likelihood estimates of p and ψ . Also, we can use Bayesian methods using uniform $(0, 1)$ prior distributions for p and ψ . Computer packages for carrying out data analysis for such methods are available from the packages **WinBUGS** and **R**.

As previously mentioned, we shall see in this chapter that both log-linear and logistic models for the parameters can incorporate the effect of covariates. For example, if ψ_i is the probability of occupancy for site i , then Tyre et al. (2003) used the logistic model

$$\text{logit}(\psi_i) = \beta_0 + \beta h_i,$$

where h_i is the value of the habitat variable h ; other covariates can be added. The authors assumed that $h_i \sim N(0, 1)$, thus also incorporating a Bayesian approach. They were particularly interested in the parameter $1 - p$, the probability of not detecting an individual at a site and particular occasion. This probability is referred to below as the probability of a false-negative error.

4.2.2 Variable Parameters

Karavarsamis and Huggins (2019, 2020) developed the following two-stage conditional model with ψ_s , the probability that site s ($s = 1, \dots, S$) is occupied by

a species, and p_{st} , the probability that the species is observed at site s on visit t , given it is present at site s . Then $\theta_s = 1 - \prod_{t=1}^T (1 - p_{st})^T$ is the probability of at least one detection over T visits at site s , given the site is occupied. Let $y_{st} = 1$ if an individual was detected at site s on occasion t and 0 otherwise, and let $y_s = \sum_{t=1}^T y_{st}$ denote the total number of occasions a species is observed at site s . Let f_0 be the number of sites where none of the species was detected, and let $r = S - f_0$ be the number of sites where they were. We reorder the sites as $i = 1, \dots, r, r+1, \dots, S$, where $1, \dots, r$ denote the sites at which at least one detection occurred and $r+1, \dots, S$ the remaining sites at which no sightings occurred. Let $z_s = I(y_s = 0)$ be the indicator of no detections at site s . The contribution to the full likelihood for site s is

$$L(\psi_s, \{p_{st}\}) = (1 - \psi_s \theta_s)^{z_s} \left\{ \psi_s \prod_{t=1}^T p_{st}^{y_{st}} (1 - p_{st})^{1-y_{st}} \right\}^{1-z_s}.$$

If the sites are independent, the overall likelihood is $\prod_s L(\psi_s, \{p_{st}\})$.

It is helpful to set $\eta_s = \psi_s \theta_s$ giving the above likelihood

$$\begin{aligned} L(\eta_s, \{p_{st}\}) &= (1 - \eta_s)^{z_s} \eta_s^{1-z_s} \times \left\{ \frac{\prod_{t=1}^T p_{st}^{y_{st}} (1 - p_{st})^{1-y_{st}}}{\theta_s} \right\}^{1-z_s} \\ &= L_1(\eta_s) L_2(\{p_{st}\}), \end{aligned}$$

which can be used to simplify the calculations for complex models when both $\logit \psi_s$ and $\logit p_{st}$ are expressed as linear regressions in covariates. Here, L_1 is a Bernoulli likelihood, while L_2 is the likelihood of the p_{st} conditional at least one sighting on site s . Using the conditional model, estimation methods and asymptotic variances were then given for the parameters. The authors considered the special case of $p_{st} = p_s$, the time-independent case.

4.2.3 Heterogeneous Detection Probabilities

If p_s is the probability of detection in site s , MacKenzie et al. (2006, Chapter 5) considered a heterogeneous detection probability model mimicking the capture-recapture model M_h of Chap. 12 with varying catchability (cf. Wintle et al., 2004). We saw in the previous section that we can allow for heterogeneous presence probabilities by allowing h , the habitat variable, to be a random variable with a prior distribution. We can do the same for p_s by assuming a prior density function $f(p; \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is the vector of the parameters of the (mixing) distribution. Conditional on a site being occupied, the probability of obtaining y detections at a site may be computed by integrating the binomial likelihood over g , namely, the

conditional distribution

$$g_c(y; \boldsymbol{\theta}) = \Pr(y | \boldsymbol{\theta}) = \int_0^1 \text{Binomial}(y; T, p)g(p; \boldsymbol{\theta})dp.$$

The unconditional, zero-inflated, distribution is now

$$g(y; \boldsymbol{\theta}, \psi) = \psi g_c(y; \boldsymbol{\theta}) + (1 - \psi)I(y_s = 0),$$

leading to the likelihood function

$$L(\psi, \boldsymbol{\theta} | \{n_k\}) = \prod_{k=0}^T [g(y; \boldsymbol{\theta}, \psi)]^{n_k},$$

where n_k is the number of sites at which k detections were observed. Instead of the binomial (conditional) distribution used in the previous section, we can have a general probability distribution $f(p; \boldsymbol{\theta})$. Royle (2006) considered a number of different distributions for g . For example, the beta distribution with parameters α, β leads to the integrated distribution

$$g_c(y; \boldsymbol{\theta}) = \frac{\Gamma(T+1)}{\Gamma(k+1)\Gamma(T-y+1)} \cdot \frac{\Gamma(\alpha+k)\Gamma(T+\beta-k)}{\Gamma(\alpha+\beta+T)} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}.$$

Other models are the logit normal, where $\text{logit}(p)$ has a normal distribution, and using finite mixtures.

In using such distributions, there is a problem encountered in identifying ψ , which parallels what happens in model M_h (Link, 2003) where very similar distributions $f(p; \boldsymbol{\theta})$ can lead to very different estimates of ψ . Royle (2006) looked at this problem and noted that it is possible to construct two distributions, say g_1 and g_2 , that yield the same marginal cell probabilities for the data but different ψ 's, so that ψ is not identifiable across mixtures.

The reader is referred to MacKenzie et al. (2006, Chapter 6) for practical issues like choosing the number of sites versus the number of surveys and the type of design used. Also, the authors considered multi-season models (see also MacKenzie & Nichols, 2004), modeling dynamic processes, multiple species models, and community-level studies. General design questions are discussed further in Sect. 4.6.

4.2.4 Model Checking

Given the large number of papers looked at for this book, very few provided methods for detecting departures from models. Overall, goodness-of-fit methods have their place, but are generally not focused on specific departures of interest (e.g., non-

independent observations). Occupancy models are tricky as there are two quite different components, presence and detection. Further general checks are considered below.

Goodness-of-Fit Test

MacKenzie and Bailey (2004) described a goodness-of-fit statistic based on the observation histories of *Plethodon glutinosus* complex (group) in Great Smoky Mountains National Park from site s ($s = 1, \dots, S$). We have the usual binary variables $z_s \sim \text{Bernoulli}(\psi_s)$ for site s where ψ_s is the probability of occupancy ($z_s = 1$) of members of the complex and $y_{st} | z_s \sim \text{Bernoulli}(p_{st})$, where $y_{st} = 1$ with probability p_{st} of the complex being observed at site s on the t th visit ($t = 1, \dots, T$), given present ($z_s = 1$). Let $\mathbf{y}_s = (y_{s1}, \dots, y_{sT})'$. It was assumed that there were no false-positive errors. If $n_s = \sum_{t=1}^T y_{st}$ is the number of times that members of the group are observed at site s , then the probability of observing a particular capture sequence \mathbf{y}_s is

$$\Pr(\mathbf{y}_s) = \begin{cases} \psi_s \prod_{t=1}^T p_{st}^{y_{st}} (1 - p_{st})^{1-y_{st}}, & \text{if } n_s > 0 \\ (1 - \psi_s) + \psi_s \prod_{t=1}^T (1 - p_{st}), & \text{if } n_s = 0. \end{cases}$$

If there are $T = 3$ visits, a possible encounter history $X_s = (101)$ for site s means the presence of the complex is observed in the site on visits 1 and 3, but not on visit 2. The likelihood is then

$$L(X_1, \dots, X_S | \{\psi\}, \{p_{st}\}) = \prod_{s=1}^S \Pr(X_s).$$

A logistic regression can be used with covariates, which reduces the number of parameters, and the above likelihood can be maximized to obtain maximum likelihood estimates of unknown parameters.

To construct a goodness-of-fit test, let O_h be then the number of sites observed to have detection history h and E_h be the expected number of sites with history h summed across sites, according to the current model being used, namely,

$$E_h = \sum_{s=1}^S \Pr(X_s = h).$$

Assuming no missing observations, there are 2^T possible sequences of ones and zeros in a \mathbf{y}_s . Then, the Pearson chi-square statistic is

$$\chi^2 = \sum_{h=1}^{2^T} \frac{(O_h - \widehat{E}_h)^2}{\widehat{E}_h},$$

where \widehat{E}_h is E_h with the parameters estimated.

There are two problems with the above test statistic, which questions its ability to have a chi-square distribution. Firstly, E_h involves pooling across sites assuming parameters across sites are constant, which is unlikely to hold. Secondly, some of the \widehat{E}_h may be relatively small (< 2) for even moderate T (say ≥ 5). The authors suggested using a parametric bootstrap technique and listed the steps involved to test for significance. Repeatedly, new sets of detection histories are generated from the model, which is refitted to each, and the chi-square statistic is recomputed in order to get a sense for what range of values for χ^2 to expect if the model were true. The coefficient of overdispersion (variance inflation factor) can be estimated, and various scenarios were investigated using simulation.

Similar test statistics have been used in posterior predictive checks for Bayesian analyses.

Using Residuals

With regression models, underlying assumptions can be investigated using residuals and associated plots. Warton et al. (2017) used the previous model of MacKenzie and Bailey (2004) along with the covariate models

$$\text{logit}(\psi_s) = \mathbf{x}'_i \boldsymbol{\beta} \text{ and } \text{logit}(p_{st}) = \mathbf{v}'_{st} \boldsymbol{\gamma},$$

for certain covariates \mathbf{x}_s and \mathbf{v}_{st} . Here, p_{st} is again the probability of detection of some individuals from the group at site s and visit t , given the site is occupied. Detection events are assumed independently with no possibility of false positives, i.e., sites with at least one detection are truly occupied ($z = 1$), and the likelihood is the product of the $\Pr(y_s)$ across the S sites as above. Parameters can be estimated by maximum likelihood or using a Bayesian approach.

Having set up the model, the authors considered testing for the following: validity of logistic regressions, independence of detections, constant occupancy across surveys, and no heterogeneity of measurements. They began with the usual goodness-of-fit test described above. Because not all detection histories will be observed a large number of times, χ^2 can be poorly approximated by a chi-square distribution, and MacKenzie and Bailey (2004) suggested estimating its null distribution by simulation, which is computationally intensive. Repeatedly, new sets of detection histories are generated from the fitted model (a parametric bootstrap), the model is refitted to each, and the chi-square statistic is recomputed, in order to get a sense for what range of values for χ^2 to expect if the model were true. The test is helpful in finding problems with the detection component, but has little success in finding violations of the occupancy component.

The test can be implemented using the **R** package *AICmodavg* in the *mb.gof.test* function, which allows for fitted unmarked occupancy-detection model objects. A faster version of the code can be found in the **PRESENCE** software, which can be accessed from the **RPresence** package (MacKenzie and Hines, 2015, version 1.1.2 or later). It is downloadable from <http://www.mbr-pwrc.usgs.gov/software/presence.html> (under Downloads, install both **PRESENCE** and **RPresence**). If a

Bayesian model is used, it also has a chi-square test using a posterior distribution and is referred to as the Bayesian P-value (e.g., Royle et al., 2007b). The authors discussed the need for aggregating data in some situations.

Warton et al. (2017) applied Dunn-Smyth residuals (Dunn & Smyth, 1996) to detection and occupancy residuals using appropriate cumulative distribution functions. Detection residuals are computed from the n_s as violations are more likely to affect these. If the detection probability parameter is constant across repeat surveys and equal to p_s for site s , then the total number of detections across T surveys would be distributed as a truncated binomial with parameters T and p_s (zero excluded), so its cumulative distribution function is

$$F(n_s \mid n_s > 0) = \frac{1}{1 - (1 - p_s)^T} \sum_{k=1}^{n_s} \binom{T}{k} p_s^k (1 - p_s)^{T-k}.$$

If p_s is not constant, then

$$F(n_s \mid n_s > 0) = \frac{1}{1 - \prod_{t=1}^T (1 - p_{st})} \sum_{y \mid 1 \leq y_{st} \leq n_s} \prod_{t=1}^T p_{st}^{y_{st}} (1 - p_{st})^{1-y_{st}},$$

where $y_{st} = \sum_{t=1}^T y_{st}$ is the total number of detections in detection history \mathbf{y} .

Occupancy residuals require the indicator variable $j_s = I(n_s > 0)$ for whether or not any detections are observed at a site. Then

$$F(j_s) = \begin{cases} 1 - \psi_s + \psi_s \prod_{t=1}^T (1 - p_{st}) & \text{if } j_s = 0, \\ 1 & \text{if } j_s = 1. \end{cases}$$

This expression holds if we set $p_{st} = p_s$, giving the product term $(1 - p_s)^T$.

For an omnibus residual incorporating both aspects of detection and occupancy, we have

$$F(n_s) = \begin{cases} 1 - \psi_s + \psi_s \prod_{t=1}^T (1 - p_{st}), & n_s = 0, \\ (1 - \psi_s) + \psi_s \sum_{y \mid 0 \leq y_{st} \leq n_s} \prod_{t=1}^T p_{st}^{y_{st}} (1 - p_{st})^{(1-y_{st})}, & n_s = 1, \dots, T. \end{cases}$$

The general method for finding a residual z for a discrete random variable x , say, is to find z to satisfy

$$\Phi(z) = (1 - u)F(x) + uF(x),$$

where $\Phi(\cdot)$ and $F(\cdot)$ are the cumulative distribution functions of a standard normal variable and of x (respectively), $F_-(x)$ is the previous value of $F(\cdot)$, and u is a value randomly generated from the standard uniform distribution. Here, $F(x)$ can be one of the cumulative distribution functions described above. If the model for x is correct, these residuals will have a standard normal distribution. They can therefore

be used just like residuals from linear regression and, for example, can be plotted against predictors or fitted values to check for no trend. They can be compared to the standard normal using a (standard) normal quantile plot. Because of the discreteness of the data (possibly obscuring patterns), the residuals can be smoothed with 95% confidence bands around the smoothed residuals and use the smoother as a basis for diagnosing lack of fit rather than the residuals themselves. (The authors used the `gam` function from the **MGCV** package (Wood, 2011).) Further details and computer code are given in the “Supporting Information” of Warton et al. (2017).

Wright et al. (2019) considered whether residuals can separately assess detection and presence when the two components are intimately connected. They used the same model as that of MacKenzie and Bailey (2004) and Warton et al. (2017) above, but with $t = 1, \dots, T_s$ for site s . Again, it was assumed that there were no false-positive errors and heterogeneity could be accounted for by assuming

$$\text{logit}(\psi_i) = \mathbf{x}'_i \boldsymbol{\beta}, \text{ and possibly } \text{logit}(p_{ij}) = \mathbf{v}'_{ij} \boldsymbol{\alpha},$$

where \mathbf{x}_i provides site-level covariates and \mathbf{v}_{ij} provides both site-level and visit-level covariates. These covariates can be used for sites not visited. Residuals are needed for both parts of the model, namely, the z_s (presence) and the y_{st} (detection). If $n_s = \sum_{t=1}^{T_s} y_{st}$, then the posterior binary distribution $\Pr[z_s = 1 | \mathbf{y}_s]$ is 1 for $n_s > 0$ and the following probability from Bayes’ theorem and $\Pr(\mathbf{y}_s)$ above when $n_s = 0$, namely,

$$\frac{\psi_s \prod_{t=1}^{T_s} (1 - p_{st})}{(1 - \psi_s) + \psi_s \prod_{t=1}^{T_s} (1 - p_{st})}.$$

Sampling from the posterior distribution of the occupancy states are generated as random draws from Bernoulli distribution with probabilities just described at each MCMC iteration. After convergence, the collection of draws approximates the posterior distribution of the latent occupancy states and can be used to make inference about other quantities of interest.

Getting now to the residuals, occupancy residuals are $o_s^{(k)} = z_s^{(k)} - \psi_s^{(k)}$, where $z_s^{(k)}$ is the k th draw from the posterior distribution of the occupancy state at site s . Similarly, the detection residuals are

$$d_{st} | z_s^{(k)} = 1] = y_{st}^{(k)} - p_{st}^{(k)},$$

which are only defined for “occupied” sites based on the latent state $z_s^{(k)}$ for a given posterior draw k . By conditioning on a posterior draw of the occupancy state $z_s^{(k)}$, sites without any detections ($\mathbf{y}_s = \mathbf{0}$) still contribute detection residuals for some posterior draws because these sites might actually be occupied ($z_s^{(k)} = 1$). The use of such residuals, such as in investigating unaccounted spatial structure, is discussed further by the authors.

4.2.5 Detectability, Covariates, and Detection-Only Data

In this section, we look more closely at the role of covariates in modeling detectability, which are used extensively in the literature, and also sound a note of warning. For many species, the available data consists of just “presence” data (locations where the species or evidence of it has been observed), together with environmental data. Phillips and Elith (2013) noted that “Recently published papers disagree on whether probability of presence is identifiable from such presence-background data alone.” Using simulation, they showed that additional information is required. They referred to two groups of methods: those based on a strong assumption that species’ true probability of presence exactly matches a given parametric form that had highly variable predictions, but with much larger RMS error than methods that take population prevalence (the fraction of sites in which the species is present) as an additional parameter. From their experiment, they recommended against the methods using the strong assumption as it is unlikely to be true in practice, and an estimate of prevalence is necessary (except in special cases) for identifying probability of presence.

Hastie and Fithian (2013) considered this problem, and we follow their discussion and notation. Let \mathcal{D} denote a domain or region with graphical sites denoted by x , and at each site, $\mathbf{z} = \mathbf{z}(x)$ records the values of some geographical attributes or features. Let the binary variable y denote presence (1) or absence (0) at sites. Define $\pi(\mathbf{z})$ to be the marginal density function of \mathbf{z} , $\pi_1(\mathbf{z})$ to be the (conditional) density at presence sites, and $\pi_0(\mathbf{z})$ to be the density for absence sites. If the overall presence occurrence probability is $\psi(y = 1)$ and hence absence $\psi(y = 0) = 1 - \psi(y = 1)$, then we have the conditional occurrence probability at a site, given we observe feature \mathbf{z} , which is given by

$$\psi(y = 1 | \mathbf{z}) = \frac{\psi(y = 1)\pi_1(\mathbf{z})}{\pi(\mathbf{z})}.$$

Also,

$$\pi(\mathbf{z}) = \pi_1(\mathbf{z})\psi(y = 1) + \pi_0(\mathbf{z})(1 - \psi(y = 1)).$$

Presence-background data consists of a random sample of values of \mathbf{z} from $\pi_1(\mathbf{z})$, as well as a separate sample from $\pi(\mathbf{z})$ (possibly the entire background distribution), which directly inform us about the densities π_1 and π . However, even if both of these distributions were fully known, we have a problem as we do not know π_0 in the second equation above and therefore the occurrence probability $\psi(y = 1)$ is unknown. We do not have data to estimate it, so it has to be modeled in some way, and this carries on to $\psi(y = 1 | \mathbf{z})$ from the first equation above. If the model

$$\text{logit}[\psi(y = 1) | x, \beta] = \beta_0 + x\beta_1$$

is used, the problem gets passed from $\psi(y = 1)$ to β_0 . The authors then showed that the model by Royle et al. (2012) has an identifiability problem.

In conclusion, we mention a commonly used computer package called **MAX-ENT** (cf. Phillips & Dudík, 2008; Elith et al., 2011; and Yackulic et al., 2013). This package is now becoming a standard approach for presence-only data analysis and is mentioned further below. It is a general-purpose method for making predictions or inferences from incomplete information, and its origins lie in statistical mechanics (Jaynes, 1957). It is an algorithmic tool that produces a probability density surface which maximizes entropy, given constraints that are imposed by the collection of vectors of environmental variable values at the sites at which the species has been observed. It does have some shortcomings (Royle et al., 2012).

4.2.6 *Plants and Variable Plot Sizes*

Ståhl et al. (2020) discussed presence-absence sampling for estimating plant density using survey data with variable plot size. Assuming plant locations follow a Poisson model with mean plant density λ , the probability p that at least one plant will occur on a plot with area a is one minus the probability of zero plants, namely, $p = 1 - \exp[-a\lambda]$. Here, p was estimated by $\hat{p} = m/n$, where m out of n plots have plants present, leading to the estimate $\hat{\lambda} = \log(1 - \hat{p})/a$, with large sample variance $p/[na^2(1 - p)]$. Other modifications were mentioned as well as a generalized log-linear model. A model for two consecutive visits to a plot was also developed.

4.3 Some General Models

Initially, there is the practical issue of the sampling method to be used, namely, sampling with or without replacement. In general, sampling with replacement is usually the only feasible method in occupancy studies as individuals are not usually uniquely identified. Kendall and White (2009) noted that some care is needed with the sampling. The key assumption is that a species is available for detection during each search of an occupied site. The species may be present but not available for detection. Some sites are so large that some kind of subsampling is needed. The authors demonstrated that when multiple samples per site are defined by searching different locations within a site, absence of the species from a subset of these spatial subunits induces estimation bias when locations are exhaustively assessed or sampled without replacement. Also, this bias can be removed by choosing sampling locations with replacement or if the species is highly mobile over a short period of time. They concluded that: “Where detection effort within a site is allocated across space rather than time, exhaustive assessment or sampling without replacement is vulnerable to positive bias in occupancy estimators when some individual sampling locations are unoccupied at the time of sampling.”

Guillera-Arroita (2011) also discussed this problem and said that the presence of bias in the estimator of site occupancy will depend on how spatial subunits are sampled. He found that sampling without replacement does not induce bias in the estimator of site occupancy when each spatial subunit has a constant probability of occupancy, regardless of the occupancy status of other subunits within the same sampling site. However, sampling with replacement may in this case induce bias in the site occupancy estimator. We now consider some models.

We begin with the basic model following MacKenzie et al. (2002), where there are S specific sites and a survey at site s is performed at T specific times with n_t the number of sites where the species is detected at time t ($t = 1, \dots, T$). Let ψ_s be the probability that site s is occupied by an individual of the particular species. Given that the species is present at the site, let p_{st} be the conditional probability of detecting an individual of the species at site s in time t . For example, if the detection history for site i is $\mathbf{x}_i = (100110)$, the species is detected in only times 1, 4, and 5 for the $T = 6$ visits. Then the probability of this, given the individual is detected in at least one survey, is

$$\psi_i p_{s1}(1 - p_{s2})(1 - p_{s3})p_{s4}p_{s5}(1 - p_{s6}).$$

If the species was not detected at all on site s , that is, with detection history (000000), then it was either present but not detected with probability $\psi_s \prod_{t=1}^6 (1 - p_{st})^6$ or not present with probability $(1 - \psi_s)$. Adding these gives us

$$\Pr(\text{Not detected at site } s) = \psi_s \prod_{t=1}^6 (1 - p_{st}) + 1 - \psi_s.$$

Assuming independence of the sites, the product of all the terms (one for each site) constructed in this manner gives us a likelihood function. As it stands, the above model has too many parameters, and the number needs to be reduced by adding constraints.

For example, one method assumes constancy across sites so that $p_{st} = p_t$ and $\psi_s = \psi$ giving the likelihood

$$L(\psi, \mathbf{p}) = \left[\psi^{n_s} \prod_{t=1}^T p_t^{n_t} (1 - p_t)^{n_t - n_t} \right] \times \left[\psi \prod_{t=1}^T (1 - p_t) + (1 - \psi) \right]^{N_s - n_s}.$$

We note in passing that if ψ_c is the probability that a species is present at a particular site, given not detected, then

$$\begin{aligned} \psi_c &= \Pr(\text{species present} \mid \text{species not detected}) \\ &= \frac{\Pr(\text{species present and not detected})}{\Pr(\text{species not detected})} \\ &= \frac{\psi \prod_{t=1}^T (1 - p_t)}{\psi \prod_{t=1}^T (1 - p_t) + (1 - \psi)}. \end{aligned}$$

Another method that reduces the number of parameters uses covariates and a log-linear or logistic model, where ψ_s may be some function of site characteristics such as habitat type or patch size, while the p parameters may be functions of weather conditions. For example, with v covariates, we can use

$$\text{logit}(\psi_s) = \beta_0 + \beta_1 x_{s1} + \dots + \beta_v x_{sv}.$$

Weir et al. (2005) applied the above approach to anuran (frog and toad) detection and site occupancy.

Because ψ_s does not change over time as the population is closed, appropriate covariates (auxiliary variables) would be time constant and site specific, whereas covariates for detection probabilities could be time varying and site specific (such as air or water temperature). This method is very similar to a capture-recapture model (Chap. 12) where sites are analogous to “individuals” with the difference that both the number of individuals and the number of sites at which the species is never detected over T sampling occasions are both known. Also, the focus is different, which is to estimate the fraction of those sites that the species actually occupies, and we then have the additional parameters ψ_s . They applied the methods to anuran species.

Several models are possible depending on the sampling program. For example, Nielson et al. (2011), in considering aquatic plants and using stratified sampling, suggested setting $p_{st} = p_s$, letting T_s be the number of surveys conducted at site s and y_s be the number of these where the species is detected. Then the detection history at site s , given the species is detected in one or more surveys, is modeled as

$$\psi_s p_i^{y_s} (1 - p_s)^{T_s - y_s}.$$

The probability that the species is not detected at site s is

$$\psi_s (1 - p_s)^{T_s} + (1 - \psi_s).$$

It was assumed that $\psi_s = \psi$ and $p_s = p$, but possible variation in the parameters was included using logistic regression with covariates and possibly random effects.

Goijman et al. (2015) used a multi-season regional analysis from 2003 to 2012 of multispecies occupancy in Argentina using transects and a Bayesian hierarchical model. They used data from large-scale bird monitoring program carried out in agroecosystems of central Argentina, in portions of the Pampas and Espinal ecoregions. They evaluated the effect of land use and land cover extent on 263 bird species and tested predictions that species diversity is positively related to habitat heterogeneity. The usual model was applied. We have for the process model $z_{ist} = 1$ with probability ψ_{ist} if species i is present at site j and year t and 0 otherwise, given present. For the observation model, $y_{iskt} = 1$ with probability p_{iskt} if the species i at site s is detected from point k on the transect at the center of an observation circle of 200 m radius on year t , given present, and 0 otherwise. We

then have

$$z_{ist} \sim \text{Bernoulli}(\psi_{ist}) \text{ and } y_{iskt} \sim \text{Benoulli}(\psi_{ijt} p_{iskt}).$$

$\text{Logit}(\psi_{ist})$ and $\text{logit}(p_{iskt})$ were expressed as linear in covariates, and those covariates that were significantly correlated were discarded. Some independent uninformative prior distributions were used for some of the parameters followed by applying Markov chain Monte Carlo to posterior distributions.

Track Surveys with Correlation

Aing et al. (2011) developed a Bayesian hierarchical occupancy model for track surveys conducted in a series of linear, spatially correlated, sites and applied it to otter data, as follows.

Let $y_{sjk\ell} = 1$ if the ℓ th observer records a track in site s ($s = 1, \dots, S$) on the t th day of the j th snowfall event and 0 otherwise. Let $z_s = 1$ if the site is occupied with probability ψ_s , where a site is considered to be occupied if it falls within the home range of at least one otter. Survey periods consist of the first t days after a snowfall event, but prior to a subsequent snowfall event (or other event that might erase tracks, e.g., wind or melting). In the application, otter tracks when present were readily observable from the air. However, observations made by different observers on the same day were often inconsistent, and it was not uncommon for observers to record a track one day and not record a track a day later (within the same site and survey period). In addition to potential problems related to locating and correctly identifying otter signs, it was suspected that tracks observed near boundaries of the plots were occasionally assigned to the wrong unit (false positives). The following assumptions were made:

- (1) Site-level occupancies, (z_1, \dots, z_S) , do not change during the course of the study.
- (2) Conditional on $z_s = 1$, a track is laid down with probability θ_s on any given day following a snowfall event.
- (3) Tracks remain present until the end of the survey period.
- (4) The probability of detecting a track is the same recorded on the first and third days and not falsely recorded on the second day (i.e., detection probabilities do not depend on how many tracks are laid down).
- (5) Conditional on the presence or absence of a track, correct detection and false detection probabilities (p_ℓ and e_ℓ , respectively) vary by observer ℓ , but they are constant across sites and sampling instances. For a single observer, we drop the subscript ℓ .
- (6) Conditional on true presence or absence of a track, observations made by different observers are independent.
- (7) Conditional on true presence or absence of a track, observations made by the same observer (on different sampling dates) are independent. In other words, observers are not influenced by having flown the river on a prior date.

To construct probabilities for various observations, consider one site s surveyed by a single observer during each of the first 3 days of a survey period. Suppose the observer records a track the first and third days but no track the second day. We denote this outcome as (101). Then, the five possible probabilities for site s are:

$$\psi_s \theta_s p(1-p)p = \Pr(\text{site occupied, track laid day 1, and correctly recorded days 1 and 3, and not recorded on day 2}).$$

$$\psi_s(1-\theta_s)\theta_s e(1-p)p = \Pr(\text{site occupied, track first laid on day 2 and falsely recorded on day 1, not recorded on day 2, and correctly recorded on day 3}).$$

$$\psi_s(1-\theta_s)^2\theta_s e(1-e)p = \Pr(\text{site occupied, track first laid on day 3 and falsely recorded on day 1, not falsely recorded on day 2, and correctly recorded on day 3}).$$

$$\psi_s(1-\theta_s)^3e(1-e)e = \Pr(\text{site occupied, no track laid, and falsely recorded on days 1 and 3, and not falsely recorded on day 2}).$$

$$(1 - \psi_s)e(1 - e)e = \Pr(\text{site not occupied, track falsely recorded on days 1 and 3, and not falsely recorded on day 2}).$$

To extend this approach to additional observers and survey periods requires that one consider all surveys at a particular site simultaneously; although track-laying processes are assumed to be independent among survey periods, occupancy status induces correlation among observations from the same site. To construct the likelihood, let $x_{sjk} = 1$ if a track is laid in site s on the k th day of j th snowfall event and 0 otherwise. From assumption (2), we have $x_{sjk} | z_s \sim \text{Bernoulli}(\theta_s)$. Let $w_{sjk} = 1$ if a track is present in site s on the k th day of the j th snowfall event and 0 otherwise. The w 's are obtained directly from the x 's, namely, $w_{sj1} = x_{sj1}$, and for $k > 1$, $w_{sjk} = 1$ if $w_{sj,k-1} = 1$ and x_{sjk} otherwise. Finally, from assumptions (4–7), we have that $y_{sjk} | w_{sjk}$ are mutually independent and

$$y_{sjk\ell} | w_{sjk} \sim \text{Bernoulli}(w_{sjk}p_\ell + (1 - w_{sjk})e_\ell).$$

We thus have three levels of parameters describing (i) occupancy ($\{\psi_s\}$); (ii) availability of tracks for detection (due to movement), conditional on occupancy ($\{\theta_s\}$); and (iii) recorded presence-absence of tracks, conditional on their availability for detection ($\{e_\ell, p_\ell\}$). Spatial dependence was introduced by assuming $\text{logit}(\psi_s) = \alpha_0 + \alpha_s$ by defining contiguous sites as neighbors and placing an intrinsic conditional autoregressive (CAR) prior distribution on the α_s 's. Rather than determine the marginal likelihood for y_{sjk} by integrating over the latent variables, the authors used

Markov chain Monte Carlo to numerically integrate over the latent variables using a Bayesian formulation of the problem.

Aftereffects of a Wildfire

Chelgren et al. (2011), in studying the aftereffects of a wildfire, used generalized linear mixed models (GLMM) with fixed and random effects and multilevel spatial structures to compare burned and unburned areas. They also made use of subplots and began with the standard survey of S sites and T samples per site and for site s the parameters ψ_s for the detection probability and p_{st} ($t = 1, \dots, T$) for the detection probabilities, given present. Here, $y_{st} = 1$ if the species is detected at site s and 0 if not detected at time t ($t = 1, \dots, T$). There are two ways that $y_{st} = 0$ for all t and site s , namely, either site s was unoccupied, or it was occupied and all T search attempts failed to detect the species even though it was present. Hence, the likelihood for site s is

$$L(\psi_s, \{p_{st}\} | \{y_{st}\}) = (1 - \psi_s)w(s) + \psi_s \prod_{t=1}^T p_{st}^{y_{st}} (1 - p_{st})^{1-y_{st}},$$

where $w(s) = 1$ if $y_{st} = 0$ for all t and 0 otherwise. It was assumed that sites are closed to changes in occupancy over the course of the study and there were no false detections. Including a subscript for species, they then modeled the ψ and p parameters using logistic models with extensive covariates, where some of the covariate regression parameters have prior distributions. They also described two goodness-of-fit tests. Details are available from their paper. An earlier paper by Russell et al. (2009) investigated avian responses to prescribed fire using a Bayesian hierarchical model of multispecies and logistic regression.

4.3.1 Using the Robust Model

MacKenzie et al. (2003) expressed the previous type of model in matrix form including allowing for time variation, missing observations, and covariates. Here, the experiment is divided up into S sites with T primary periods, and within each primary period, there are multiple surveys (secondary “periods”). This approach uses the idea of the robust design of Pollock (1982), where the population is closed within each primary period but possibly open between primary periods. It extends the model of MacKenzie et al. (2002). It is discussed further below under the topic of “Closure Assumption.”

Within each primary period t ($t = 1, \dots, T$), detections were made with k_t surveys at a site with no false detections when absent, with each site having a T vector of 1’s and 0’s indicating the usual presence and absence in period t . The detection history for the k_t surveys of site s and period t , with probability of presence ψ_t at primary period t , is denoted by \mathbf{X}_{st} and over all primary periods as \mathbf{x}_s for

site s . We now see how detection histories for the S sites can be used to estimate site occupancy, colonization, and local extinction probabilities.

The following parameters are independent of the site s . Let γ_t be the probability that a site unoccupied at primary sampling period t is occupied by the species at period $t + 1$ (a form of “local colonization”), and let ε_t be the probability that a site occupied by the species at t is unoccupied at $t + 1$ (a form of “local extinction”). Let p_{tj} be the probability of detection, given present, for survey j ($j = 1, \dots, k_t$) within primary period t . For example, for two primary periods each with three secondary periods, we can consider as an example an overall detection history for site s as (001, 000). Here, the species was detected at the site in the first primary period ($t = 1$) during the third survey only and was never detected in the secondary period ($t = 2$). Then, for period 1 and $k_1 = 3$ surveys,

$$\Pr(\mathbf{x}_{s1} = 001) = \psi_1(1 - p_{11})(1 - p_{12})p_{13}.$$

Following this, there are two possibilities that would result in the species not being detected at $t = 2$: either it continued to occupy the site and was undetected, or it became locally extinct. The probability of this occurring would be expressed as

$$\Pr(\mathbf{x}_{s2} | \mathbf{x}_{s1}) = (1 - \varepsilon_1) \prod_{j=1}^3 (1 - p_{2j}) + \varepsilon_1.$$

The probability of observing the full sequence is then given by

$$\Pr(\mathbf{X}_s = 001, 000) = \Pr(\mathbf{x}_{s1})\Pr(\mathbf{x}_{s2}).$$

Using several examples, the authors showed how to use matrices and vectors to express the model more elegantly.

The overall likelihood takes the form

$$L(\psi_1, \{\varepsilon_t\}, \{\gamma_t\}, \{p_{jt}\} | \{\mathbf{X}_s\}) = \prod_{s=1}^S \Pr(\mathbf{X}_s).$$

As already noted, an important assumption of the model is that all parameters are constant across sites at any given time. The probability of occupancy at period t , ψ_t , can be calculated recursively from the relationship

$$\psi_t = \psi_{t-1}(1 - \varepsilon_{t-1}) + (1 - \psi_{t-1})\gamma_{t-1}.$$

The authors suggested some reparameterizations to highlight certain parameters, as well as considering missing observations and covariates. In addition to using simulation, they applied the model to northern spotted owls and tiger salamanders.

Open Population

The good feature of the robust model referred to above is that it can be used for populations that are open between primary periods provided it is closed within a primary period. This means that incomplete detectability can be incorporated using secondary samples as repeat counts, as demonstrated by Kéry et al. (2009) in the following development, and using the previous notation. Let c_{sjt} be the number of counts from site s ($s = 1, \dots, S$), survey j ($j = 1, \dots, J$), and in year (primary period) t , ($t = 1, \dots, T$), where the population at each site is assumed to be demographically closed each year during the sampling period. Let N_{st} be the size of the population size at site s , year t , and these are assumed to have independent Poisson distributions with parameters λ_{st} . As usual, we then have two processes:

$$\text{State process : } N_{st} \mid \lambda_{st} \sim \text{Poisson}(\lambda_{st}),$$

$$\text{Observation process : } c_{sjt} \mid N_{st} \sim \text{Binomial}(N_{st}, p_{sjt}),$$

where p_{sjt} is the probability of detection at site s , survey j , and year t . To incorporate trend in population sizes, the authors used the models

$$\log(\lambda_{jt}) = \alpha_s + r_s(t - 1)$$

$$\text{logit}(p_{sjt}) = \alpha_{sjt} + \beta_w x_{sjtw},$$

where r_s is the annual population growth rate for site s and x_{sjtw} is w th covariate with value x_{sjtw} . Random effects can also be added to the above models. Bayesian methods were applied with suitable “vague,” independent priors for all parameters. The method was applied to the Dutch sand lizard *Lacerta agilis* from 1994 to 2005 using 208 sites and the computer package **BUGS**, with some code in an online appendix. Some aspects of study design were also mentioned.

4.3.2 Spatial Models

Although spatial models are considered in more detail in Chap. 14, we first introduce them here as more recent occupancy models are including spatial and temporal components. Also, the role of autoregression is mentioned here. They can also be applied to animal signs (e.g., snow tracks, Whittington et al., 2015; tracks on plates, with skunks, Hackett et al., 2007). We begin with an important data set from the North American Breeding Bird Survey (BBS), a survey that has been conducted for more than 45 years across much of North America. With such a long period, population trends can also be investigated, and Bled et al. (2013) developed spatially explicit models of temporal population change for selected species of birds to assess bird population variations and develop predictive maps for bird population trends. By overlaying a grid on the BBS survey area using degree blocks of latitude and

longitude as their cells, they estimated the trend for a theoretical route contained within each individual grid cell.

The spatial effect was modeled via a Gaussian CAR model (Besag et al., 1991). This can be thought of as a spatial analog of an autoregression model in time series, in the sense that the model is described by relating values of the state process to neighboring values (in the case of time-series, preceding values). The observed count, y_{st} , on route s at year t is assumed to have a Poisson distribution with mean λ_{st} , which depends on a year-specific intercept, an observer effect, and a spatial effect at the level of the cell encompassing the corresponding route, namely,

$$y_{st} \sim \text{Poisson}(\lambda_{st}) \quad \text{with } \log(\lambda_{st}) = \alpha_t + \beta_{c(s),t} + w_{K(s,t)}.$$

Here, α_t is the year-specific intercept, and $w_{K(s,t)}$ is the observer effect for the observer K on the route s during year t . This is because observed patterns of change in counts do not only reflect changes in population sizes, but also changes in the pool of observers (or other factors affecting detection). The parameter $\beta_{c(s),t}$ is the spatial effect for the cell c that encompasses the route s in year t . Both α_t and $w_{K(s,t)}$ are given respective prior normal distributions $N(0, 0.01)$ and $N(0, \sigma^2)$, where σ has a uniform distribution on $[0, 10]$. The spatially correlated random effect $\beta_{c(s),t}$ is expressed as a CAR model where the spatial effect of the cell c is based on the grid cells that share a common boundary with cell c .

If n_{ct} is the expected abundance for a theoretical route in cell c and year t , then the trend is defined as an interval-specific geometric mean of yearly changes in population size, expressed as a percentage. Hence, the trend from year t_a to year t_b for cell c is

$$\Delta_{c,ab} = 100 \left[\left(\frac{n_{c,t_b}}{n_{c,t_a}} \right)^{1/(t_b-t_a)} - 1 \right] \text{ with } \log(n_{ct}) = \alpha_t + \beta_{ct} + \sigma^2/2.$$

Bled et al. (2013) used a Bayesian analysis to obtain posterior samples of n_{ct} for each cell and each year, from which they computed the trends of interest. They compared their method with the previous BBS methods for three case studies and used **WinBUGS** for the computations.

Such large-scale studies need careful planning of sites and routes (Sauer et al., 2003). Johnson et al. (2013) considered spatial occupancy models for large data sets and explicitly accounted for spatial autocorrelation and introduced the restricted spatial regression (RSR) model. They assumed there were no false positives, that is, an individual was regarded as present when it wasn't, or misspecifying the species. Let $z_s = 1$ if site s ($s = 1, \dots, S$) is occupied with probability ψ_s and 0 otherwise. Given an individual is present, let $y_{st} = 1$ with probability p_{st} if the individual is detected, at site s and time t , and 0 otherwise. We have the usual equations

$$y_{st} | z_s, p_{st} \sim \text{Bernoulli}(z_s p_{st}) \text{ and } z_s | \psi_s \sim \text{Bernoulli}(\psi_s).$$

For computational reasons, namely, increased computational efficiency and greater flexibility through a data augmentation approach, the authors used the probit link $\Phi^{-1}(\psi_s) = \mathbf{x}'_s \boldsymbol{\gamma}$, where \mathbf{x}_s are covariate values, Φ is the cumulative distribution of a standard normal distribution, and Φ^{-1} is the quantile function. Let \tilde{z}_s be a continuous latent version of the binary occupancy process z_s , and let $\tilde{z}_s | \boldsymbol{\gamma} \sim N(\mathbf{x}'_s \boldsymbol{\gamma}, 1)$. Let z_s be the indicator of $\tilde{z}_s > 0$ so that

$$z_s | \boldsymbol{\gamma} \sim \text{Bernoulli}(\Phi(\mathbf{x}'_s \boldsymbol{\gamma})).$$

Analogous to \tilde{z}_s , let \tilde{y}_{st} be a continuous version of y_{st} . Because $y_{st} = 0$ with probability 1 if $z_s = 0$, then

$$\tilde{y}_{st} | z_s, \boldsymbol{\beta} \sim z_s N(\mathbf{x}'_{st} \boldsymbol{\beta}, 1) + (1 - z_s) \delta_0$$

with δ_0 a degenerate distribution with point mass 1 at zero. Defining y_{st} to be the indicator of $\tilde{y}_{st} > 0$, it follows that

$$y_{st} | z_s, \boldsymbol{\beta} \sim \text{Bernoulli}(z_s \Phi(\mathbf{x}'_{st} \boldsymbol{\beta})).$$

The authors then discussed spatial regression models in more detail including a conditionally autoregressive CAR model and a restricted spatial regression. Bayesian inference was also used. The method was applied to aerial censusing of caribou.

Gardner et al. (2010) introduced a coarse-scale model to investigate wolverine occupancy based on their tracks surveyed in Alaska using aerial surveys. Because wolverines naturally occur at low densities, are difficult to see, and range widely, monitoring wolverines over large, remote areas the size of Interior Alaska, USA, is a challenging task for managers.

Let $y_{sj} = 1$ if there is an observed occurrence of a track given present for sample unit s ($s = 1, \dots, S$) and transect j (one flight transect across the sample unit; $j = 1, \dots, n_s$) and 0 otherwise. Let $x_s = 1$ for the true presence and 0 otherwise so that

$$\Pr(\{y_{sj}\} | \{x_s\}) = \prod_{s=1}^S \prod_{j=1}^{n_s} (x_s \theta_{sj})^{y_{sj}} (1 - x_s \theta_{sj})^{1-y_{sj}},$$

where p_{sj} is the probability of sighting the track, given present. It was assumed that

$$\text{logit}(p_{sj} = \gamma_0 + \gamma_1 d_{sj}),$$

where d_{sj} is the date (no. of days from 20 Feb.). The $\{x_s\}$ was viewed as a spatial hierarchical model allowing for autocorrelation, and it was assumed that

$$x_s | w_s, h_s \sim \text{Bernoulli}(\xi_s).$$

Also,

$$\text{logit}(\xi_s) = \alpha_0 + \alpha_1 v_s + \alpha_2 \phi_s + \sigma_w w_s,$$

where v_s is the percentage of the sample unit ≤ 305 m elevation and computed for all sample units, including unsampled ones, ϕ_s is the log-mean of a Poisson regression model of human influences (e.g., roads-major trails, buildings, mines) observed within each sample unit, and h_s is the human influence. Also, w_s is a spatially autocorrelated random effect. Accounts for h_s for unsampled units were not available for unsampled units, but predicted using a spatially autocorrelated model for unsampled units, namely,

$$h_s | u_s \sim \text{Poisson}(\exp(\phi_s)),$$

where $\phi_s = \beta + u_s$. It was assumed that both $\{w_s\}$ and $\{u_s\}$ have a conditionally autoregressive (CAR) model and the use of a CAR model as a random effect, namely,

$$w_s | \mathbf{w}_{-s} \sim N(\mu_s, \tau_s^2),$$

where \mathbf{w}_{-s} is the set of w 's except w_s . Also,

$$\mu_s = \frac{\beta_w}{|B_s|} \sum_{j \in B_s} w_j, \text{ and } \tau_s^2 = \frac{1}{|B_s|},$$

where τ_s is the inverse of the standard deviation, with B_s representing the set of neighbors of the s th sampling unit and $|B_s|$ the number of neighbors. If any two sample units shared a border, they were defined as neighbors so that an interior sample had six neighbors. The $\{w_s\}$ and $\{u_s\}$ were assumed to be independent and both with multivariate normal distributions. Prior distributions were put on all the parameters, and posterior distributions were obtained so that Markov chain Monte Carlo methods could be used.

Broms et al. (2014) modeled spatial autocorrelation using an intrinsic conditional autoregressive (ICAR) model and also with the restricted spatial regression (RSR) model. Both spatial models could be readily applied to any other gridded, presence-absence data set using a newly introduced R package. The RSR model provided the best inference and was able to capture small-scale variation that the other models did not. Spatial occupancy models applied to atlas data showed southern ground hornbills strongly depend on protected areas.

Autologistic and related automodels have been implemented approximately as autocovariate regression and can provide simple and direct modeling of spatial population processes with spatial autocorrelation. The autologistic model extends the logistic model for presence-absence data to allow dependence between nearby observations. However, Bardos et al. (2015) raised some concerns about studies

employing neighborhood weighting schemes inconsistent with automodel definitions. Correction methods were described.

4.3.3 Penalized Likelihood

The above models depend heavily on maximum likelihood estimation. A variation on this is so-called penalized likelihood described by Moreno and Lele (2010) and Hutchinson et al. (2015). As above, p_{st} is the probability of detecting the species on visit t ($t = 1, \dots, T$), given that site s is occupied, and ψ_s is the probability of site occupancy at site s . Let $z_s = 1$ if site s is occupied and 0 when the site is not occupied and $y_{st} = 1$ if the species is observed at t and site s and 0 otherwise. Using covariates x and w denoting the occupancy and detection covariates, respectively, we have the following probability distributions:

$$\begin{aligned} z_s &\sim \text{Bernoulli}(\psi_s), \text{ with } \text{logit}(\psi_s) = \alpha_0 + \alpha_1 x_s, \\ y_{st} \mid z_s &\sim \text{Bernoulli}(z_s p_{st}), \text{ with } \text{logit}(p_{st}) = \beta_0 + \beta_1 w_{st}. \end{aligned}$$

Other covariates can be added. The parameters are given by $\boldsymbol{\theta} = (\alpha_0, \alpha_1, \beta_0, \beta_1)'$, and likelihood function is

$$L(\boldsymbol{\theta}) = \prod_{s=1}^S \left\{ \psi_s \prod_{t=1}^T [p_{st}^{y_{st}} (1 - p_{st})^{1-y_{st}}] + (1 - \psi_s) I(y_{i\cdot} = 0) \right\}, \quad (4.1)$$

where the indicator function $I(\cdot)$ is 1 if the statement in brackets is true and 0 otherwise and $y_{i\cdot} = \sum_{t=1}^T y_{st}$. A correction shrinkage term λ or “penalty” is added to the log-likelihood before maximization, where λ is a function of the number of visits, the number of sites, and initial estimates of certain parameters. The penalty helped to stabilize the computations.

Moreno and Lele (2010) proposed a different penalty function λ and used prior distributions for the covariates. They compared the maximum likelihood estimates (MLE) and the penalized estimates (MPLE) using simulation. It was found that for $T = 5$, the estimates were quite similar (because the penalty function was small with large T) and standard errors and confidence intervals based on MPLE were somewhat shorter than for MLE. Overall, it was found that MPLE performed better than MLE.

Hutchinson et al. (2015) proposed three penalized models of the form

$$\log(L(\boldsymbol{\theta})) - \lambda g(\boldsymbol{\phi})$$

to be maximized, where $g(\boldsymbol{\phi})$ is a function of the various parameters $\boldsymbol{\phi}$ from the occupancy and logistic parts of the model. Two methods were proposed for choosing λ : cross-validation and a method called MPLE. Across all of the synthetic data

sets, the penalized estimation methods were found to have lower mean squared error than the maximum likelihood estimates. Problems with maximum likelihood estimation, especially with small data sets, were discussed, and in this case, the penalized method was found to be particularly useful. All three models along with cross-validation were to be included in an update of the *unmarked* R package.

4.3.4 Testing for Occupancy Differences

Assuming there are no false positives and that ψ and p are constant, (4.1) reduces to

$$\begin{aligned} L(\psi, p) &= \prod_{s=1}^S \left\{ \psi p^{d_s} (1-p)^{T-d_s} + (1-\psi) I(d_s = 0) \right\} \\ &= [\psi^{S_D} p^d (1-p)^{TS_D-d}] (1 - \psi p^*)^{S-S_D} \end{aligned}$$

where d_s ($= y_{s.}$) is the number of detections at site i , $d = \sum_s d_s$, S_D is the number of sites where the species was detected at least once, $I(\cdot)$ takes the value 1 when the expression in brackets is true and 0 otherwise, and $p^* = 1 - (1-p)^T$ is the probability of detecting the species in at least one of the T surveys carried out at an occupied site. Guillera-Arroita and Lahoz-Monfort (2012) give the following method for determining S and T to test for a species difference in occupancy at a given power of the test.

The maximum likelihood estimates are given by MacKenzie et al. (2006): 95, and Guillera-Arroita et al. (2010), and satisfy

$$\hat{\psi} = \frac{S_D}{S \hat{p}^*}, \quad \frac{\hat{p}}{\hat{p}^*} = \frac{d}{S_D T}.$$

These estimates are valid when

$$\left(\frac{S - S_D}{S} \right) \geq \left(1 - \frac{d}{ST} \right)^T;$$

otherwise,

$$\hat{\psi} = 1, \quad \hat{p} = d/(ST).$$

Then $\hat{\psi}$ is asymptotically normal with mean ψ and asymptotic variance

$$\sigma^2 = \text{var}[\hat{\psi}] = \frac{\psi}{S} \left\{ (1-\psi) + \frac{1-p^*}{p^* - Tp(1-p)^{T-1}} \right\} = \frac{\psi}{S} [(1-\psi) + F].$$

When $p^* \rightarrow 1$ (perfect detection), the above variance tends to a binomial variance $\psi(1 - \psi)/S$.

If we have two independent data sets with ψ_i ($i = 1, 2$), we can test $H_0 : D = \psi_1 - \psi_2 = 0$ using $\widehat{D} = \widehat{\psi}_1 - \widehat{\psi}_2$, which is approximately normal $N(0, \sigma_1^2 + \sigma_2^2)$ when H_0 is true. This is the same as a Wald test. Guillera-Arroita and Lahoz-Monfort (2012) showed that the size of S to achieve a certain power $1 - \beta$ for a test of H_0 is given by

$$S = (f_1 + f_2) \left(\frac{z_{\alpha/2} + z_\beta}{\psi_1 - \psi_2} \right)^2,$$

where $f_i = \psi_i(1 - \psi_i + F_i)$ and z_ϕ gives the upper ϕ tail of the normal $N(0, 1)$ distribution. Because the above formula is based on approximations, its outcome represents a lower bound. They also give a method to provide an optimum survey design in terms of minimum total survey effort ($E = TS$) for choosing T to achieve a given power.

4.3.5 Surveys After the First Detection?

The question of whether surveys should continue after the first detection was asked by Guillera-Arroita and Lahoz-Monfort (2017). For instance, further records after the first detection do not add any information directly about the occupancy status of sites, so they could seem unnecessary. However, the authors concluded that additional records can help in the estimation of species detectability and, through this, lead to a better estimation of species occupancy. We shall see below that it is sometimes recommended that only the first detection is recorded, which leads to a type of removal model. MacKenzie and Royle (2005) concluded that there was “strong evidence that a removal design is much more efficient than a standard design for estimating occupancy” if the detection probability is constant. However, they noted that they expected the data yielded by removal designs to provide less flexibility for modeling, particularly for exploring potential sources of variation in detection probability, and hence suggested that a removal design could be less robust in general.

Guillera-Arroita and Lahoz-Monfort (2017) compared the mathematical models for the standard and optimal removal designs, assuming initially for simplicity that occupancy (ψ) and detectability (p) are constant. They found that their asymptotic results indicated that the removal design is more efficient than the standard design in most scenarios under optimal conditions. The removal design is worse when occupancy and detection probabilities are small, especially when p and T , the number of surveys per site, are small. It is also more sensitive than the standard design to reductions in the cumulative detectability level $p^* = (1 - (1 - p)^T)$.

The removal design is much better when T and p are large as $p^* \approx 1$ and the removal design avoids “wasting” effort in revisiting sites where the species is already detected. However, the authors’ simulations revealed that the removal design is more severely affected by reductions in survey effort. Also, unmodeled variable detectability degraded the performance of both designs in a similar fashion. They concluded from their simulations that “when spatial heterogeneity in detectability is unaccounted for, the standard design continues outperforming the removal design when cumulative detectability is not close to perfect, but that the gap in performance decreases as the amount and impact of heterogeneity increases.”

Time-to-Detection (TTD) Models

Recording just the time to the first detection, essentially the removal method previously discussed, is referred to as a TTD (time-to-detection) model discussed by Henry et al. (2020). They compared using a single survey TTD model and standard models with a higher number of replicates using simulations and real data. It was found that for inconspicuous species, single survey TTD models can perform better or equally as well as the standard models with a higher number of replicates. Results from the field data showed that a single TTD survey was not adequate to reliably estimate occupancy, detection, and species richness, especially in rare and inconspicuous species. Increasing the number of TTD surveys to two replicates improved the models substantially. They also noted that “A single TTD survey may be sufficient in some study designs but is unlikely to be sufficient in most multispecies field scenarios where communities are made up of species that have a wide range of detection and occupancy probabilities.” However, Halstead et al. (2018) noted that both double-observer and metapopulation occupancy designs are costly because they require multiple observers or multiple visits to sites and supported a single TTD. Of course, given a certain amount of effort, more sites could be surveyed. With regard to species occupancy, they mention that the event of interest is the initial detection of the species. Survey duration is important here, and some knowledge about this is needed in planning an experiment. The authors investigated the effectiveness of the method with respect to amphibians and reptiles.

With regard to plants, only one visit to a site seems most appropriate, and this can be used with a TTD model. Garrard et al. (2008) considered using this model to estimate the amount of survey time required to achieve a pre-specified confidence that a plant species will be detected at a site if it is present. They used a failure-time approach and assumed that the detection times are exponentially distributed and that the probability of a species undetected at time t being discovered in the next time step is constant for a given species in a given survey (by a given observer). Here, there is the hazard rate λ where, in a flora study, the chance of detecting a species is unaffected by the amount of time that has passed. In each new increment of time spent searching, the chance of finding a particular plant species remains the same.

For a given search area, plants do not become more or less detectable within the duration of an individual survey. Based on these assumptions, the mean time to detection ($\mu = 1/\lambda$) may be modeled as a function of site and survey covariates, such as time of year, vegetation density, or observer experience. With incomplete

detectability, we are interested in “true” absences, that is, those actually absent and not those present but not detected (false negatives). If p is the probability of being detected given present, then $p = 1 - e^{-\lambda t}$.

The authors used an exponential-binomial mixture distribution adjusted for true absences giving the likelihood for a single site as follows.

$$L(y | \psi, \lambda) = \begin{cases} \psi e^{-\lambda t} + (1 - \psi), & y > T(\text{non-detection}), \\ \psi \lambda e^{-\lambda t}, & y \leq T, \end{cases}$$

where y is the observed time to detection; ψ is the probability of occupancy of a species at a site; λ is the rate at which a species, undetected at time t , will be detected in the next time step; and T is the duration of the survey. The likelihoods have to be multiplied for each site. Posterior probabilities of presence were obtained, and the required survey effort such that $\text{Pr}(\text{absent} | \text{not detected}) = 0.95$ was considered. In their appendix, they gave code for simulating data and calling up the **WinBUGS** model in **R**. Further work on this topic is given by Garrard et al. (2015).

Halstead et al. (2018) applied the above model to amphibians and reptiles using prior distributions for all the parameters. They found that the method does not work so well with rare species and the consequent low numbers of detections.

Loonam et al. (2021) assessed the robustness of TTD models for estimating unmarked wildlife abundance using remote cameras. They used simulated walk models to test the effects of violating the assumptions of the time-to-event model under four scenarios: (1) incorrectly estimating movement speed, (2) violating closure, (3) individuals moving within simplified territories (i.e., movement restricted to partially overlapping circles), and (4) individuals clustering in preferred habitat. The TTD model was found to be robust to closure violations, territoriality, and clustering when cameras were placed randomly. However, the model failed to estimate abundance accurately when movement speed was incorrectly estimated or cameras were placed nonrandomly with respect to habitat. They considered both the basic model and a second version that adjusts the density estimate for spatial variation in density using habitat covariates. The number of animals in camera viewsheds N was assumed to have a Poisson distribution and the TTD an exponential distribution.

4.4 Key Assumptions

It is helpful to take a breather and relook at the key assumptions using occupancy methods as follows:

- (1) The occupancy state of a site does not change during the sequence of repeated surveys, i.e., the population is closed. This is the most important assumption and can be readily violated if, for example, the species undergoes seasonal changes in habitat use or migrations.

- (2) The sites are independent. This can be achieved if the sites selected are a random selection without replacement from a set of primary sites.
- (3) The detection of the species in each survey of a site is independent of detections during other surveys of the site. This precludes any observer effects or temporal variations due to say the weather. There is however a learning curve associated with an observer described under the heading of a “start-up” effect (Sauer & Link, 2011).
- (4) Individuals are not falsely detected. This would mean that a species is correctly identified. It is mentioned later that such false positives can seriously bias occupancy estimates.

Depending on the model, further assumptions may be made. For example, the probability of occupancy may be assumed to be constant across all sites, and/or the probability of detection p may be constant for all occasions. Sometimes, an average parameter value can be used. Alternatively, a Bayesian method can be applied where parameters have prior distributions.

4.4.1 Model Adequacy

MacKenzie and Bailey (2004) considered testing the adequacy of the model of MacKenzie et al. (2002) described above with S sites and T samples per site. For example, with $S = 5$ sites, a detection history for site 1 might be $\mathbf{x}_1 = 10100$ with probability

$$\Pr(\mathbf{x}_1 = 10100) = \psi_1 p_{11}(1 - p_{12})p_{13}(1 - p_{14})(1 - p_{15}),$$

where p_{1j} is the probability detecting the complex, given presence, in the j th survey of site 1. Using covariates, let O_h be the number of sites observed to have detection history h and E_h be the expected number of sites with history h according to the above model. Generally, E_h equates to the sum of the estimated probabilities of observing h , namely,

$$E_h = \sum_{i=1}^S \Pr(\mathbf{x}_i = h),$$

where parameters are replaced by their maximum likelihood estimators. Assuming no missing observations, there are 2^T possible detection histories (a sequence of T 0's and 1's) that may be observed and tested with a goodness-of-fit test

$$X^2 = \sum_{h=1}^{2^T} \frac{(O_h - E_h)^2}{E_h}.$$

Because some underlying assumptions will not be satisfied and some E_h will be small, X^2 will not have a chi-square distribution. As an alternative, the parametric bootstrap may be used to determine whether the observed value is unusually large. MacKenzie and Bailey (2004) gave a procedure for carrying this out. They also provided an estimate of an overdispersion parameter which can be used in model selection. Their model was applied to terrestrial salamander data, to hypothetical examples, and to a Monte Carlo simulation study. They used a program **PRESENCE**, specifically developed for analyzing site occupancy data, which may be downloaded from <http://www.proteus.co>.

4.4.2 Closure Assumption

Rota et al. (2009) looked particularly at the closure assumption and considered testing it using a removal-type approach, but based on the method of MacKenzie et al. (2003) described above in Sect. 4.3.1. They used Pollock's (1982) robust design, which uses secondary sampling periods nested within primary sampling periods. Here, populations are assumed to be closed to demographic changes between secondary sampling periods and open to demographic changes between primary sampling periods. A species is only recorded when first sighted, so after sighting, it is effectively “removed.” One main reason for this is to avoid the situation where a species detected during one survey may be easier to detect during subsequent surveys once its location is known, so that the surveys are not independent. In applying the robust model to occupancy, individual site visits are considered as primary sampling periods, and multiple surveys conducted during each site visit as secondary sampling periods. Sites are open to change in occupancy between site visits, but are closed to changes in occupancy during site visits.

Transitions in occupancy between primary sampling periods can be estimated by fitting dynamic models to data collected using γ , the probability that an unoccupied site at time t will become occupied at time $t + 1$, and ε , the probability that an occupied site at time t will become unoccupied at time $t + 1$. For example, using a so-called two-season dynamic removal model that extends the one-season model of MacKenzie et al. (2006, 202), Rota et al. (2009) considered the following likelihood function for S sites:

$$\begin{aligned} L(\psi_1, p_1, p_2, \gamma, \eta \mid j_{s1}, j_{s2}, y_{s1}, y_{s2}) \\ = \prod_{s=1}^S \left\{ \psi p_1^{y_{s1}} (1 - p_1)^{j_{s1} - y_{s1}} \left[(1 - \varepsilon) p_2^{y_{s2}} (1 - p_2)^{j_{s2} - y_{s2}} + \varepsilon I(y_{s2} = 0) \right] \right. \\ \left. + (1 - \psi) I(y_{s1} = 0) \left[\gamma p_2^{y_{s2}} (1 - p_2)^{j_{s2} - y_{s2}} + (1 - \gamma) I(y_{s2} = 0) \right] \right\}, \end{aligned}$$

where ψ is the probability a site is occupied during time 1; p_1 and p_2 are the conditional probabilities of detecting a species, given presence, during times 1 and 2, respectively; and j_{s1} and j_{s2} are the number of surveys until the first detection of a species at site s during times 1 and 2, respectively. If a species remains undetected at site s during time t , then $j_{st} = 2$. Also, y_{s1} and y_{s2} are binary indicators of whether a species is detected ($y = 1$) or not ($y = 0$) at site s during times 1 and 2, respectively; I is an indicator function where $I = 1$, if y_{s1} or $y_{s2} = 0$ for the two cases, and $I = 0$, otherwise.

A likelihood ratio test Λ , say, for closure between primary units is a test of the hypothesis H_0 that $\varepsilon = \gamma = 0$ versus the open hypothesis that ε and/or $\gamma > 0$. Under standard regularity conditions, $-2 \log(\Lambda)$ would be asymptotically distributed as χ^2 , but in this case, H_0 has parameters that are on the boundary of the parameter space ($\varepsilon = \gamma = 0$), and the limiting distribution is a mixture of χ^2 and zeros (Self & Liang, 1987; see also Pledger, 2000: 437–438 for similar examples). Rota et al. (2009) detailed a parametric bootstrap method, and an approximate method of obtaining the mixing proportion for this distribution is given in Appendix S1 of their paper.

Otto et al. (2013) considered the situation when sampling may violate the closure assumption. They considered a design that allowed one to differentiate among two processes that violate the closure assumption during a sampling season, namely, repeated destructive sampling events that result in either short- or long-term site avoidance by the target species and sampling occurring over a time period during which nonrandom movements of the target species result in variable occupancy status. Sample sites were selected using “cells” as a first step and then transects within cells, and the robust model that incorporated primary and secondary periods as described previously in Sect. 4.4.2 was used. They utilized the so-called dynamic occupancy models (MacKenzie et al., 2003) described previously to estimate initial occupancy probabilities (ψ_1) and time-specific extinction and colonization probabilities (ε_t , γ_t) and thereby derive estimates of time-specific occupancy probabilities (ψ_{t+1}).

The authors considered some alternative models based on various hypotheses common to any species that occurs seasonally or may be influenced by the sampling process. For example, the first model was that occupancy varied as a function of the total number of woody cover objects along each transect ($\psi_1(cover)$), but with no site extinction or colonization ($\varepsilon = \gamma = 0$) occurring throughout the duration of the study. Other models included initial occupancy and extinction varying spatially as a function of woody cover, trend models using covariates, and several models involving different types of detection probabilities. In all, 14 models were considered using program **MARK** and Akaike’s information criterion, adjusted for small sample size, to rank the models. They found evidence of a chronic decrease in occupancy, an increase in local extinction, and a temporary decrease in detection probability, associated with repeated sampling disturbances in their salamander system. They also found evidence for changes in seasonal occurrence of salamanders that was independent of the sampling process.

Bled et al. (2013) considered dynamic occupancy models to analyze large-scale atlas data including the estimation of local colonization and persistence probabilities. They also accounted for spatial autocorrelation using conditional autoregressive models and autologistic models.

Staggered Arrival and Departure Times

Kendall et al. (2013) gave a statistical model for relaxing the closure assumption in occupancy models using staggered arrival and departure times at each site. We begin with the parameters ψ_s , the probability that site s ($s = 1, \dots, S$) is occupied by the species, and p_{st} , the probability of detection at site s at time t , given available for detection at that time. We now add two more parameters for single entry to and single exit from each site by the species of interest.

The first parameter β_{st} is the probability that the species enters the study area between sampling occasions t and $t + 1$, given that site s is occupied during the season. Here, β_{s0} is the probability the species enters in time for sampling occasion 1. Then

$$\sum_{t=0}^{T-1} \beta_{st} = 1.0.$$

The second parameter, d_{st} , is the probability that the species will depart the site (and therefore be unavailable for detection) before sampling occasion $t + 1$, given it is available for detection at site s on sampling occasion t . The terms “entry” and “departure” are used more generally to define a period over which the species is available for detection. For example, if detection of an amphibian species is based on calling, entry time is when they first begin calling for the season. Using 1 for detection and 0 for non-detection, the authors give two examples of structure probabilities for detection histories over $T = 3$ sampling occasions, namely,

$$(001) \quad \psi_s [\{\beta_{s0}(1 - p_{s1})(1 - d_{s1}) + \beta_{s1}\} p_{s2}(1 - d_{s2}) p_{s3}]$$

$$(000) \quad \psi_s (1 - p_s^*) + (1 - \psi_s),$$

where

$$p_s^* = \beta_{s0} p_{s1} + \{\beta_{s0}(1 - p_{s1})(1 - d_{s1}) p_{s2} + \beta_{s1} p_{s2}\}$$

$$+ \{\beta_{s0}(1 - p_{s1})(1 - d_{s1})(1 - p_{s2})(1 - d_{s2}) p_{s3}$$

$$+ \beta_{s1}(1 - p_{s2})(1 - d_{s2}) p_{s3} + \beta_{s2} p_{s3}\}.$$

The derived parameter p_{st}^* describes the probability of being detected at least once across the T sampling occasions, given the species is present for the first time in each sampling occasion.

Because the entry and departure processes complicate the structure of the model, it is perhaps clearer to view the species as transitioning within the season among

three availability states: (1) not yet arrived, (2) present and available for detection, and (3) departed for the season. The following 3×3 matrices include \mathbf{D}_{st} , which describes the probabilities of transition from row state r to column state c , and \mathbf{P}_{st} and \mathbf{Q}_{st} , diagonal matrices that denote the probability of detection or non-detection, respectively, for a site in a given state, namely:

$$\begin{aligned}\mathbf{D}_{st} &= \begin{bmatrix} (1 - e_{st}) & e_{st} & 0 \\ 0 & (1 - d_{st}) & d_{st} \\ 0 & 0 & 1 \end{bmatrix}, \\ \mathbf{P}_{st} &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & p_{st} & 0 \\ 0 & 0 & 0 \end{bmatrix}, \\ \mathbf{Q}_{st} &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & (1 - p_{st}) & 0 \\ 0 & 0 & 1 \end{bmatrix}, \\ \mathbf{n}'_0 &= [1 \quad 0 \quad 0],\end{aligned}$$

where \mathbf{n}_0 is the availability vector just prior to the field season (i.e., the species always begins in the “not yet arrived” state). For ease of presentation and computation, matrix \mathbf{D}_{st} includes a conditional entry probability (i.e., given the species has not yet entered the site by sampling period t):

$$e_{st} = \frac{\beta_{st}}{1 - \sum_{v=0}^{t-1} \beta_{sv}}$$

for $t = 1, \dots, T - 1$ (implies $e_{s,T-1} = 1.0$). Finally, for just prior to the season (i.e., before the first sample), we define $d_{s0} = 0$ and $e_{s0} = \beta_{s0}$. We can now use matrix notation to write the probability structure for a given detection history, h_s for site s , across T sampling occasions. For cases where the species is detected at least once,

$$\mathbf{P}(h_s) = \psi_s \mathbf{n}'_0 \left[\prod_{t=1}^T \mathbf{D}_{s,t-1} (\mathbf{P}_{st})^{x_{st}} (\mathbf{Q}_{st})^{1-x_{st}} \right] \mathbf{1}_3,$$

where x_{st} is an indicator variable that is 1 when the species is detected and 0 otherwise (e.g., $(\mathbf{P}_{st})^0$ is the identity matrix) and $\mathbf{1}_3$ is a column vector of 1’s of dimension 3. For histories of all zeros,

$$\mathbf{P}(h_s) = \psi_s \mathbf{n}'_0 \left[\prod_{t=1}^T \mathbf{D}_{s,t-1} \mathbf{Q}_{st} \right] \mathbf{1}_3 + (1 - \psi_s).$$

Maximum likelihood estimates of the parameters were obtained using a program **PRESENCE** available online in the supplementary material in Kendall et al. (2013), and Bayesian inference can also be used. With this open model, not all the parameters can be estimated individually, for example, β_{s0} and p_{s1} are confounded, and β_{s1} cannot be estimated without additional assumptions. By constraining one of these three parameters equal to something estimable (e.g., set $p_{s1} = p_{s2}$), the confounding can be resolved. Similarly, $d_{s,t-1}$ and p_{st} are confounded, and $\beta_{s,T-1}$ cannot be estimated without additional assumptions (e.g., set $p_{sT} = p_{s,T-1}$). As we have seen with models in this book, we can resolve these issues by using covariates or random effects to reduce the number of parameters. It was assumed that occupancy, entry, departure, and detection probabilities are the same at each site unless differentiated in the model by group or individual site covariates.

The above model was extended by Chambert et al. (2015) to include multi-season occupancy dynamics and showing how estimates and standard errors of derived parameters of interest can be obtained. Their emphasis in using this approach was to investigate inter-annual variation, especially trends, in both occupancy (spatial distribution) and within-season residence patterns (phenology).

4.5 Some Alternative Methods

Some of the earlier approaches used a two-step method where the probability of detecting the species is estimated in the first step and is then used in a second step to estimate the occupancy parameter. One ad hoc method using this approach was proposed by Geissler and Fuller (1987). They assumed that the detection probability does not change after the first detection of the species. However, once observers have detected the species at a site, they may be more or less likely to detect the species in subsequent surveys. For example, once the observer has found an indication that the species is present on a site, they may be able to use the information to make detection easier in future visits. This is reminiscent of trap-happy individuals in capture-recapture experiments where they are more likely to be caught in the future once caught for the first time. One method around this problem mentioned below is to record just the first time an individual is detected. This is essentially a removal method previously mentioned.

Another ad hoc method with a slightly different sampling scheme to that in the previous paragraph for monitoring spotted owls was given by Azuma et al. (1990) in that once detection occurred, the site was not revisited, reminiscent of the removal method of Sect. 1.1. Nichols and Karanth (2002) used the idea of the closed capture-recapture method for monitoring tiger and their prey species in India. Here, the detection histories for sites where a species was detected at least once are regarded as the capture histories of individuals encountered during a mark-recapture experiment. Then estimating the number of occupied sites where the species was not detected is completely analogous to estimating the number of individuals in the population that were never captured.

4.6 Choosing Design Parameters

General comments about the planning of a study and having clear objectives are given in Sect. 1.10. We now look at specific details and refer to some helpful comments by Henderson (2021, Chapters 11 and 12). Once the sampling unit is selected, the next question relates to its size as the spatial scale with occupancy is important as bigger units will have a higher probability of occupancy. Also, the amount of survey effort required needs to be considered as bigger units require more effort. A unit should be big enough for there to be a reasonable probability of the species being present (they suggest 0.2–0.8 in most situations). The role of the closure assumption needs to be taken into consideration as this may not hold too well for smaller units. There is also the question of whether the study will be repeated in the future (e.g., next season) so that size consistency is important for, say, studying trends. Clearly, all units in the study area should have a non-zero probability of being selected. The timing between the searches at a site and overall time for the study have to be carefully selected, and this will depend on biological (e.g., avoiding movement of species between sites) and logistical considerations for the sampling process (e.g., access).

How many units should there be? It is a question of balancing the number S of units (sites) in a study and the number of T surveys (searches) per site. In addition, there may be more than one overall study, for example, a one-season study or multiple-season studies. It is important to try and have independent searches as much as possible. For example, the same observer revisiting a site may remember where to look. Using different observers may introduce an observer effect, which usually exists, depending on a number of factors such as the level of individual training, eyesight, and hearing, though covariates can sometimes be used to allow for observer differences. Environmental conditions can also play a role. As a general rule, MacKenzie and Royle (2005) suggested that for a rare species it is more efficient to survey more sites less intensively, while for a common species survey fewer sites but more intensively.

For design purposes, it is usual to assume that the occupancy probability ψ and the probability of detection p for a given search, given present (occupied), are both constant and there are no false detections. Then MacKenzie and Royle (2005) suggested two approaches for designing a study: (1) aim to achieve a desired level of precision for minimal total survey effort or (2) minimize the variance for a given total number of surveys. This means determining the values of S and T to most efficiently achieve the study's objectives, given the assumed values of ψ and p .

4.6.1 Three Designs

Three types of design were considered by MacKenzie and Royle (2005): (a) a standard design where S sites are each surveyed T times, giving $\tau = TS$ total surveys; (b) a type of double sampling design, where S_T sites are surveyed T times

and S_1 sites surveyed once (with the second round of sampling performed to select sites at which the repeated survey will be conducted); and (c) a removal design where S sites are surveyed up to a maximum of T times, but surveying halts at a site once the species is detected. We now need to choose parameters like S and T to optimize some criterion. We focus mainly on the standard design to give the general idea.

Let $p^* = 1 - (1 - p)^T$, the probability of detecting at least one individual; then it transpires that

$$\begin{aligned}\text{var}[\hat{\psi}] &= \frac{\psi}{S} \left[(1 - \psi) + \frac{(1 - p^*)}{p^* - Tp(1 - p)^{T-1}} \right] \\ &= \frac{\psi}{S} f_1(T), \text{ say.}\end{aligned}$$

We can then rearrange the above equation to give

$$\tau = \frac{\psi T}{\text{var}[\hat{\psi}]} f_1(T),$$

which, for a given precision $\text{var}[\hat{\psi}]$, we can differentiate $f_2(T) = Tf_1(T)$, setting it zero and solving for T analytically or numerically leading to an optimal value of τ , the total number of surveys. Alternatively, if the criterion is to minimize $\text{var}[\hat{\psi}]$ for a given τ , we minimize $f_2(T)\psi/\tau$ with respect to K and substitute this value to give τ . We note that both cases reduce to considering $f(T) = Cf_2(T)$, where C does not depend on T , so that the optimal T , T_{\min} , is the same in both cases.

If costs are taken into account, one can use a possible cost function for the study such as

$$\text{costs} = c_0 + S[c_1 + c_2(T - 1)],$$

where c_0 is a fixed overhead cost, c_1 is the cost of conducting the first survey of the site, and c_2 is the cost to subsequent surveys. We can then minimize with respect to T the cost for a given variance or the variance with a given cost. They recommended that sampling sites should be surveyed at a minimum of three times when the detection probability is high (>0.5 per survey). The authors also applied the method to the removal and double sampling designs.

For the standard design, Sliwinski et al. (2016) suggested the following method for finding T . If P is the probability of not detecting an individual in T searches individual from a single search given presence, then the probability of not detecting any individuals in T searches is $P = (1 - p^*)^T$. One criterion is to choose P small, say no greater than $P = 0.1$. We then have $T \geq \log(d) \cdot \log[(1 - P)^{-1}]$, and the authors give a number of graphs of P versus T . Of course, we need to have some idea of P (and P may not be constant), which may be available from previous surveys or from a pilot survey using small T .

4.6.2 Bayesian Optimal Design

Guillera-Arroita et al. (2014) showed how a simple two-stage design can significantly improve the efficiency of an occupancy study. Using S sites, it is assumed that the occupancy probability ψ and the probability of detection p , given present, are both constant and there are no false detections. Then the likelihood function for the S sites is

$$L(\psi, p) = \prod_{s=1}^S \left\{ \psi p^{d_s} (1-p)^{T_s - d_s} + (1-\psi) I(d_s = 0) \right\},$$

where T_s is the number of surveys carried out at site s , d_s is the number of them that resulted in detection, and $I(\cdot)$ takes the value 1 when the expression in brackets is true and 0 otherwise. This is a zero-inflated binomial model, where $1 - \psi$ is the zero-inflated probability.

In the single-stage design, the whole study is designed prior to data collection, where typically the same amount of survey effort T is applied to every site. Hence, the likelihood for the model can be written as

$$L(\psi, p) = \left\{ \psi^{S_d} p^{d_t} (1-p)^{TS_d - d_t} \right\} (1 - \psi p^*)^{S - S_d},$$

where $S_d = \sum_{s=1}^S I(d_s \neq 0)$ is the number of sites where the species was detected at least once, $d_t = \sum_{s=1}^S d_s$ is the total number of detections, and $p^* = 1 - (1-p)^T$ is the probability of detecting the species in at least one of the T surveys carried out at an occupied site.

The inverse of the expected information matrix for $L(\psi, p)$ can be used to obtain

$$\text{var}[\hat{\psi}] = \frac{\psi}{S} \left(1 - \psi + \frac{1 - p^*}{p^* - Tp(1-p)^{T-1}} \right).$$

The study's aim was to choose a design to minimize the above variance with respect to T , the number of replicates for a given total survey effort ($E = TS$), or equivalently minimize the total survey effort required to achieve a target estimator precision. Assuming particular values for the parameters leads to locally optimal designs, whereas assigning prior distributions to parameter values leads to Bayesian optimal designs. Given a prior distribution $f(\psi_0, p_0)$ for (ψ, p) , the Bayesian design approach chooses S to minimize the expected asymptotic variance of the occupancy estimator over that prior given by

$$E(\text{var}[\hat{\psi}]) = \int \int \frac{T\psi_0}{E} \left\{ (1 - \psi_0) = \frac{1 - p_0^*}{p_0^* - Tp_0(1-p_0)^{T-1}} \right\} \cdot f(\psi_0) d\psi_0 dp_0,$$

where $p_0^* = 1 - (1 - p_0)^T$.

Guillera-Arroita et al. (2014) gave two tables showing optimal values for both the above two methods. They then introduced a two-stage approach where $100r\%$ of the total survey effort (E) is employed in the first stage and the remaining effort is used in the second stage. For simplicity, we assume that the effort corresponding to the second stage is allocated to sites different from those sampled in the first stage. The optimal amount of replication for the second stage (T_2), based on the updated parameter information, might be different from that identified for the first stage (T_1) based on the initial parameter estimates. The two-stage procedure is as follows:

- (1) Based on the initial estimates (ψ_0, p_0) , choose the design for stage 1, i.e., identify the optimal T_1 , and set S_1 to be the nearest integer to Er/T_1 .
- (2) Collect data (summarized by S_{d_1}, d_{t_1}), and obtain the maximum likelihood estimates $(\hat{\psi}_1, \hat{p}_1)$.
- (3) Based on the updated estimates $(\hat{\psi}_1, \hat{p}_1)$, choose the design for stage 2, i.e., identify the optimal T_2 , and set S_2 to be the integer part of $E(1 - r)/T_2$.
- (4) Collect data summarized by (S_{d_2}, d_{t_2}) , and analyze the full data set from both stages, to obtain the final estimates $(\hat{\psi}_2, \hat{p}_2)$.

We choose the optimal T_1 as the amount of replication that would minimize the variance of the occupancy estimator in one stage if our initial estimates (ψ_0, p_0) were true, whereas the optimal T_2 is the amount of replication in the second stage that would minimize the variance of the overall occupancy estimator (i.e., when the full data set estimates from the first stage $(\hat{\psi}_1, \hat{p}_1)$ were true). The resulting likelihood for the two stages combined is

$$L(\psi, p) = \left\{ \psi^{S_{d_1}} p^{d_{t_1}} (1 - p)^{T_1 S_{d_1} - d_{t_1}} \right\} (1 - \psi p_{T_1}^*)^{S_1 - S_{d_1}} \\ \times \left\{ \psi^{S_{d_2}} p^{d_{t_2}} (1 - p)^{T_2 S_{d_2} - d_{t_2}} \right\} (1 - \psi p_{T_2}^*)^{S_2 - S_{d_2}},$$

where S_{d_i} and d_{t_i} are, respectively, the number of sites where the species was detected at least once and the total number of detections in stage i and $p_{T_i}^* = 1 - (1 - p)^{T_i}$. The inverse of information matrix for the combined design was derived to obtain $\text{var}[\hat{\psi}_2]$. This variance is conditional on the outcome of the first stage $(\hat{\psi}_1, \hat{p}_1)$ since this outcome determines the design for the second stage (T_2 and S_2). The unconditional variance of the occupancy estimator can then be computed by taking the expectation of the variance over all the possible outcomes of the first stage

$$\text{E}[\text{var}(\hat{\psi}_2)] = \sum \sum \text{var}[\hat{\psi}_2 | \psi, p, T_1, T_2, S_1, S_2] \cdot \Pr(\hat{\psi}_1, \hat{p}_1 | \psi, p, S_1, T_1),$$

where the double summation is over all values of $\hat{\psi}_1$ and \hat{p}_1 . The authors then compared the single and combined designs and in the latter determined how to find the optimal value of r , the design fraction. Prior distributions for the initial values ψ_0 and p_0 were also used for this. It was evident that the two-stage design was more

robust to poor initial estimates than the single-stage design. A full discussion of the assumptions used was given.

4.7 Adaptive Methods for Rare Species

Determining occupancy can be problematic with rare species because of the paucity of data. This is particularly a problem when the emphasis is on a community of species and how the abundance and distribution of species depend on environmental covariates. Ovaskainen and Soininen (2011) discussed this problem and introduced a comprehensive multivariate logistic regression model using covariates for the probability of presence, thus using the community information to help with the rare species. We note that for rare species, the choice of design and parameters is particularly important, and we now consider some alternative approaches below.

4.7.1 Two-Phase Design

Pacifci et al. (2012) developed a two-phase sampling approach for rare species by using a design that places more effort in areas with higher predicted probability of occurrence. They assumed two requirements: (i) that variation in occurrence is associated with spatially varying habitat information, such that interest lies in examining the relationship between a suite of covariates and their influence on occupancy, and (ii) the availability of fully observable covariate information throughout the study area, possibly obtained through remotely sensed data.

In the first phase, we use a traditional, single-season occupancy model in which a simple random sample of size n_1 is selected (without replacement) from a sample frame composed of S sites. At each site, $T > 1$ independent surveys are taken, and the number of surveys $y \in \{0, 1, \dots, T\}$ resulting in the detection of one or more individuals of the species being recorded. Then sampling n_1 sites yields a vector of observations: $\mathbf{y} = (y_1, \dots, y_{n_1})$. Let $z_s = 1$ if the species is present at site s with probability ψ_s , $z_s = 0$ for a species absence, and p be the probability of being detected at site s given present at the site. Then

$$y_s | z_s, p \sim \text{Binomial}(T, z_s p) \text{ and } z_s | \psi_s \sim \text{Bernoulli}(\psi_s),$$

and $\text{logit}(\psi_s)$ can modeled as a linear regression of site-specific covariates, namely,

$$\text{logit}(\psi_s) = \beta_0 + \boldsymbol{\beta}' \mathbf{x}_s.$$

Under conditional independence among sites, the marginal or integrated likelihood for the observations (integrating out the latent occupancy state z_s) is

$$\begin{aligned}\Pr(\mathbf{y}) &= \prod_{s=1}^{n_1} \Pr(y_s \mid \psi_s, p) \\ &= \prod_{s:y_s > 0} \binom{T}{y_s} \psi_s p^{y_s} (1-p)^{T-y_s} \prod_{s:y_s=0} [\psi_s (1-p)^T + (1-\psi_s)].\end{aligned}$$

The above model provides maximum likelihood estimates for p and the β coefficients.

In the second phase, let $m = S - n_1$ denote the number of remaining non-sampled locations. At such a location j , say, covariate measurements from the site can be used in the logistic regression model to predict ψ_j , namely,

$$\hat{\psi}_j = \exp[\hat{\beta}_0 + \hat{\beta}' \mathbf{x}_j] / (1 + \exp[\hat{\beta}_0 + \hat{\beta}' \mathbf{x}_j]).$$

We then calculate the probability of inclusion for each site j as a logistic-type ratio

$$\pi_j = \hat{\psi}_j / \sum_{j=1}^m \hat{\psi}_j.$$

The second phase of sampling consists of selecting n_2 of the remaining m sites without replacement such that each site j has the associated sampling weight π_j . Estimation of the associated parameters, p and the β 's, is then carried out by maximizing the likelihood using the complete data ($n_1 + n_2$ sites). The second phase of sampling needs to be conducted during a “single season,” and thus, we are assuming that there is closure for the entire duration of the study when using the two-phase sampling approach.

Pacifci et al. (2012) considered the separate problem of estimating the number of occupied sites S_{occ} directly using three different types of site: (i) those that were sampled and a detection occurred, (ii) those that were sampled and no detections occurred, and (iii) those that were not sampled. For type (i), $z_s = 1$, and for type (ii), we can have either the site being unoccupied or the site being occupied but no detections occurred. Using Bayes' rule, this gives us

$$\Pr(z_s \mid y_s = 0) = \frac{\psi_s (1-p)^T}{\psi_s (1-p)^T + (1-\psi_s)},$$

where ψ_s is given by $\hat{\psi}_s$ described above. The third type of site is simply a prediction for a non-sampled location j , given a new covariate value and estimates of the β 's, so that the probability of occurrence is $\hat{\psi}_j$ given above. The goal is to estimate S_{occ} using a frequentist version of the Bayesian estimator $S\psi_{fs} = \sum_{s=1}^S z_s$, where $f s$ is

for “finite sample.” Then

$$\begin{aligned}\widehat{S}_{occ} &= \sum_{s:y_s>0}^{n_1} E[z_s \mid y_s > 0, \mathbf{x}_s] + \sum_{s:y_s=0}^{n_1} E[z_s \mid y_s = 0, \mathbf{x}_s] + \sum_{j=1}^m E[z_j \mid \mathbf{x}_j] \\ &= \sum_{s:y_s>0}^{n_1} 1 + \sum_{s:y_s=0}^{n_1} \frac{\widehat{\psi}_s(1 - \widehat{p})^T}{\widehat{\psi}_s(1 - \widehat{p})^T + 1 - \widehat{\psi}_s} + \sum_{j=1}^m \widehat{\psi}_j.\end{aligned}$$

The authors considered an alternative approach to estimating S_{occ} directly using a single sample of n sites assuming constant ψ . They set up the likelihood

$$\begin{aligned}L(\psi, p, S_{occ} \mid \mathbf{y}, S, n) &= \left[\frac{S_{occ}!}{(S_{occ} - n)!} p^{y_s} (1 - p)^{TS - y_s} \right] \\ &\quad \times \left[\frac{S!}{S_{occ}!(S - S_{occ})!} \psi^{S_{occ}} (1 - \psi)^{S - S_{occ}} \right],\end{aligned}$$

where $y_s = \sum_{s=1}^n y_s$, and use maximum likelihood to estimate ψ , p , and S_{occ} .

They also developed a measure of predictive ability to compare their new two-phase sample design with the traditional simple random sample and single-season occupancy approach using simulation. The two-phase sampling approach outperformed the traditional single-season approach in the majority of the scenarios investigated. Allocating 25% or 50% to the first phase of sampling for the two-phase approach resulted in the lowest overall error rate, and the difference was the largest between the two approaches in two habitats when the detection probability was high (0.75). It should be emphasized that the two-phase method is dependent upon the availability of covariate information across the entire landscape. Remotely sensed data are becoming increasingly more available.

4.7.2 Conditional Design

An alternative adaptive design, referred to as a conditional design, for dealing with populations of rare and cryptic species, namely, those that occur on only a small portion of the landscape and cannot be detected reliably during a single survey, was given by Specht et al. (2017) and compared with the standard and removal designs. The design involved visiting S sites once, and sites where the species of interest is encountered during the first survey are visited an additional $(T - 1)$ times to better estimate the detection probability.

We now set up the notation for the three designs. Once again, ψ is the probability of being present, and p is the probability of detection given present. For the standard occupancy model, let T_S be the number of replicate surveys for each site s ($s = 1, \dots, S$). For the removal model, removal sampling is carried out at each

site for R sites until the required species is detected or until T_R replicate surveys are carried out. For the conditional model, T_C sites are surveyed once, and sites where occupancy is confirmed during the first visit are resurveyed $T_C - 1$ times. With regard to sampling effort, the total effort measured in terms of surveys taken is for the standard design

$$E_S = S_S T_S.$$

For the other two designs, the effort will be random around their expected values with

$$E_C = S_C [1 + \psi p(T_C - 1)]$$

for the conditional design and

$$E_R = S_R \left[T_R(1 - \psi) + \frac{\psi p^*}{p} \right]$$

for the removal design, where $p^* = 1 - (1 - p)^T$, the probability of at least one detection at an occupied site across T surveys. For the conditional design, replicate surveys only occur at sites where the species was detected in the first visit, so $p^* = p$. For the conditional and removal designs, the distribution of survey effort between sites and surveys is affected by the unknown true state of the study system. This means that more sites will be visited under these two designs for a given total effort (E) and maximum number of surveys per site (T), as some sites will not be revisited T times. These differences in the distribution of sampling effort mean that the three survey designs differ in values of key statistics: expected numbers of sites with positive detections (S_D), total detections (d), and supplemental detections ($d_+ = d - S_D$).

For the conditional model, let T_D be the number of sites with positive detections, d the total detections, and $d_+ = d - S_D$ the number of supplemental detections. The likelihood function for S sites and $T - 1$ possible replications is then

$$\begin{aligned} L(\psi, p) &= \left(\frac{S!}{S_D!(S - S_D)!} (\psi p)^{S_D} (1 - \psi p)^{S - S_D} \right) \\ &\times \left(\frac{(S_D(T - 1))!}{d_+!(S_D(T - 1) - d_+)!} p^{d_+} (1 - p)^{S_D(T - 1) - d_+} \right). \end{aligned}$$

The maximum likelihood estimates of the parameters are

$$\hat{p} = \frac{d_+}{(T - 1)S_D}$$

and

$$\hat{\psi} = \frac{S_D}{S\hat{p}} = \frac{S_D^2(T-1)}{d_+ S},$$

where, for given T and S , (S_D, d_+) are sufficient statistics. Like the standard and removal estimators, conditional occupancy can also produce boundary estimates of $\hat{\psi} = 1$ if p^* is too small (i.e., if there are too few supplemental detections). However, these sampling scenarios are highly improbable and are likely indicative of assumption violations. To compare the three designs, their asymptotic variances of $\hat{\psi}$ were derived, and the optimal design was obtained in each case for a given effort E . For example, for the conditional model, the optimal T is

$$T_{Copt} = 1 + \sqrt{\frac{1-p}{p^2\psi(1-p\psi)}}$$

suitably rounded off to an integer. Assuming equal costs between initial and replicate surveys, the authors compared estimates of optimal T for standard, removal, and conditional sampling designs for ψ and $p = 0.1$ to 0.9 at 0.1 increments.

All designs required more visits when p is low so as to provide more information for estimating detection. Using standard errors, the conditional design provided the most precise occupancy estimates when occupancy was low ($\psi \leq 0.2$), while the removal design provided the best precision when $\psi \geq 0.4$. The conditional design provided the most precise estimates of p when $\psi \leq 0.5$, but the standard design was superior when ψ or p was high ($p \geq 0.7$). Even with a low level of effort, differences among designs were generally small when effort was distributed optimally between sites and surveys for each design, and differences were expected to diminish further with increasing effort because the variances are inversely proportional to sampling effort.

Specht et al. (2017) concluded that the conditional design required the largest T to achieve an optimal design when occupancy is low, whereas the removal design required the largest T when occupancy is high. The standard design never requires the most replicates, but under the conditional or removal designs, most sites will be surveyed fewer than T times so the standard design often averages more visits per site. Further comparisons are given in their article. The method can also be used for pilot studies.

4.8 Occupancy and Abundance

In Sect. 4.1.4, we referred to the RN model for estimating abundance occupancy, and we now wish to make some further comments. If N is the population size for area A , it is reasonable to expect N to be related in some way to ψ , the probability of

being on the area. Following Royle and Dorazio (2009: Chapter 4), suppose that N has a discrete distribution $g(N; \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a (possibly vector-valued) parameter. As A decreases, we would expect N to decrease so that as $A \rightarrow 0$, $f(N; \boldsymbol{\theta})$ will concentrate on 0 or 1. This suggests that small spatial units might therefore be preferred if obtaining abundance information from occupancy is desired. Since the occupancy probability $\psi = \Pr(N > 0) = 1 - g(0; \boldsymbol{\theta})$, then given $\hat{\boldsymbol{\theta}}$ the maximum likelihood estimate of $\boldsymbol{\theta}$, we have that the maximum likelihood estimate $\hat{\psi}$ of ψ is given by $1 - g(0; \hat{\boldsymbol{\theta}})$. For example, if N has a Poisson distribution with parameter λ , then $\psi = 1 - e^{-\lambda}$, and, as $\lambda \rightarrow 0$, $\psi \approx \lambda$.

We now return to the model involving S sites sampled, with each site surveyed T times. Given a certain species is present on site s , let y_{st} ($t = 1, 2, \dots, T$) be the number of distinct individuals counted on site s at time t . Let N_s be the population size at site s . If individuals are detected independent of one another with probability r , then $P_s = 1 - (1 - r)^{N_s} = P(N_s)$, say, is the probability that at least one individual is detected. Here, P_s is used instead of p_s as the former is the probability of detection conditional on N_s so that when $N_s = 0$, we can have $P_s = 0$. Here, p_s of the previous models is the conditional probability of detection *at an occupied site*, that is, conditional on $N_s > 0$. If $y_s = \sum_t y_{st}$ is the total number of detections on site s , then

$$\begin{aligned} P_s &= \Pr(y_s > 0) \\ &= \Pr(y_s > 0 \mid N_s = 0)\Pr(N_s = 0) + \Pr(y_s > 0 \mid N_s > 0)\Pr(N_s > 0) \\ &= 0 + \Pr(y_s > 0 \mid N_s > 0)\Pr(N_s > 0) \\ &= p_s \psi. \end{aligned}$$

If we are interested in more than one species, we can simply add another subscript to denote the species. For example, Dorazio (2007) applied the method to frog species using logistic regression with weather covariates (temperature, wind, and humidity) and added an extra subscript for year to observe the dynamics of the population. When there is more than one species, there can be interest in species interaction and cohabitation (e.g., Fidino et al., 2019). Another study involving several species is given by Dorazio and Royle (2005a) who developed likelihood functions and provided an estimate of the number of species. They used an alternative model based on a reparameterization of the model for estimating species richness and accumulation that allows it to be fitted by Markov chain Monte Carlo (MCMC) sampling without any numerical integration (Dorazio & Royle, 2005b). Species' interaction and richness is discussed in Chap. 5.

Blanc et al. (2014) considered combining capture-recapture (CR) and occupancy data to estimate abundance as follows. Essentially, observation data from CR modeling is used for estimating abundance, and presence/absence data is used for estimating occupancy. To see a connection between the two, let N be the size of a population, where $N \sim \text{Poisson}(\lambda)$, and the probability of presence is $\psi = \Pr(N > 0)$, leading to $\lambda = -\log(1 - \psi)$. However, in order to treat N as a

fixed rather than a derived parameter, the authors used the complete data likelihood approach proposed by Schofield and Barker (2014) to fit a CR model with individual heterogeneity on the detection process p_i for individual i .

Linden et al. (2017) considered the relationship of occupancy and density for low-density carnivores across large landscapes and combined remote cameras with baited hair snares. They fitted occupancy and Royle-Nichols models (Sect. 4.1.5) to species detection-non-detection data collected by cameras and fitted spatial capture-recapture (SCR) models to individual encounter data obtained by genotyped hair samples.

Using DNA

DNA techniques can be incorporated into the above methods. For example, Doi et al. (2019), using environmental DNA (eDNA) metabarcoding along with a multispecies approach, described how to incorporate false-negative errors in species detection. The detection probabilities were modeled using logistic regression and an analysis of variance-type format with interactions, as well as using prior distributions for some of the parameters. They used the Markov chain Monte Carlo (MCMC) method in JAGS software version 4.2.038 to obtain samples from the posterior distribution of parameters. Posterior samples were obtained from 3 independent chains of 100,000 iterations after a burn-in of 100,000 thinning at intervals of 100. The convergence of MCMC was affirmed by determining if the appropriate \hat{R} statistic for each parameter of interest was less than 1.1. The topic eDNA is discussed further in Sect. 4.11.1.

4.8.1 Population Trend

Hierarchical models have been developed for inference about abundance and detection that explicitly account for a metapopulation design, and we consider one such model from Kéry and Royle (2010). They considered two classes of models. First, they regarded the data as independent binomial counts and modeled abundance and detectability based on a product-binomial likelihood. Second, they used the more complex detection-non-detection data for each site to form encounter history frequencies. They then extended both models to directly estimate population trends over multiple years. We consider the multinomial model first.

The aim is to select a substantial number of sites (say at least 30) and on each site carry out a sequence of surveys over time with at least three (T) surveys per site being the recommended number, though sometimes a single survey is used. Let N_s be the population size at site s , which is unknown (latent) because of being incompletely observed. If $T = 3$ and π_s is the probability of detection for a given survey and site s , then the detection process has 2^3 possible histories (h), for example, $h = 100$, and the probability of that history is $\pi_h = p_1(1 - p_2)(1 - p_3)$. We now do some relabeling and let n_{sk} ($k = 0, 1, \dots, 7$) be the number of each type of encounter history for site s , where we use the integers k here to index the eight

nominal categories representing possible encounter histories; n_{s0} is the number of individuals never encountered, i.e., those having encounter history $h = (000)$. Then $N_s = n_{s0} + \sum_{k=1}^7 n_{sk}$, and the joint distribution of $\mathbf{n}_s = (n_{s0}, \dots, n_{sk})'$ is the multinomial distribution

$$f(\mathbf{n}_s | N_s, \mathbf{p}) \propto \frac{N_s!}{n_{s0}!} \prod_{k=0}^7 \pi_k^{n_{sk}}.$$

By pooling data, we do lose some information, for example, in going from $k = 7$ to $T = 3$.

For the binomial model, let y_{st} be the counts of unique individuals during each of the T survey periods for site s , and let $\mathbf{y}_s = (y_{s1}, \dots, y_{sT})'$ with distribution a product of binomials, namely,

$$f(\mathbf{y}_s | N_s, \mathbf{p}) \propto \prod_{t=1}^T \frac{N_s!}{y_{st}!} p_t^{y_{st}} (1 - p_t)^{N_s - y_{st}}.$$

Variation in the N_s can be modeled using a log-linear model with covariates to account for spatial variation in abundance that is not captured by these covariates. They also modeled $y_{st} \sim \text{Poisson}(\lambda_{st})$, where

$$\log(\lambda_{st}) = \log(\lambda_0) + rt,$$

for a population trend, and added a normally distributed site random effect into the linear predictor for abundance. This allows for overdispersion relative to the conditional Poisson distribution. Vague priors were used for the parameters. They gave several different ways to proceed. For example, with the Poisson assumption, we have that n_{sk} is Poisson with mean $\lambda_s \pi_{sk}$, which is computationally convenient for analysis whether by likelihood methods or Markov chain Monte Carlo. Their method was applied to a farmland and a woodland species, skylark (*Alauda arvensis* L.) and willow tit (*Parus montanus* Conrad). Both are widespread but moderately rare species.

Earlier, Kéry et al. (2005) used a modeling approach similar to that above, but considered using some prior distribution $f(N_i; \boldsymbol{\theta})$, depending on an unknown parameter(s) $\boldsymbol{\theta}$, for the N_i . They set up a data model conditional on N_i , then obtained the joint distribution by multiplying by $f(N_i; \boldsymbol{\theta})$, and integrated (summed) out N_i to get a data likelihood free of N_i to estimate the required parameters. They suggested some methods for estimating N_i . For example, the maximum likelihood estimate of $\boldsymbol{\theta}$ can be used to obtain an estimate of the expected number of individuals in each site based on the values of site-specific covariates. Alternatively, N_i can be estimated by the posterior mean of N_i , given the observed counts at site i . Their paper should be consulted for further details.

4.8.2 Random or Elusive Aggregated Populations

Zhou and Griffiths (2007) considered estimating abundance from occupancy data for random or aggregated elusive populations. Suppose the population area A is divided into A/a cells of the same shape and size a . For randomly distributed individuals in a population of size N , the number x present in each cell is Binomial($N, a/A$), and the probability of there being zero individuals is $(1 - a/A)^N$. If ψ is the probability that a cell has at least one individual, then $\psi = 1 - (1 - a/A)^N$. Because of randomness, this model is essentially equivalent to assuming that x is Poisson with mean $\lambda = Na/A$, and the probability of presence is $\psi = 1 - \exp[-\lambda]$.

For the binomial model, let p_0 be the probability of detecting an individual of a species, given it is present in a cell (site), per survey. The detection probability of species members, P say, is the probability of detecting one or more individuals in the site so that $P = 1 - (1 - p_0)^n$, where n is the number of individuals in the site. Repeated surveys are carried out in S sites, and, for the s th such site with T_s surveys, the species is detected in n_s of them. Let ψ_s be the probability that a species is present in site s and P_s be the probability of one or more individuals being detected in site s , given present. We then have for site s the likelihoods

$$L(\psi_s, P_s) | T_s, n_s = \begin{cases} \binom{T_s}{n_s} P_s^{n_s} (1 - P_s)^{T_s - n_s} \psi_s, & n_s > 0, \\ (1 - P_s)^{T_s} \psi_s + (1 - \psi_s), & n_s = 0. \end{cases}$$

Assuming P and ψ are constant (say the average cell values) and substituting for ψ , we have the combined likelihood

$$\begin{aligned} L(N, P | \{T_s, n_s\}, A, a) = & \\ & \prod_{s=1:n_s>0}^{C_1} \left\{ \binom{T_s}{n_s} P^{n_s} (1 - P)^{T_s - n_s} \left[1 - \left(1 - \frac{a}{A} \right)^N \right] \right\} \\ & \times \prod_{s=1:n_s=0}^{C_0} \left\{ (1 - P)^{T_s} \left[1 - \left(1 - \frac{a}{A} \right)^N \right] + \left(1 - \frac{a}{A} \right)^N \right\}, \end{aligned}$$

where C_1 is the total number of sampled cells when $n_s > 0$ and C_0 is the total number of sampled cells when $n_s = 0$. The above likelihood can be used to obtain maximum likelihood estimates of N and P where $\text{logit}(P)$ can be modeled using a linear regression on covariate variables such as site characteristics and sampling gear. In practice, a simple random sample without replacement of size $C = C_0 + C_1$ will be used.

Zhou and Griffiths (2007) then extended the above approach to a negative binomial distribution instead of the binomial, where x the number in a site has

probability function

$$f(x) = \binom{x+k-1}{k-1} \pi_i^k (1-\pi)^x$$

and π is the probability that an individual is present in the site. For this model,

$$\text{E}[x] = \frac{k(1-\pi)}{\pi}, \text{ and } \text{var}[x] = \frac{k(1-\pi)}{\pi^2}.$$

Using Poisson notation, $\text{E}[x] = \lambda = Na/A$ and

$$\Pr(x=0) = \pi^k = \left(\frac{k}{k+\lambda}\right)^k = \left(\frac{Ak}{Ak+aN}\right)^k$$

so that $\psi = 1 - \Pr(x=0)$. Since we now have two confounding parameters k as well as N , which are not easy to separately estimate, a different procedure is used to sample s cells using additional neighboring sites to obtain information on the aggregation pattern. Simulation studies were carried out that showed the usefulness of the method; the effects of changing the sizes of various parameters, the number of surveys, and the sample size; and the method's advantages such as compared with Royle and Nichols' (2003) model.

Linden et al. (2017) considered the relationship between occupancy and abundance for low-density carnivores across large landscapes to investigate the usefulness of occupancy in this regard as applied to fishers (a small, carnivorous mammal native to North America). They fitted occupancy data using the RN model of Royle and Nichols (2003) to species detection-non-detection data collected by cameras and a spatial capture-recapture (SCR) model to individual encounter data obtained by genotyped hair samples (see Fuller et al., 2016, for further background details).

4.9 Observation Errors

Miller et al. (2012) investigated observation error in anuran call surveys and mentioned two kinds of error, false negative (non-detection when actually present), which we focused on above, and false positive (detection when not present), which had not received much attention at the time. A common form of false positive is species misidentification. Despite increasing use of occupancy models to account for variation in the detection process in, for example, anuran call surveys, little is understood about the mechanisms driving this process. The authors listed many environmental-type factors affecting detectability including substantial species variation. They also mentioned data collection methods where false positives are most likely to arise such as large-scale volunteer-based surveys, interviews with local experts, use of historical records, call surveys, computer algorithms to detect

species from recordings, and laboratory assays. The greatest effect occurs when the occupancy probability is low.

Molinari-Jobin et al. (2012) considered the possible errors using the monitoring of the Eurasian lynx in the Alps as an example. They used the so-called SCALP classification of reliability of distributional records that uses three categories: (C1) “hard fact” data (e.g., dead or captured lynx), (C2) confirmed data (e.g., tracks verified by an expert), and (C3) unconfirmed data (e.g., any kind of direct visual observation such as sightings and calls, which cannot be verified, and old field signs, e.g.). A grid of large quadrats or sites (100 km^2) was used. They contrasted two periods to outline the effect of the categorization in time, 1995–1997 versus 2006–2008, with each year representing a sampling occasion. Two assumptions were made: (1) Lynx distribution remained unchanged within each of the two periods, and (2) owing to the large number of persons and organizations, every quadrat was monitored so that the years without a lynx can be recorded as a zero rather than as a missing value in the detection history fed into the site occupancy model. Three parameters were defined: ψ (the usual occupancy probability), ϕ (persistence or “survival” probability), and γ (colonization) probability. In the first 3-year period and site s , they defined $z_{s1} = 1$ if there was at least one record and 0 if no lynx is recorded, so that

$$z_{s1} \sim \text{Bernoulli}(\psi).$$

For the second 3-year period,

$$z_{s2} | z_s \sim \text{Bernoulli}(z_{s1}\phi + (1 - z_{s1})\gamma).$$

For the observation process, let p_{st} be the annual detection probability for site s and 3-year period t . The observed data y_{st} , the number of years with a lynx detection in site s and 3-year period t , is a binomial random variable, namely,

$$y_{st} \sim \text{Binomial}(3, p_{st}).$$

The authors fitted a separate model to each of the three data sets (C1, C2, and C3) and assumed that probabilities of first-year occupancy (ψ), survival (ϕ), and colonization (γ) were constant over all the sites. It was also assumed that the detection probability differed only by a 3-year period, but not among sites nor among years within a period. They fitted the model in the Bayesian framework using Markov chain Monte Carlo simulation techniques implemented in the WinBUGS software. A variation of this approach with colonization, a multi-season model, was given by Broms et al. (2016a).

4.9.1 False Positives

False positives occur when a species is recorded as present when it is not present or there is species misidentification and false negatives when a species is not detected when actually present. Often, it is assumed that there are no false-positive detections. However, using a particular field simulation, Miller et al. (2012) found that false-positive errors made up 5% of all positive detections. Distance and ambient noise both proved to be important predictors, with increases in either variable associated with decreasing detection probabilities. Distance, day, ambient noise, and observer abilities were the most important factors explaining false-negative detections, while distance and observer ability were the overall predictors of false-positive errors, but ambient and competing species also affected error rates for some species. Apparently, it sometimes happens that a species is not recorded if there is any doubt about its identity. This converts many false positives into false-negative results and thus will tend to inflate the rate of false negatives.

There is mounting evidence that false-positive errors may be more problematic than previously thought, and there is a current lack of methods addressing this form of observation error. False-positive observations in studies of rare species may cause an overestimation of the distribution or abundance of the species. Because call surveys rely on auditory cues, they appear to be more susceptible to false-positive errors than methods relying on captures or sightings of individuals. Miller et al. (2012) found that “false-positive detections to be most attributable to differences in species, observers, and distance.” The authors referred to some other studies highlighting substantial errors due to false positives, and we now consider a model for misclassification due to Royle and Link (2006) that generalizes a model of MacKenzie et al. (2002) in Sect. 1.8.1.

Suppose $s = 1, \dots, S$ sites are classified according to some characteristic having two possible classes “occupied” ($z = 1$) and “not occupied” ($z = 0$) (though we can extend the theory to K possible classes). We denote the true class of site s as z_s and the observed class of site s for time t as y_{st} ($t = 1, \dots, T$). It is assumed that the y_{st} ($t = 1, \dots, T$) are independently and identically distributed so that $y_s = \sum_t y_{st}$ is a sufficient statistic. Let $\psi = \Pr(z_s = 1)$ be the occupancy probability. It is assumed that the sites are independent with respect to occupancy status and occupancy status does not change during the period of sampling. For each y_{st} , we have the classification or misclassification probabilities

$$p_{k\ell} = \Pr(y_{st} = k \mid z_s = \ell),$$

which sum to 1 for each z_s . Hence, we have p_{00} (not detected not occupied), p_{10} (detected not occupied), p_{01} (not detected but occupied), and p_{11} (detected and occupied). Also, $p_{00} = 1 - p_{10}$ and $p_{01} = 1 - p_{11}$, and p_{10} is the probability of a

false positive. Let $\mathbf{p} = (p_{11}, p_{10})'$; then we have the likelihood function

$$L(\mathbf{p}, \psi \mid \{y_s\}) \propto \prod_{s=1}^S \left\{ \left[p_{11}^{y_s} (1 - p_{11})^{T-y_s} \right] \psi + \left[p_{10}^{y_s} (1 - p_{10})^{T-y_s} \right] (1 - \psi) \right\}.$$

This likelihood can be used to obtain maximum likelihood estimates of ψ , p_{10} , and p_{11} . The authors noted that

$$L(p_{11}, p_{10}, \psi) = L(p_{10}, p_{11}, 1 - \psi)$$

so that, for example, a detection rate of 70% at occupied sites and false detection rate of 40%, with 80% of sites occupied, have the same likelihood as 40% detection rate at occupied sites and 70% false detection rate, with 20% of sites occupied. To get round this identifiability problem, it was suggested that a restriction of $p_{11} > p_{10}$ was sensible in that p is higher for occupied sites than the misclassification probability is for unoccupied sites. This likelihood symmetry implies also that, as p_{11} gets close to p_{10} , the maximum likelihood estimate becomes unstable. The usual conventional model gets round this problem by assuming $p_{10} = 0$. In conclusion, Royle and Link (2006) discussed briefly the question of extending the model to K possible classes, instead of just 2. They also showed that the probability of falsely concluding that a site is occupied in T samples, given a false-positive probability p_{10} , is $1 - (1 - p_{10})^T$, which increases rapidly as a function of p_{10} and T .

Species error can also be incorporated. For example, suppose p_a is the probability of correctly detecting a species at an occupied site and p_b is the probability of incorrectly detecting a species (at either an occupied site or an unoccupied site), and suppose that the two types of detection events are independent; then the likelihood function is the same as before except that $p_{11} = p_a + p_b - p_a p_b$ and $p_{00} = p_b$, so that p_a and p_b can be estimated. This topic of species error is considered further in Sect. 4.9.1.

Fitzpatrick et al. (2009) used the same model, focusing on low-density populations (e.g., “founder” populations of invasive species), but interpreted T to be the number of observers, thus allowing for observers to differ in their probabilities $p_{11}(t)$ and $p_{10}(t)$ of detection. They also provided sample data (infested hemlock from a tiny insect called HWA), R code, and Excel spreadsheets in a supplement to their paper. They concluded that in their study, “heterogeneity in detection associated with variation in abundance of HWA and differences in their ability of observers to detect low-density populations, rather than misclassification, is the factor most likely to be driving the differences in detection between sites.”

Gormley et al. (2011) compared information from incidental sightings with data from designed field surveys and gave a unifying framework for the case of the invasive sambar deer *Cervus unicolor* in Victoria, Australia. They compared a standard occupancy model with a presence-only habitat suitability model using the package **MAXENT** along with field sampling consisting of fecal pellet surveys, sign surveys, and camera trapping. A Bayesian state-space occupancy model was used to

predict the probability of suitable habitat from the field data. Occupancy models can be extended in a number of directions to incorporate other population variables such as dispersal and growth using matrix models when investigating species invasions (Hooten et al., 2007).

Chambert et al. (2015) described how the theory can be applied to three types of design that they call the site confirmation, calibration, and observation confirmation designs. In the first design, there are three types of observation: no detection, ambiguous detection, and unambiguous detection. The other two designs require additional information. Likelihoods are given for all three designs, as well as **R** and **BUGS** code needed for implementation. Guillera-Arroita et al. (2017) considered a similar problem concerning both false-positive and false-negative errors and noted that ambiguities in the data mean that estimation of relevant parameters might be confounded unless additional information is available to resolve those uncertainties. They introduced a two-stage occupancy-detection model to deal with the errors, though there are identifiability issues if only survey data prone to false positives is available. In this case, at least two sources of extra information are required for reliable estimation, or else one can set plausible bounds on false detection rates as prior information in a Bayesian setting. The methods were applied to eDNA data.

Ruiz-Gutiérrez et al. (2016) suggested using auxiliary data (e.g., from opportunity surveys, Sect. 4.16) and provided an integrated, computationally efficient Bayesian hierarchical model to correct for false positives. The auxiliary data are derived from a series of tests or experiments where it is known that an event has not occurred, but it is reported as detected or determined as having occurred during the experiment. Clement (2016) considered the design of occupancy studies when false-positive detections occur, and Clement et al. (2014) developed a model combining acoustic detectors and MistNet data to estimate probability of presence when survey data from bats may include false positives.

4.10 Further Multiple Detection Models

We have already discussed models in which there are replicate observations at a site, and we now consider further models that have some variations. Miller et al. (2011) considered the case where false positives include the situation where an animal is detected, but the species is misidentified, using a single-season method. They noted that the model of Royle and Link (2006) had some problems such as correlation among parameters in the model. Also, there was the inability to distinguish heterogeneity in the true-positive detection probabilities among sites from heterogeneity due to positive detections, leading to the identifiability problems mentioned above.

We now consider two methods by Miller et al. (2011). The true occupancy state of the s th site ($s = 1, \dots, S$), z_s , is assumed to come from one of the K occupancy states. Observations from the s th site on the t th visit, y_{st} , are classified into one of the L observation states that differ in the probability of being a false-positive

detection. We define $p_{\ell k} = \Pr(y_{st} = \ell \mid z_s = k)$ and $\psi_k = \Pr(z_s = k)$, where $\sum_k \psi_{sk} = 1$. The likelihood can then be expressed as

$$L(\{p_{\ell k}\}, \{\psi_{sk}\}) \propto \prod_{s=1}^S \left(\sum_{k=0}^{K-1} \left[\left\{ \prod_{t=1}^T p_{y_{st} k} \right\} \psi_{sk} \right] \right).$$

They demonstrated use of the likelihood when two types of detections occur, where the first type may include false-positive detections (“uncertain detections”), but the second does not (“certain detections”). For example, when sampling a site to determine whether it is occupied by a species, one might consider an indirect observation based on a sign to be uncertain such as a scat or tracks and a direct observation to be certain (e.g., visual encounter).

An observation y_{st} from the s th site on the t th visit is either 2, for a certain detection; 1, for an uncertain detection; or 0, for no detection, so that $L = 3$. The model can then be parameterized using three parameters: the probability of (incorrectly) detecting the species at a site given the site is unoccupied (p_{10}), the probability of detecting the species at a site given the site is occupied (p_{11}), and the probability that a detection is classified as certain given that the site is occupied and the species was detected (b). This leads to Table 4.1. In Appendix B of their paper, the authors show how the likelihood can be applied to another sampling design where $K = 4$ occupancy states and $L = 4$ observation states occur.

Miller et al. (2011) introduced a second method, which we call model 2 (the “multiple detection method model”). This involves visiting a site a further T_1 times in addition to the T visits, but using a second detection method (multiple devices are described in Sect. 4.19) below. Observations of the s th site on the v th visit, w_{sv} , are classified into one of the M observation states. The probability of detecting a species, conditional on whether a site is occupied or not, for each of the two methods, is, respectively,

$$p_{\ell k} = \Pr(y_{st} = \ell \mid z_s = k) \quad \text{and} \quad \tau_{mk} = \Pr(w_{sv} = m \mid z_s = k).$$

The likelihood is

$$L(\{p_{\ell k}\}, \{\tau_{mk}\}, \{\psi_{sk}\}) \propto \prod_{s=1}^S \left(\sum_{k=0}^{K-1} \left[\left\{ \prod_{t=1}^T p_{y_{st} k} \right\} \times \left\{ \prod_{v=1}^{T_1} \tau_{w_{sv} k} \right\} \psi_{sk} \right] \right).$$

Table 4.1 Values of $\Pr(y = \ell \mid z = k)$, $y = 0, 1, 2$, and $z = 0, 1$

| True state | $\Pr(y = 0 \mid z)$ | $\Pr(y = 1 \mid z)$ | $\Pr(y = 2 \mid z)$ |
|----------------------|---------------------|-------------------------|---------------------|
| $z = 0$; unoccupied | $1 - p_{10}$ | p_{10} | 0 |
| $z = 1$; occupied | $1 - p_{11}$ | $(1 - b) \times p_{11}$ | $b \times p_{11}$ |

Table 4.2 Values of $\Pr(y = \ell | z = k)$, $y = 0, 1$, $w = 0, 1$, and $z = 0, 1$

| True state | Uncertain $\Pr(y = 0 z)$ | Detection $\Pr(y = 1 z)$ | Certain $\Pr(w = 0 z)$ | Detection $\Pr(w = 1 z)$ |
|----------------------|-------------------------------|-------------------------------|-----------------------------|-------------------------------|
| $z = 0$; unoccupied | $1 - p_{10}$ | p_{10} | 1 | 0 |
| $z = 1$; occupied | $1 - p_{11}$ | p_{11} | $1 - r_{11}$ | r_{11} |

When two methods are used, the first may include false-positive detections but the second does not. For example, the first method could be auditory call surveys with some level of uncertainty and the second method direct handling under the assumption that species can be identified with certainty. Data obtained from the first detection are of the type envisioned by Royle and Link (2006) described above, while data from the second method follows the standard design of MacKenzie et al. (2002), where false positives are assumed not to occur.

The full detection probabilities for all state combinations are found in Table 4.2, with p specifying detection probabilities for the uncertain detection method and r specifying detection probabilities for the certain method. Here, p_{10} is the probability of a false-positive detection using the first method; p_{11} is the probability of a true-positive detection using the first method; and r_{11} the probability of a true-positive detection using the second method. When only one visit can be conducted using the certain method, for all parameters to be identifiable, at least two visits must be conducted using the uncertain detection method and vice versa, so that $T + T_1 > 3$. It is not necessary that both methods be employed at all sites. When sampling is incomplete, one assumes that the sites where the second detection method was used are representative of both false-positive and true-positive detection probabilities for all sites in the study. Computational details are given by Miller et al. (2011) in a series of appendices available online. All analyses were done using **R** (version 2.10.1). They also discussed the connection of their model with other models given in the literature, which are mainly special cases. Improvement using their models was greatest when occupancy was low, true-positive detection probabilities were small, and the number of sampling occasions was larger.

A novel idea is to use dogs that detect scat (Thompson et al., 2012). The use of scats can be effectively combined with DNA methods (e.g., Morin et al., 2018). Using DNA as a tag is discussed in Sect. 4.11 and in Sect. 5.5.10. Keeping et al. (2018) found that ground counts using tracks could be more useful for large herbivores than aerial censusing using distance methods.

4.10.1 Combining Detection Types

Miller et al. (2013) considered a similar problem of non-detection (false negatives) and misclassification (false positives) to that discussed in the previous section when

estimating occurrence parameters for gray wolves in northern Montana from 2007 to 2010. They considered the special case where detections can be divided into those that are certain (i.e., probability that a detection is a false positive is zero) and uncertain detections. Their primary data source for the analysis was observations by deer and elk hunters, reported as part of the state's annual hunter survey. This data was supplemented with data from known locations of radio-collared wolves. They extended previous methods which allow for false-positive errors in single-season data to models used to estimate occupancy dynamics across multiple seasons.

Two designs are mentioned. The first design, (a), is where either certain or uncertain detections can occur during a single sampling occasion (the multiple detection state model). For example, during an avian point-count survey, observers may consider visual observations of morphologically cryptic species uncertain because of the potential for misidentification, but auditory observations certain if the call is distinct. The second design, (b), occurs when only one observation type may occur during any given sampling occasion so that detections during a sampling occasion either are all uncertain or are all certain (multiple detection method model). For example, scat surveys can be regarded as uncertain due to potential species misidentification, but certain if individuals are handled in some way. In both designs, a site must be occupied for certain detections to be recorded, but there is some possibility when an uncertain detection is recorded that the site is actually unoccupied (i.e., false-positive detection).

The basic model used consists of three types of parameters: (1) the initial state distribution, (2) the between-season transition probabilities, and (3) the detection probabilities. In the simplest case with two occupancy states (where a site is either occupied or not occupied), the initial probability of a site being occupied in time 0 is denoted by ψ_0 , and the probability of being unoccupied by $1 - \psi_0$. The probability an unoccupied site in time t will be occupied in time $t + 1$ is γ_t , and the probability it will remain unoccupied is $(1 - \gamma_t)$. Similarly, the probability an occupied site in time t will be unoccupied in $t + 1$ is ε_t , and the probability it will remain occupied is $(1 - \varepsilon_t)$.

We first consider the case (a) with multiple detection states where both uncertain and certain detections can occur during a single sampling occasion. For unoccupied sites, by definition, certain detections do not occur, so that two possible observations can occur: an uncertain detection or no detection. The probability of a false-positive detection occurring for an unoccupied site is defined to be p_{10} , and the probability of no detection is $(1 - p_{10})$. For occupied sites, no detections, certain detections, and uncertain detections can occur. Let $1 - p_{11}$ denote the probability of not detecting the species. The probability that the detection will be certain, α , is conditional on detecting the species at an occupied site. The probability of an uncertain detection is $p_{11}(1 - \alpha)$ and of a (true-positive) certain detection is $p_{11}\alpha$.

Now consider the multiple detection method design (b) where individual sampling occasions will include either all certain detections or all uncertain detections. When the uncertain method is used, species will be detected at sites that are unoccupied with a false-positive detection probability p , while no detection will occur with probability $(1 - p)$. For occupied sites, the true-positive detection

probability is p , and the probability of a false-negative error is $(1 - p)$. When the certain method is used, the probability is 1 that no detections will occur for unoccupied sites. If the site is occupied, the probability of a true-positive detection is r and of not detecting the species is $(1 - r)$. We now show how to multiply up the probabilities for an example encounter history $h = 000, 101$ where a site is sampled on three occasions over two seasons. Here, the species was not detected at all during the first season, and uncertain detections were recorded in the first and third occasions during the second season. The probability of this is

$$\begin{aligned}\Pr(h = 000, 101) &= (1 - \psi_0)(1 - p_{10})^3(1 - \varepsilon_1)p_{10}^2(1 - p_{10}) \\ &\quad + (1 - \psi_0)(1 - p_{10})^3\varepsilon_1[p_{11}(1 - \alpha)]^2(1 - p_{11}) \\ &\quad + \psi_0(1 - p_{11})^3\varepsilon_1p_{10}^2(1 - p_{10}) \\ &\quad + \psi_0(1 - p_{11})^3(1 - \varepsilon_1)[p_{11}(1 - \alpha)]^2(1 - p_{11}).\end{aligned}$$

Because no certain detections occurred, it is possible that the site could have been either occupied or unoccupied in each of the two time periods so that the probability is the sum of the probabilities for each of the four possible state combinations: unoccupied in both seasons, unoccupied and then occupied, occupied and then unoccupied, or occupied in both seasons. In each case, the probability calculation is the product of (1) the probability of being in the initial state, (2) the probability of observing a set of detections conditional on the starting state, (3) the probability of being in a state in the second season conditional on the initial state, and (4) the probability of observing a set of detections conditional on the state of the site in the second season. To calculate the probabilities of encounter histories that include additional seasons involves iterating the (3) and (4) steps for each additional season.

As a further example, consider another site where the observed encounter history is $h = 210, 011$, that is, both certain and uncertain detections occurred during the first season, but only uncertain detection occurred in the second. This has probability

$$\begin{aligned}\psi_0(p_{11}\alpha)[p_{11}(1 - \alpha)]^2(1 - p_{11})\varepsilon_1p_{10}^2(1 - p_{10}) \\ + \psi_0(p_{11}\alpha)[p_{11}(1 - \alpha)]^2(1 - p_{11})(1 - \varepsilon_1)[p_{11}(1 - \alpha)]^2(1 - p_{11}).\end{aligned}$$

Because a certain detection occurred during the first season, we have only two possibilities for the true state of the site over the two seasons. The first possibility is that the site was occupied in the first period but transitioned into being unoccupied in the second. In this case, detections during the second season would be false positives. Alternatively, the site could have been occupied in both seasons. Further examples are given involving two detection methods employed on separate occasions where both certain and uncertain methods are used. Covariates can also be incorporated such as, for site s and time t ,

$$\text{logit}(\varepsilon_{st}) = \boldsymbol{\beta}' \mathbf{x}_{st}.$$

The authors considered a general model where there are more than two occupancy states and several possible observation states (i.e., the set of discrete observations that can be made during a visit to a site). Here, we have S sites monitored for T seasons with the s th site visited T_{st} times during the t th season. The true occupancy state of the s th site in the t th season, z_{st} , is one of the K discrete occupancy states. Observations of the s th site on the r th visit in the t th season, y_{srt} , are classified into one of the L observation states that differ in the probability of being a false-positive detection. The probability of recording an observation conditional on the true occupancy state is $\pi_{srt} = \Pr(y_{srt} = \ell | z_{st} = k)$.

A second sampling method can also be incorporated. Because we have such a general model, it is not surprising that it includes a number of previous models as special cases. As with such complicated models, the way forward is to use matrices. Here, the likelihood of the full set of parameters, given the full set of encounter histories, can be formulated using the general likelihood for multi-season occupancy and involves three components: the initial state distribution, transition probabilities among states, and state-specific detection probabilities. The methods were applied to wolves, and all models were fitted using **PRESENCE** v 3.1 (Hines et al., 2010). Details are given by Miller et al. (2013).

Falke et al. (2012) used an integrated habitat occupancy model to simultaneously quantify habitat change, site fidelity, and local colonization and extinction rates for larvae of a suite of Great Plains stream fishes in the Arkansas River, eastern Colorado, USA, across 3 years. The model had three states, dry (and thus unoccupied (0)), wet but species not detected (1), or wet and species detected (2).

Miller et al. (2015) presented the most comprehensive test of occupancy estimation methods at the time, using more than 33,000 auditory call observations collected under standard field conditions and where the true occupancy status of sites was known, a rare situation. Basic occupancy estimation approaches were biased when two key assumptions were not met, namely, no false positives occur (e.g., no recording of an individual as present when absent) and that no unexplained heterogeneity in detection parameters occurs (e.g., probability of detecting a species given the site is occupied does not vary with site location). The greatest bias occurred for dynamic parameters (i.e., local colonization and extinction), and in many cases, the degree of inaccuracy would render results largely useless. They found that even small departures from the two assumptions could have severe effects on parameter estimation.

One of the problems of having temporal replication at sampling sites is that it is time-consuming and costly. Gómez-Rodríguez et al. (2012) considered a lower-cost alternative based on a double sampling procedure. Basically, a number of sites were surveyed once, then some of these had repeated surveys to estimate detection probabilities, and this information was applied to sites surveyed only once. This means selecting those absence records that really represent sites not occupied by the species. The sampling unit was a pond, and amphibian species were investigated. The models used and further descriptive details are given by the authors.

4.10.2 Multiple Detection Devices

Detection can be carried using several different methods or devices. For example, Nichols et al. (2008) considered multiple detection models to estimate parameters associated with both large- and small-scale occupancy. They used a sampling design in which K different detection or sampling devices are deployed at each of S sample sites (stations) for T occasions (visits). Each station samples a sample unit, which may be larger in area than the station itself. Let

$$p_t^{(k)} = \Pr(\text{detection at occasion } t \text{ by device } k \mid \text{sample unit occupied and}$$

species present at immediate sample station site)

$$\psi = \Pr(\text{sample unit occupied})$$

$$\theta_t = \Pr(\text{species present at immediate sample site at occasion } t \mid \text{sample unit occupied})$$

The two occupancy parameters, ψ and θ_t , permit the modeling of occupancy at two different spatial scales. The basic occupancy parameter, ψ , corresponds to species occurrence at the larger scale, where member(s) of the species have some non-negligible probability of being present at the sample unit and hence exposed to the detection devices. It applies to all of the sample occasions.

The occupancy parameters for the smaller scale, θ , refer to the presence of member(s) of the target species at the sample station site where the detection methods are collocated, conditional on species presence in the sample unit. The product $\psi\theta_t$ represents the probability of small-scale occupancy, indicating presence of individual(s) of the species at the local site exposed to detection devices at sampling occasion t . The product $\psi(l - \theta_t)$ indicates occupancy at the large scale, but not at the small scale, for sampling occasion t . The species may be temporarily unavailable for detection because members are not in the immediate vicinity of the detection devices.

As an example, consider a study with $K = 3$ detection devices and $T = 2$ sampling occasions, with the species of interest detected by devices 2 and 3 at sampling occasion 1 and only by device 2 at sample occasion 2. Then

$$\Pr(h_s=010, 000) = \psi \left[\theta_1(1 - p_1^{(1)})p_1^{(2)}(1 - p_1^{(3)})[(1 - \theta_2) + \theta_2 \prod_{k=1}^3 (1 - p_2^{(k)})] \right].$$

The sum within the interior brackets specifies the two possible ways of obtaining no detections (000) during occasion 2. The first term, $(1 - \theta_2)$, is the probability that the species was not locally present at the sample site on occasion 2. The second term, $\theta_2 \prod_{k=1}^3 (1 - p_2^{(k)})$, denotes the probability that a member(s) of the species

was present at the sample site at occasion 2 but simply not detected by any of the devices.

As a further example, consider the detection history for a site with no detections by any device at either sample occasion, 000 000. The probability associated with this history can be written as

$$\begin{aligned} \Pr(h_s = 000, 000) &= (1 - \psi) + \psi(1 - \theta_1)(1 - \theta_2) + \psi\theta_1(1 - \theta_2) \prod_{k=1}^3 (1 - p_1^{(k)}) \\ &\quad + \psi(1 - \theta_1)\theta_2 \prod_{k=1}^3 (1 - p_2^{(k)}) + \psi\theta_1\theta_2 \prod_{k=1}^3 (1 - p_1^{(k)})(1 - p_2^{(k)}). \end{aligned}$$

Nichols et al. (2008) described how this is put together. Here, h_s is the detection history at site s , and the likelihood from all the data takes the form

$$L(\psi, \boldsymbol{\theta}, \mathbf{p}; \mathbf{H}) \propto \prod_{s=1}^S \Pr(h_s),$$

with the possibility of incorporating covariates with the parameters. Estimates of ψ and \mathbf{p} can be found by collapsing the detection data from the three devices at each sampling occasion to reflect detection by at least one device or non-detection by all devices. For example, the detection history (010,000) would simply be collapsed to (10). The first three device-specific detection entries are collapsed to 1 for the first occasion, while the second three entries are collapsed to 0 for the second occasion. The authors applied the method to striped skunk (*Mephitis mephitis*) and red salamanders (*Pseudotriton ruber*) and compared various submodels.

4.11 DNA Methods

With the technological advances in genetics, DNA methods are becoming more extensively used with the availability of genetic tags and markers (cf. the review by Luikart et al., 2010). We mention just a few applications relevant to this chapter. For example, Fisher et al. (2016) used the model for several devices (Sect. 4.2.2) for grizzly bears to estimate missed detections when an animal does not leave a hair sample with viable DNA by sampling the hair trap and the area around it using camera trapping. Although the authors used the term “occupancy” for consistency, occurrence at a site for mobile animals should be interpreted as “site use” rather than permanent residence. Here, there were $k = 2$ sampling devices and $t = 8$ sampling occasions. They used a systematic sampling design consisting of 10-km \times 10-km grid cells, plotted on a digital landscape, and within each cell, they deployed a fixed sampling site that remained in place for the season. Sampling sites consisted of a

hair trap—a scent-lured tree wrapped with barbed wire (noninvasive genetic tagging or NGT) and a camera trap placed 6–10 m away to image the hair trap and the surrounding area. Animals at the hair trap were fully exposed to the camera trap.

For mobile animals, it was assumed that a species available for sampling has a non-zero probability of being present at the sample unit within the sampling period. The detection area of the cameras was greater than point detection at the hair trap; barring camera failures (treated as missing data), there were no occasions where a bear was sampled at a hair trap without being sampled by a camera. This differs from the (Nichols et al., 2008) scenario wherein either device could fail to detect a species at a site.

One problem we look at is to how to account for and estimate genetic detection errors. Fisher and Bradbury (2014) provided a model using a two-method approach to estimating the detection error during hair trapping. Hair trapping for NGT requires that an animal (1) encounter a site, (2) interact with a trap, and (3) leave a hair sample from which DNA is extracted. Each step inherently has variability and a potential for error. They referred to the DNA tagging as noninvasive genetic tagging (NGT) and combined the method with camera trapping. The authors surveyed marten (*Martes americana*), fisher (*Pekania pennanti*), and wolverine (*Gulo gulo*) occurrence via hair trapping and camera trapping for 3 repeated monthly surveys at 66 sites in the Rocky Mountains of Alberta, Canada. Camera traps photographed the hair traps and surrounding sampling site. Although details are not given of the mathematical models used, capture-recapture analysis of “marked” individuals accounted for false absences to estimate the present but undetected component of a sample population, while detection-non-detection data from unmarked animals can be analyzed in an occupancy framework.

4.11.1 eDNA

Genetic tags using environmental DNA (eDNA), a noninvasive method, are being increasingly used along with metabarcoding in population studies, particularly with the detection of rare species (e.g., Wilcox et al., 2013) and invasive species (e.g., Takahara et al., 2013, invasive fish), and estimating biodiversity (e.g., Stauffer et al., 2021, investigated how many replicates are needed to accurately estimate fish biodiversity using environmental DNA on coral reefs). With fisheries, DNA fragments can be filtered from the water to determine species’ presence. Evans et al. (2016) noted that: “Because eDNA persists in water for days to weeks after organisms are removed from controlled experimental systems, it contains a catalog of species present in the recent past.” Degradation of DNA is an important issue. For example, Thomsen et al. (2012) found that DNA becomes undetectable within 2 weeks after removal of animals, indicating that DNA traces are near contemporary with presence of the species.

Environmental DNA is DNA extracted from an environmental sample without isolating the target organism and has been characterized as a mixture of genomic

DNA from many different organisms, which is often degraded into small fragments. Short DNA fragments allow the use of degraded DNA from environmental samples. Smart et al. (2015) demonstrated that eDNA sampling can be an order of magnitude more sensitive than traditional methods and illustrated that traditional- and eDNA-based surveys can provide independent information on species distributions when occupancy surveys are conducted over short timescales. They modeled the detection probabilities of trapping- and eDNA-based sampling methods using hierarchical generalized linear models and incorporating temporal and spatial random effects. The models were fitted using Markov chain Monte Carlo Bayesian modeling software **JAGS** v.3.4.0. An offshoot of eDNA is iDNA which is invertebrate-derived DNA where terrestrial vertebrates can be detected via their DNA that was ingested by invertebrates such as leeches. They also combined leech iDNA and camera trap data in a single occupancy model.

Rees et al. (2014) reviewed the eDNA literature for aquatic species. Freshwater ecosystems are among the most endangered habitats on Earth, with thousands of animal species known to be threatened or already extinct. Jerde et al. (2011) noted that detection probabilities, particularly for rare species, are low in aquatic environments where organisms are hidden beneath the water's surface. Fish surveillance programs traditionally employ nets or electrofishing gear, which usually have low capture probabilities per target organism, so that they are reliable indicators of occurrence only for species present at moderate-to-high abundance. Darling and Blum (2007) and Dejean et al. (2012) considered invasive species, and the latter authors, using bullfrogs as an example, found that the eDNA approach permits the early detection of alien invasive species at very low densities and at any life stage. This is particularly important for the detection of rare and/or secretive aquatic species. They noted that the method is valuable for species detection and surpasses traditional amphibian survey methods in terms of sensitivity and sampling. Some problems using DNA tags generally are discussed in Seber and Schofield (2019, 17–20), though some problems will be ameliorated with advances in the technology and the use of on-site methods. We now consider how DNA can be incorporated into some of the above methods.

Ficetola et al. (2015) considered the effect of replication levels using eDNA metabarcoding data and applying standard presence/absence models. They recommended that at least eight replicates should be performed if detection probability is not high, such as in ancient DNA studies. Environmental DNA (eDNA) is prone to both false-positive and false-negative errors, particularly false positives where a species may be apparently detected in areas when it is actually absent. Lahoz-Monfort et al. (2016) reviewed statistical methods to account for such errors in the analysis of eDNA data and used simulations to compare the performance of different modeling approaches. Ficetola et al. (2016) considered the importance of controlling for false detection from the early steps of eDNA analyses. They asked the question: “How many positives do we need to avoid stating that a species is present at a site where it is actually absent?”

Let p_{10} be the probability of a false detection, p_{11} the probability of a true detection, and K the number of PCR (polymerase chain reactions) analyses on one

sample. If we consider that a taxon is present if detected in at least i PCR replicates from the same sample, then for x false positives,

$$\Pr(x \geq i \mid \text{site not occupied}) = \sum_{j=i}^K \binom{K}{j} p_{10}^j (1 - p_{10})^{K-j}.$$

Given a value of p_{10} , we can use the above equation to determine K to satisfy a given value of the left-hand side probability. Alternatively, we can compute the probability that the species is present, given the observed number of detections (x) and the estimated probabilities of occupancy (ψ), true detection (p_{11}), and false detection (p_{10}). This probability is (Lahoz-Monfort et al., 2016, eq1)

$$\Pr(\text{site occupied} \mid x) = \frac{\psi p_{11}^x (1 - p_{11})^{K-x}}{\psi p_{11}^x (1 - p_{11})^{K-x} + (1 - \psi) p_{10}^x (1 - p_{10})^{K-x}}.$$

Schmidt et al. (2013) developed a three-level hierarchical model using the concept of components of the detection probability as introduced earlier and in Sect. 7.8. It is similar to that used by Nichols et al. (2008) in Sect. 4.10.2 with a change in the definitions of the parameters and to using eDNA. Let $z_s = 1$ if the species is present with probability ψ_s at site (pond) s and 0 otherwise; let $a_{sj} = 1$ if DNA of the species is present in water sample j from site s is with probability θ_{sj} and 0 otherwise, given the species is present; and let $y_{sjk} = 1$ with probability p_{sjk} if the species is present in the PCR analysis k , given the DNA is present in sample j at site s . We thus have

$$\begin{aligned} z_s &\sim \text{Bernoulli}(\psi_s) \\ a_{sj} \mid z_s &\sim \text{Bernoulli}(z_s \theta_{sj}) \\ y_{sjk} &\sim \text{Bernoulli}(a_{sj} p_{sjk}). \end{aligned}$$

With the detection probability, it is assumed that there are no false positives. This type of model with components of the detection probability is discussed in more detail in Sect. 7.8, and a slight variation without covariates was used by Hunter et al. (2015) to study invasive Burmese pythons. For a list of references and an extensive example of using eDNA, see Buxton et al. (2021).

Smith and Goldberg (2019) considered the use of eDNA in accounting for multiples scales and false positives and noted that most past models assumed that no false positives occurred, which is not the case at the PCR replicate level. However, they used so-called qPCR replicates, a site confirmation approach that relies on an additional detection method (that does not have false positives) and can be used to estimate the rate of false positives. The above notation of Schmidt et al. (2013) was used for their model, except combining availability within the filter sample and qPCR detection into a single detection parameter p_{sjk} . This probability is now the detection probability given that eDNA is in the filter sample and is amplified by

qPCR. They estimated the probability of false positives at the qPCR replicate level using a method of Miller et al. (2011) described above for handling two types of data.

Doi et al. (2019), using environmental DNA (eDNA) metabarcoding along with a multispecies approach, described how to incorporate false-negative errors into species detection. The detection probabilities were modeled using logistic regression and an analysis of variance-type format with interactions as well as using prior distributions for some of the parameters. They used the Markov chain Monte Carlo (MCMC) method in JAGS software version 4.2.038 to obtain samples from the posterior distribution of parameters. Posterior samples were obtained from 3 independent chains of 100,000 iterations after a burn-in of 100,000 thinning at intervals of 100. The convergence of MCMC was affirmed by determining if the appropriate \hat{R} statistic for each parameter of interest was less than 1.1.

4.12 Presence-Only Data

Explaining the distribution of a species by using local environmental features is a long-standing ecological problem. Often, available data are collected as a set of presence locations only, thus precluding the possibility of a desired presence-absence analysis. This can happen, for example, when museum records, personal records, unpublished reports, and the published literature are used. Presence-only records may provide data on the distributions of rare species, but commonly suffer from large, unknown biases due to their typically haphazard collection schemes. This topic will be continued in discussing citizen science data in Sect. 4.16.7.

Presence-only data can provide a point pattern over a region, and then local environmental features can be used to explain what is driving this point pattern. Its use arises in many places in the rest of this chapter. Phillips et al. (2006) introduced the use of the maximum entropy method **MAXENT** for modeling species geographic distributions with presence-only data. It estimates the density of environmental covariates conditional on species presence. Fithian and Hastie (2013) showed that there are close links between **MAXENT**, the inhomogeneous Poisson process (IPP), and logistic regression for presence-only data.

4.12.1 Pseudo-Absence Problem

The above topic of presence-only data is related to the so-called pseudo-absence problem where artificial absence data are created when absence data are missing. Here, a random sample of locations is taken from the region of interest, and although the presence or absence of a species at the background locations is unknown, they provide a summary of the landscape, against which one can compare the observed presences. Given just the background information from a site, we do not know

whether a species is present or absent at the site, and this can be regarded as a missing observation. This lends itself to using the EM algorithm (Ward et al., 2009).

We note that creating pseudo-absence data can be a problem with rare and endangered species, and Engler et al. (2004) and Hazen et al. (2021) describe methods for doing this. Barbet-Massin et al. (2012), using simulation, used a broad-brush approach to providing guidelines on how, where, and how many pseudo-absences should be generated to build reliable species distribution models. Mathematical aspects of the models used are not given, and the reader is referred to their paper for further details. They noted that there is evidence that presence-absence models tend to perform better than just presence-only models, though presence-absence models are increasingly used when only presence data is available by creating pseudo-absences. Chakraborty et al. (2011) argued that presence-only data are not inferior to presence-absence data, as presence-only data offer a complete census, whereas presence-absence data, since confined to a specified set of sampling sites, contain less information.

4.12.2 Poisson Point Process Models

Warton and Shepherd (2010) considered a Poisson point process for dealing with presence-only data. Assuming no species detection errors, suppose we have a population area \mathcal{A} of area A , where a species is observed to be present at n locations with ℓ_i being the location of the i th location, and ℓ is the vector of the ℓ_i . Then the ℓ_i may be modeled as a realization of an inhomogeneous Poisson process (Cressie, 1993) parameterized by a first-order intensity function $\lambda_i = \lambda(\mathbf{x}_{\ell_i}) = \exp(\boldsymbol{\beta}' \mathbf{x}_{\ell_i})$ that includes the covariate measurements \mathbf{x}_{ℓ_i} at location ℓ_i and the parameters $\boldsymbol{\beta}$. Here, $\lambda(\mathbf{x}_\ell)$ denotes the limiting, expected density of individuals (the number of individuals of the species per unit area). From Cressie (1993, 651), the log-likelihood function for the point process model is

$$\log L(\boldsymbol{\beta} \mid \ell) = \sum_{i=1}^n \boldsymbol{\beta}' \mathbf{x}_{\ell_i} - \mu_A - \log(n!),$$

where $\mu_A = \int_{\mathcal{A}} \lambda(\mathbf{x}_\ell) d\ell$, the mean intensity of the point process over region \mathcal{A} . Then, using numerical quadrature, $\mu_A \approx \sum_{i=1}^m w_i \lambda_i$, and the above likelihood is approximately proportional to a weighted Poisson likelihood

$$L_{\text{ppm}}(\boldsymbol{\beta}; \ell, \ell_0, \mathbf{w}) = \sum_{i=1}^m w_i (z_i \log(\lambda_i) - \lambda_i),$$

where $z_i = I(i \in [1, \dots, n]) / w_i$, $\ell_0 = (\ell_{n+1}, \dots, \ell_m)'$ are the quadrature points, the vector $\mathbf{w} = (w_1, \dots, w_m)'$ stores all the quadrature weights, and $I(\cdot)$ is the

indicator function. The authors also introduce a zero term into the regression so that

$$\log(\lambda_i) = \beta_0 + \sum_{j=1}^k x_{ij}\beta_j$$

and $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_k)'$. Quadratic expressions were also used for $\log(\lambda_i)$.

Warton and Shepherd (2010) showed that pseudo-absence logistic regression and Poisson point process models are asymptotically equivalent for a large number of pseudo-absences and only for pseudo-absences that are either regularly spaced or located uniformly at random over \mathcal{A} . This was done by redefining ℓ_0 as a set of pseudo-absence points where m is large. As the number of pseudo-absences increases, all parameter estimators except for the intercept in the logistic regression model converge to the maximum likelihood estimators of the Poisson process model. They also considered the questions of how to choose quadrature points and how to calculate the quadrature weight w_i at each point ℓ_i .

Warton et al. (2013) noted that when there is just presence data available, it can be subject to bias due to observers being more likely to visit and record sightings at some locations than others. They used the point process model and an alternative, model-based bias correction approach especially for species distributions. It involved two steps, namely, modeling the observer bias and then conditioning on a common level of observer bias at all locations to predict species distributions.

4.12.3 Geographic Distribution

Dorazio (2012) mentioned several models for presence-only surveys that have been developed to predict the geographic distribution of a species. This is done by combining measurements of covariates of occurrence at locations where the species is known to be present with measurements of the same covariates at other locations where species occurrence status (presence or absence) is unknown. Using the methodology of the previous paragraphs, Dorazio (2012) briefly described how $\mu_A = \int_A \lambda(\mathbf{x}_\ell) d\ell$ mentioned above can be approximated using numerical quadrature. If $N(A)$, the number on area A , comes from the Poisson point process, then

$$\Pr(N(A) = n) = \exp[-\mu_A] \frac{\mu_A^n}{n!}.$$

He also linked up his approach with what he calls CA binary regression models, which are formulated for samples of two distributions, the conditional distribution of the covariates at presence-only locations ($f(\mathbf{x} | z = 1)$) and the unconditional distribution of the covariates at all locations ($f(\mathbf{x})$, where $z = 1$ if the individual is

present and 0 otherwise. He established the asymptotic equivalence of the Poisson and CA models. Using Bayes' rule,

$$g(\mathbf{x} | z = 1) = \psi(\mathbf{x}, \boldsymbol{\alpha}) f(\mathbf{x}) / C(\boldsymbol{\alpha}),$$

where $\psi(\mathbf{x}, \boldsymbol{\alpha}) = \Pr(z = 1 | \mathbf{x})$ is the conditional probability of species presence given the covariate measurements \mathbf{x} and parameters $\boldsymbol{\alpha}$. Also, $C(\boldsymbol{\alpha})$ is the normalizing constant for $g(\mathbf{x} | z = 1)$ to be a valid distribution, namely,

$$C(\boldsymbol{\alpha}) = \int \psi(\mathbf{x}, \boldsymbol{\alpha}) f(\mathbf{x}) d\mathbf{x} = \Pr(z = 1),$$

the unconditional probability of species presence. Further details are given by Dorazio (2012), who also considered the influence of detection errors on the parameter estimates, especially relating to covariates.

Dorazio (2014) discussed a spatial model relating to geographic distributions but using two kind of observations: (1) detections of individuals encountered during opportunistic and citizen science surveys (i.e., presence-only observations) and (2) detections of individuals encountered during planned survey locations selected using a prescribed sampling design. Further details are given there.

4.12.4 Using Environmental Conditions

Chakraborty et al. (2011) aimed to explain the probability of presence at a location given the environmental conditions that are present there. A generalized additive binary regression model was used with, say, a logistic link where the covariates can be introduced linearly or as smoothly varying functions. One model-based strategy for presence-only data has attempted to implement a presence-absence approach by drawing so-called background samples. These are random samples of locations in the region with known environmental features, and these have been characterized as pseudo-absences that create an unobserved and artificial pattern of absences.

The authors commented that a major problem with methods that have been used is that of bias in sample selection, which can be affected by variation in site access and what they referred to as “degradation.” For example, there are sites that are adjacent to roads or along paths, near urban areas with public ownership, or with flat topography that are likely to be oversampled relatively to more inaccessible sites. Human intervention and agricultural transformation can preclude availability. Obtaining a random sample of available environments can therefore be tricky, especially as sampling tends to be sparse and irregular. The authors used an inhomogeneous Poisson process for the set of presence locations. Because of mathematical details, we take just a brief look at their model, which provides a probability model for presence locations. Their model approach, which is somewhat similar to that used by Gelfand et al. (2005), we now sketch.

They began with imagining three surfaces over the domain \mathcal{D} of interest. The first, denoted by $\lambda(\mathbf{s})$ (the intensity in the absence of degradation), is called the “potential” intensity surface and is a positive function integrable over \mathcal{D} , namely, $\lambda(\mathcal{D}) = \int_{\mathcal{D}} \lambda(\mathbf{s}) d\mathbf{s}$, where \mathbf{s} represents the coordinates of a point. Then, $g(\mathbf{s}) = \lambda(\mathbf{s})/\lambda(\mathcal{D})$ gives the potential density over \mathcal{D} . A second imagined surface is an availability surface $U(\mathbf{s})$, which is a binary surface over D such that $U(\mathbf{s}) = 1$ if location s is untransformed by land use (i.e., assuming no sampling bias) and 0 otherwise. Finally, we imagine a sampling effort surface over \mathcal{D} denoted by $T(\mathbf{s})$, also a binary surface, and $T(\mathbf{s})U(\mathbf{s}) = 1$ indicates that location s is both available and sampled. If $|A_i|$ is the area of A_i , they then set $q_i = \int_{A_i} T(\mathbf{s})U(\mathbf{s}) d\mathbf{s}/|A_i|$, which is interpreted as the probability that a randomly selected location in A_i was available and sampled. Then $\lambda(\mathbf{s})U(\mathbf{s})T(\mathbf{s})$ becomes the degradation at location \mathbf{s} , which implies that in regions where no locations were sampled, the operating intensity for the species is 0. Given n_i presence locations $(s_{i1}, s_{i2}, \dots, s_{in_i})$ within cell (site) i , $U(s_{i,j})T(s_{i,j}) \equiv 1, 0 \leq j \leq n_i, 1 \leq i \leq I$, and it transpires that the likelihood function for the inhomogeneous Poisson process $\lambda(\cdot)$ is

$$L(\lambda(\cdot); \{s_{i,j}\}) \propto \exp \left[- \int_{\mathcal{D}} \lambda(\mathbf{s}) U(\mathbf{s}) T(\mathbf{s}) d\mathbf{s} \right] \prod_{i=1}^I \prod_{j=1}^{n_i} \lambda(s_{i,j}).$$

The authors used covariates and a Bayes approach using Markov chain Monte Carlo (MCMC) methods. The theory was also extended to the problem of estimating richness with presence-only data. Computational details are given in an appendix.

In modeling species distributions, Fithian and Hastie (2013) argued that the inhomogeneous Poisson point process (IPP) mentioned previously, in comparison to maximum entropy (**MAXENT**) and logistic regression along with its various generalizations, is a more natural method of inference in presence-only studies than occurrence probability (which they argue is only defined with reference to quadrat size). They described the strong connections between the three methods (cf. Section 5 of their paper for a summary), where the asymptotic equivalence of IPP and logistic regression has already been described above. Renner and Warton (2013) proved the equivalence of **MAXENT** and IPP. Fithian and Hastie (2013) argued that “a presence-only data set only affords us direct knowledge of the expected number of specimen sightings per unit area. The absolute sightings rate is reflected in the number of records in our data set, but, at best, this rate is only proportional to the occurrence rate...” Also, “Using presence-only data alone, we can at best hope to estimate a relative, not absolute, occurrence rate.”

A related problem is considered by Phillips et al. (2009) and Phillips and Elith (2013) where, as above, the aim is to estimate the probability that a species is present in (or uses) a site, conditional on environmental covariates termed background or presence-background (PB) data. Guillera-Arroita (2017), in a helpful review of the general subject area, made the comment that PB data do not contain information about sampling effort, so the data is very susceptible to estimation biases induced by sampling bias (Phillips et al., 2009). Furthermore, they cannot robustly quantify

prevalence or probabilities of occurrence (Hastie & Fithian, 2013; Phillips & Elith, 2013); from such data, one cannot tell whether few species records are due to species rarity or due to little survey effort. However, with the addition of species absence records, more information is available about sampling effort so that any inference is much more robust to biases in sampling.

4.12.5 Methods of Phillips & Elith, 2013

Let L be the landscape of interest and L_1 and L_0 be the subsets of L , in which the species is present or absent, respectively. Let P be a set of presence samples (drawn uniformly from L_1) and $|P|$ the number and B a set of background samples (drawn uniformly from L) with $|B|$ the number. Let y represent the presence ($y = 1$) or absence ($y = 0$) of species and s represent the sampling stratum; $s = 1$ for samples in P and $s = 0$ for samples in B . Assume that

$$\Pr(y = 1 \mid \mathbf{x}; \boldsymbol{\beta}) = \frac{1}{1 + \exp[-\eta(\mathbf{x})]},$$

where $\eta(\mathbf{x})$ is a function of the set of predictor variables, \mathbf{x} , described by a set of parameters, $\boldsymbol{\beta}$. For convenience, it is assumed that η is just a linear function with coefficients $\boldsymbol{\beta}$. Phillips and Elith (2013) described existing methods for determining parameters $\boldsymbol{\beta}$, and these are given below. Three of the methods (EM, SC, and SB) require, as an additional input, an estimate of the prevalence $\Pr(y = 1)$, denoted by π .

The LK Method

This method (Lele & Keim, 2006; Lele, 2009) defines the log-likelihood of the presence samples as

$$\sum_{\mathbf{x} \in P} \log[\Pr(\mathbf{x} \mid y = 1; \boldsymbol{\beta})].$$

Applying Bayes' rule to the above equation and dropping terms that do not depend on $\boldsymbol{\beta}$ give

$$\sum_{\mathbf{x} \in P} \log \left[\frac{\Pr(y = 1 \mid \mathbf{x}; \boldsymbol{\beta})}{\Pr(y = 1; \boldsymbol{\beta})} \right] \quad (4.2)$$

or

$$\sum_{\mathbf{x} \in P} \log[\Pr(y = 1 \mid \mathbf{x}; \boldsymbol{\beta})] - |P| \log[\Pr(y = 1; \boldsymbol{\beta})]. \quad (4.3)$$

The background data are used to give an empirical estimate of the second term, resulting in the expression

$$\log[\Pr(y = 1 \mid \mathbf{x}; \boldsymbol{\beta})] - |P| \log \left[\frac{\sum_{\mathbf{x} \in B} \Pr(y = 1 \mid \mathbf{x}; \boldsymbol{\beta})}{|B|} \right],$$

where the prevalence (the average probability of presence over the whole landscape) has been approximated by the average probability over the background samples. Optimization techniques can be used to find the maximum likelihood estimate of $\boldsymbol{\beta}$ for the above equation.

Equation (4.2) can be thought of as the log of a relative likelihood as it describes the probability of the presence records (in the numerator) relative to the probability averaged over all sites (i.e., the prevalence, in the denominator). If we multiply all probabilities by a constant, that constant will cancel out of the ratio. Hence, the LK method measures only relative probability of presence. Lele and Keim (2006), in extending the work of Johnson et al. (2006), discussed this problem, which can arise with exponential probability functions and gave situations when we can have an absolute probability, for example, when the covariates are not all categorical. Phillips and Elith (2013) say that there is a need for an additional condition, namely, that the species' response must be nonconstant. They also mention how an EM algorithm (method EM) can be used, though $\pi = \Pr(y = 1)$ is needed.

The SC Method

The SC method (Steinberg & Cardell, 1992) uses the fact that the log-likelihood can be written in a way that the sums are either over all of L or over L_1 , namely,

$$\begin{aligned} \sum_{\mathbf{x} \in L_0} \log[1 - \Pr(y = 1 \mid \mathbf{x}; \boldsymbol{\beta})] + \sum_{\mathbf{x} \in L_1} \log[\Pr(y = 1 \mid \mathbf{x}; \boldsymbol{\beta})] = \\ \sum_{\mathbf{x} \in L} \log[1 - \Pr(y = 1 \mid \mathbf{x}; \boldsymbol{\beta})] \\ + \sum_{\mathbf{x} \in L_1} \{\log[\Pr(y = 1 \mid \mathbf{x}; \boldsymbol{\beta})] - \log[1 - \Pr(y = 1 \mid \mathbf{x}; \boldsymbol{\beta})]\}. \end{aligned}$$

The two sums in the above equation can be estimated empirically using the samples B and P , respectively, in a similar fashion to the SK model. Combining terms and simplifying yield a pseudo-likelihood likelihood that approximates the log-likelihood, namely,

$$\sum_{\mathbf{x} \in B} -\log(1 + \exp[\eta(\mathbf{x})] + \pi |B| \eta(\boldsymbol{\mu}),$$

where $\boldsymbol{\mu}$ is the vector of means of predictor variables over P . The above equation can be maximized using standard numerical methods.

The SB Method

The scaled binomial loss (SB) method applies existing logistic model-fitting methods using a modification to the standard binomial loss function in order to deal with presence-background data (Phillips & Elith, 2011). Let $f_p = |P|/(|P| + |B|)$ be the fraction of presence samples in the training data. Following Lancaster and Imbens (1996), the presence and background data can be regarded as having been generated together as follows: each sample is drawn from L_1 with probability f_p (a presence sample) and uniformly from L with probability $(1 - f_p)$ (a background sample). Recalling that $s = 1$ for samples in B , under this use-availability model, let

$$P_U(s = 1 \mid \mathbf{x}) = \frac{1}{1 + r + \exp[-\eta(\mathbf{x}) + \log r]}, \quad (4.4)$$

where

$$r = \frac{(1 - f_p)}{f_p} \pi.$$

The SB method finds the parameters of η that maximize the likelihood of the presence and background data as described by (4.4) using an appropriate logistic regression package.

The LI Method

This method from Lancaster and Imbens (1996) is also based on (4.4), but regards the species' prevalence π as a parameter to be estimated rather than a user-supplied value. General-purpose nonlinear optimization software can be used to simultaneously estimate values for π and β according to (4.4).

Phillips and Elith (2013) compared various special cases of the above models using simulation and concluded that there is a big difference between the methods that make a strong parametric assumption (LI and LK) and those that take the species' prevalence as a parameter (SC, EM, and SB). The LI and LK methods had RMS (root mean square) errors that were greater than those of the other methods for all simulated species and approximately ten times greater for all but one of the simulated species. They also failed to adequately estimate the species' true probability of presence from presence-background data. However, they did use less data than the other methods. Differences in performance within the groups were minor. The authors recommended the use of the second group even though they require a user-supplied estimate of prevalence.

In discussing the PB method, Guillera-Arroita (2017), in a helpful review of the general subject area, makes the comment that PB data do not contain information about sampling effort, so the data is very susceptible to estimation biases induced by sampling bias (Phillips et al., 2009). Furthermore, they cannot robustly quantify prevalence or probabilities of occurrence (Hastie & Fithian, 2013; Phillips & Elith, 2013); from such data, one cannot tell whether few species records are due to species rarity or due to little survey effort. However, with the addition of species absence

records, more information is available about sampling effort so that any inference is much more robust to biases in sampling.

4.12.6 Combining Presence and Presence/Absence Data

Elith et al. (2006) used a method in which presence-only data can be used to fit the model and then one can check on its prediction ability by using presence-absence information as well. Using data from plants, mammal, and reptiles from six global regions, they set up two sets of data for each region: (1) presence-only data from unplanned surveys or incidental records, including those from museums and herbaria, and (2) independent presence-absence data from planned surveys with accurate location records. Two important features of the study were the independence of the two data sets and the totally different data collection methods for the two sets. Sixteen methods of modeling the distributions were applied to 226 species. Some tidying up and pruning of the data was carried out. As some of the modeling methods required data akin to absences, background samples (pseudo-absences, discussed previously in Sect. 4.12.4) were generated by drawing a random sample of 10,000 sites for each region. These were intended as a sample of the whole region, and it is possible that a background sample coincided with a presence record.

Two groups of methods were considered. The first group, which uses only presence records, included a so-called envelope-style method (BIOCLIM) and two distance-based methods (DOMAIN and LIVES). The second group included a number of regression approaches such as generalized linear models (GLMs) and generalized additive models (GAMs). A large number of other computer packages were also applied. **WinBUGS** was used to fit a Bayesian model with uninformative priors for all parameters. Mathematical details are not given, though a separate appendix is available from the journal website.

Guisan et al. (2007) used the same extensive data set as that above and presence-only data to compare ten modeling techniques in terms of predictive power and sensitivity to location error, change in map resolution, and sample size. They also assessed whether some species traits can explain variation in model performance. The study focused on 30 native tree species in Switzerland and used presence-only data to model current distributions, which were evaluated against independent presence-absence data. The authors found that Swiss presence-only tree data produced models that were reliable enough to be useful in planning and management applications.

Fithian et al. (2015) also considered an approach combining both presence-only and presence-absence data using the inhomogeneous Poisson process (IPP) (cf. Appendix A.1). Given a random set of S sites (point locations) in region (domain) \mathcal{D} , where \mathbf{s}_i is the two-dimensional location of site i , then for any subregion A in \mathcal{D} , let $N_S(A)$ denote the number of points \mathbf{s}_i falling into A . If S is an $\text{IPP}(\lambda)$ with

intensity $\lambda(\mathbf{s})$, then $N_S(A)$ is a Poisson random variable with mean

$$\Lambda(A) = \int_A \lambda(\mathbf{s}) d\mathbf{s}.$$

If A is a quadrat centered at \mathbf{s} , small enough that $\lambda(\mathbf{s})$ is nearly constant over A , then $\Lambda(\mathbf{s}) \approx \lambda(\mathbf{s})|A|$, where $|A|$ represents the area of A . This means that $\lambda(\mathbf{s})$ represents the expected species count per unit area near \mathbf{s} . Then integrating over the entire study region R ,

$$\Lambda(\mathcal{D}) = \int_R \lambda(\mathbf{s}) d\mathbf{s} = E[N_S(\mathcal{D})],$$

the expected population size. Normalizing $\lambda(\mathbf{s})$, we have

$$f_\lambda(\mathbf{s}) = \lambda(\mathbf{s})/\Lambda(\mathcal{D})$$

is the probability function of the distribution of individuals. The authors used the model

$$\log \lambda(\mathbf{s}) = \alpha + \boldsymbol{\beta}' \mathbf{x}(\mathbf{s}),$$

with covariate information.

The presence-only process \mathcal{T} consists of the set of all individuals observed by opportunistic presence-only sampling. Assuming they are identified correctly (not always a given), \mathcal{T} is the subset of S that remains after the sites with unobserved individuals are removed (thinned). An individual at \mathbf{s}_i is included in \mathcal{T} (i.e., is observed) with probability $p(\mathbf{s}_i)$ independent of all other individuals. As a result of the biased thinning, individuals in areas with relatively large $p(\mathbf{s})$ will tend to be over-represented relative to areas with small $p(\mathbf{s})$. This leads to

$$\mathcal{T} \sim PP(\lambda(\mathbf{s}) p(\mathbf{s})),$$

representing a “thinning” Poisson process. The authors used the model with the two parts modeled separately, namely,

$$\log(\lambda(\mathbf{s}) p(\mathbf{s})) = \alpha + \boldsymbol{\beta}' \mathbf{x}(\mathbf{s}) + \gamma + \boldsymbol{\delta}' \mathbf{z}(\mathbf{s}).$$

If the covariates comprising \mathbf{x} and \mathbf{z} are distinct and have no perfect linear dependencies on one another, then b , d , and the sum $a + c$ are identifiable, but individually a and c are not, so that $\lambda(s)$ cannot be estimated.

In combining presence-only and presence-absence data, it is assumed that there are m species. Let S_k and \mathcal{T}_k denote the species and presence-only processes for species k ($k = 1, \dots, m$). Each observation i is associated with site \mathbf{s}_i as well as covariates \mathbf{x}_i and \mathbf{z}_i . For a survey site, \mathbf{s}_i represents the centroid of a quadrat A_i , and

there we observe counts $n_{ik} = N_{s_k}(A_i)$ or binary presence/absence indicators y_{ik} with $y_{ik} = 1$ if $n_{ik} > 0$ and $y_{ik} = 0$ otherwise. For species k , the authors modeled the inhomogeneous Poisson processes $S_k \sim \text{IPP}(\lambda_k(s))$ and $\mathcal{T}_k \sim \text{IPP}(\lambda_k(s)p_k(s))$ with S_k and \mathcal{T}_k independent across species. They then used

$$\log(\lambda_k(s)p(s)) = \alpha_k + \boldsymbol{\beta}'_k \mathbf{x}(s) + \gamma_k + \boldsymbol{\delta}' \mathbf{z}(s),$$

where $\boldsymbol{\delta}$ is assumed to be constant across species. This allows parameters to be estimated. Unless there is some presence/absence data for species k , α_k is unidentifiable.

To develop the model for maximum likelihood estimation, let I_{PA} denote the set of indices i for which s_i are presence-absence survey quadrats, and let I_{PO_k} denote the indices for presence-only sites $s_i \in S_k$. Let n_{PA} be the total number of survey quadrats. For species k , the log-likelihood for the presence-absence data is

$$\begin{aligned} \ell_{k,\text{PA}}(\alpha_k, \boldsymbol{\beta}_k) = & \sum_{i \in I_{\text{PA}}} -y_{ik} \log \left(1 - e^{-\exp[\alpha_k + \boldsymbol{\beta}'_k \mathbf{x}_i]} \right) \\ & + (1 - y_{ik}) \exp[\alpha_k + \boldsymbol{\beta}'_k \mathbf{x}_i]. \end{aligned}$$

The log-likelihood for the presence-only data is

$$\begin{aligned} \ell_{k,\text{PO}_k}(\alpha_k, \boldsymbol{\beta}_k, \gamma_k, \boldsymbol{\delta}) = & \sum_{i \in I_{\text{PO}_k}} \log(\lambda_k p_k(s_i)) - \int_{\mathcal{D}} \lambda_k p_k(s) ds \\ = & \sum_{i \in I_{\text{PO}_k}} (\alpha_k + \boldsymbol{\beta}'_k \mathbf{x}_i + \gamma_k + \boldsymbol{\delta}' \mathbf{z}_i) \\ & - \int_{\mathcal{D}} \exp[\alpha_k + \boldsymbol{\beta}'_k \mathbf{x}_i + \gamma_k + \boldsymbol{\delta}' \mathbf{z}_i], \end{aligned}$$

and the second equation can be evaluated numerically. If we put γ_k and $\boldsymbol{\delta}$ equal to zero, we get the model for just the presence-only data. We need to multiply the two types of model together for all k species for estimation. Various submodels were considered along with some problems with the approach, and the authors have released an **R** package called **multispeciesPP** for performing the computations.

4.13 Multi-Scale Models

Developments of species distributions depend critically on the scales at which data is used for model building, and this topic was referred to briefly in the previous discussion. Species interact differently with the environment at different scales. For example, Gould et al. (2012) observed that the appropriate spatial scale for

monitoring amphibian populations is uncertain and likely differs among species. In their study, they used two scales: catchments (portions of watersheds containing variable amounts and types of wetlands) and sites (individual ponds or wetlands within the catchments).

Steenweg et al. (2018), in a largely descriptive paper, noted that little research has explored how different occupancy sampling designs affect occupancy-abundance relationships. The independence of occupancy estimates from spatial sampling units also depends on the sampling unit. Occupancy surveys are most directly related to abundance when cells size matches home range size, allowing one individual to occupy each cell. Smith and Goldberg (2020) commented that “A challenge for occupancy monitoring programs is to determine the appropriate spatial scale of analysis and to obtain precise occupancy estimates for elusive species” and “the choice of sampling unit remains difficult and is further complicated in landscapes with dynamic habitat.” Sampling at a landscape level can be costly, especially for rare and difficult-to-detect species. However, efficient landscape monitoring that investigates occupancy at multiple spatial scales is important for many species such as aquatic species in arid systems and amphibians. They noted that highly sensitive eDNA detection techniques can be effectively used to model occupancy at site level and landscape scales (c.f. Sect. 4.13).

Nichols et al. (2008) developed multiple detection models to estimate parameters associated with both large- and small-scale occupancy, which were considered in Sect. 4.10.2. We now look further at multi-scale models as there is widespread support for the hypothesis that ecological processes are affected by environmental variables at multiple spatial scales. Spatial variation in abundance is influenced by local- and landscape-level environmental variables, but modeling landscape effects is challenging because the spatial scales of the relationships are unknown.

Pavlacky et al. (2012) considered a similar model except that they estimated small-scale occupancy using spatial rather than temporal replicates. They permitted the simultaneous use of presence-absence data collected at two spatial scales together with a removal design to estimate the probability of detection. Occupancy at the small scale corresponded to local territory occupancy, whereas occupancy at the large scale corresponded to regional occupancy of the sample units for two bird species. The authors also accounted for the non-independence of the spatially replicated survey stations and addressed the closure assumption for the spatially replicated survey stations as follows.

Suppose we have a sample of S sites (sampled units) with each one subsampled by R (primary) survey stations to determine the presence or absence of a species. At each survey station, T (secondary) repeated survey occasions are carried out, so we essentially have a robust design (Pollock, 1982) with primary and secondary units. When the survey occasions are sampled without replacement, the robust design can be combined with a removal design in which only the first detection at a survey station is recorded. For example, with $R = 4$ and $T = 3$, one example encounter history is $H_i = 010, 101, 000, 001$, where 0 is non-detection and 1 is detection. Here, the target species was detected at survey station 1 on occasion 2 and at station 2 on occasions 1 and 3, was not detected at station 3, and was detected at survey

station 4 on occasion 3. For the removal model, we have missing data so that $H_i = 01\cdot, 1\cdot\cdot, 000, 001$. The removal design is well suited for situations where the survey occasions are not independent and where species are more or less likely to be detected in subsequent survey occasions.

Let p be the probability of detection at an occasion given the survey station is occupied; let θ be the probability of occupancy for a survey station, given the site is occupied; and let ψ be the probability of occupancy for a site. The assumptions of the multi-scale occupancy model are as follows: (1) no unmodeled heterogeneity in the probabilities of detection and occupancy, (2) each survey station is closed to changes in occupancy over the sampling period, (3) the detection of species at each survey station is independent, and (4) the target species are never falsely detected.

The occupancy parameters ψ and θ allow the estimation of occupancy at two spatial scales. The parameter ψ represents species occurrence at the larger scale and can be interpreted as the proportion of sites occupied. The parameter for the smaller scale, θ , corresponds to species occurrence at the survey stations conditional on species presence at the site and can be interpreted as the proportion of survey stations occupied when the site is occupied. The product $\psi\theta$ corresponds to the probability of small-scale occupancy, while $\psi(1 - \theta)$ corresponds to large-scale, but not small-scale, occupancy. Here, $(1 - \theta)$ represents the probability of no availability given presence, $\theta(1 - p)$ corresponds to the probability of no detection given presence and availability, and θp reflects the probability of detection given presence and availability. This parameterization relaxes the closure assumption for the spatially replicated survey stations and allows some survey stations to have negligible probabilities of detection. We have given just a brief summary of the method, and Pavlacky et al. (2012) should be consulted for practical details and computations.

Kroll et al. (2015) developed a hierarchical single-scale and multi-scale models for plethodontid salamanders. The single-scale model was a fairly standard occupancy model at the plot level that incorporated random effects to account for both the nested spatial sampling design and repeated measurements from the same unit across time. The multi-scale model involved two nested but distinct occupancy processes, one acting at the harvest unit level and the other acting at the plot level, conditional on harvest unit occupancy. Logistic covariate models were used.

Chandler and Hepinstall-Cymerman (2016) noted that the common approach for addressing the scale problem is to measure each landscape variable at multiple spatial scales, defined by polygons of various extents centered on a set of survey sites. For instance, in avian point-count surveys, each point-count location may be buffered by concentric circles with radii ranging from hundreds of meters to several kilometers and landscape variables such as forest cover are averaged over each circle. However, there are some problems with this approach as arbitrary decisions need to be made.

For simplicity, the authors assumed that the landscape can be characterized as a regular grid such that each pixel is a site with an associated value of abundance and an associated set of environmental covariates, though their method is readily adapted to other landscape configurations. Their method consists of using smoothing kernels

to average landscape variables around focal sites and applied maximum likelihood to estimate the scale parameters of the kernels and the effects of the smoothed variables on abundance as follows.

The coordinates of a site (i.e., pixel), of which there are R , are denoted by \mathbf{s} , and the abundance at a site is denoted by $N(\mathbf{s})$. The aim was to model the expected value of abundance $E[N(\mathbf{s})] = \lambda(\mathbf{s})$ as a function of the environmental covariates $\mathbf{z}(\mathbf{s})$. To address the fact that one does not know the scale at which the surrounding landscape affects $N(\mathbf{s})$ and that the effect should diminish with distance, they employed a spatial smoothing approach. Here, they used a function of the form

$$f(z(\mathbf{s}_i, \sigma)) = \sum_{\mathbf{s}_j (\neq \mathbf{s}_i) \in R} z(\mathbf{s}_j) w(\mathbf{s}_i, \mathbf{s}_j, \sigma),$$

a weighted average of the original landscape variable with weights, $w(\cdot)$, determined by an appropriate kernel. For the Gaussian (normal) case,

$$w(\mathbf{s}_i, \mathbf{s}_j, \sigma) = \frac{\exp[-||\mathbf{s}_i - \mathbf{s}_j||^2/(2\sigma^2)]}{\sum_{\mathbf{s}_j (\neq \mathbf{s}_i) \in R} \exp[-||\mathbf{s}_i - \mathbf{s}_j||^2/(2\sigma^2)]},$$

where σ is the scale parameter that determines the shape of the kernel and $||\mathbf{s}_i - \mathbf{s}_j||$ is the Euclidean distance between sites i and j . If σ is high, then the kernel is relatively flat, and landscape features far from the focal site are more influential than if σ is low. The smoothing takes place over the two-dimensional region R that encompasses the study area. They then used the model

$$\log[\lambda(\mathbf{s}_i)] = \beta_0 + \beta_1 f_1(z_1(\mathbf{s}_i, \sigma_1)) + \dots + \beta_p f_p(z_p(\mathbf{s}_i, \sigma_p)).$$

This equation includes a smoothing term for each of the p landscape covariates, but site-level covariates can be accommodated by removing the associated smoothing functions, i.e., the $f(\cdot)$ functions. For example, using one site-level covariate ($z_1(\mathbf{s})$) and one landscape-level covariate ($z_2(\mathbf{s})$), abundance data were generated from the model

$$\log[\lambda(\mathbf{s}_i)] = \beta_0 + \beta_1 z_1(\mathbf{s}_i) + \beta_2 f_2(z_2(\mathbf{s}_i), \sigma).$$

Various abundance models can be used, but for simplicity, the authors used a Poisson model so that $N(\mathbf{s}_i) \sim \text{Poisson}(\lambda(\mathbf{s}_i))$ with likelihood

$$L(\boldsymbol{\beta}, \sigma, \{N(\mathbf{s}_i)\}) = \prod_{i=1}^S \frac{\lambda(\mathbf{s}_i)^{N(\mathbf{s}_i)} \exp[-\lambda(\mathbf{s}_i)]}{N(\mathbf{s}_i)!},$$

where S , a subset of R , is the number of focal sites in the sample taken. The above likelihood assumes that the site-specific abundances, $\{N(\mathbf{s}_i)\}$, are independent after accounting for spatial dependence attributable to the smoothed landscape variables.

Maximizing the likelihood can be accomplished using numerical optimization routines available in most statistical software packages, and it can be easily modified to accommodate other discrete distributions such as the negative binomial or zero-inflated Poisson. Selecting among kernels (e.g., Gaussian vs exponential) can be accomplished using AIC or some other information criterion. **R** code for fitting the model is available in an online appendix.

Stevens and Conway (2019) raised the question of developing optimally predictive multi-scale SDMs (species distribution models) when covariate data are collected over a range of scales and the problem of how to select scales. They constructed multi-scale models by first optimizing the observational scale of each covariate using model selection and then combining scale-optimized covariates into true multi-scale models (hereafter multi-scale approach). The multi-scale approach was compared to a constrained approach that first identified the best model at each individual scale using model selection and then identified the best overall single-scale model (hereafter scale-constrained approach). Their paper is basically descriptive without mathematics and has an extensive bibliography.

4.14 Multispecies Models

Throughout this chapter, reference is made briefly to various multispecies models. As the extension is straightforward, we give just a basic example from Sauer et al. (2013), which is similar to a previous model (Zipkin et al., 2010). For species i and site j , let z_{ij} be the usual presence indicator, where $z_{ij} = 1$ with probability ψ_{ij} . For visit k , let y_{ijk} be the observed indicator variable, where $y_{ijk} = 1 \mid z_{ij} = 1$ has probability p_{ijk} . Then z_{ij} and y_{ijk} are the usual Bernoulli random variables with

$$y_{ijk} \sim \text{Bernoulli}(p_{ijk}z_{ij}).$$

Linear models in habitat covariates including a species random effect are used for $\text{logit}(\psi_{ij})$ and $\text{logit}(p_{ijk})$, along with prior distributions on the covariate parameters. A factor involving management strategy was also considered. The authors fitted the model using the software WinBUGS, which uses Markov chain Monte Carlo (MCMC) methods to estimate the posterior distributions of the parameters of interest. This occupancy topic is also mentioned in Sect. 5.3.3.

4.15 Zero-Inflated Models

It has been mentioned in various places in this chapter that we can sometimes have extra zeros in our models, often because of overdispersion or because there are two types of absences, both actual and unobserved. A mixture model of two distributions can then be used for random variable Y , say, such as a $\text{Poisson}(\lambda)$ distribution and

one with a point mass only at zero. For example, if ψ is the presence probability discussed above, then one mixture is

$$\Pr(Y = y) = \psi \exp[-\lambda] \lambda^y / y! + (1 - \psi) I(y = 0),$$

where $I(\cdot)$ is an indicator function and

$$\Pr(Y = 0) = \psi \exp[-\lambda] + 1 - \psi.$$

In general, given $f(y | \boldsymbol{\theta})$, a distribution on the integers $y = 0, 1, 2, \dots$, we define an associated zero-inflated distribution

$$\Pr(Y = y | \boldsymbol{\theta}) = \psi f(y | \boldsymbol{\theta}) + (1 - \psi) I(y = 0).$$

Wenger and Freeman (2008) used the model to estimate abundance. Candidates for $f(y | \boldsymbol{\theta})$ include a variety of well-known discrete distributions including the negative binomial and the previously mentioned binomial and Poisson distributions.

An alternative (“hurdle”) model, introduced by Cragg (1971), is when $f(y | \boldsymbol{\theta})$ is a distribution on the integers $y = 1, 2, \dots$, with 0 omitted. Potts and Elith (2006) compared five regression models to assess the relationship between the abundance of a vulnerable plant species, *Leionema ralstonii*, and the environment and found the hurdle model performed the best. A double hurdle model is given by Balderama et al. (2016).

The zero-inflated model arises in several subject areas such as time series, spatial statistics, zero-inflated Poisson (ZIP) regression, and combinations of these. ZIP models were discussed by Agarwal et al. (2002), Rathbun and Fei (2006), Lee et al. (2006), and Neelon et al. (2013). As mentioned above, there is the so-called “hurdle” model that is similar to a ZIP model but does not mix zeros from the binary and count processes. Ver Hoef and Jansen (2007) described such a model that involves a mixture of a binary distribution and a count distribution that includes zero. In their application involving harbor seal counts on ice harbor seal, the binary distribution is the absence or presence of harbor seals, and the count distribution is the number of seals. They compared the space-time ZIP and hurdle models and how they differed in terms of estimated parameters, predictions, and precision. They used the standard formulation of ZIP and hurdle models to develop Bayesian hierarchical models with space-time errors and autocorrelation to investigate haul-out patterns of harbor seals on glacial ice as follows.

Let $y_{it} = 1$ if a seal is observed at location i at time t , given that it is present, with probability p_{it} , and 0 otherwise, so that $y_{it} \sim \text{Bernoulli}(p_{it})$. A ZIP regression model is given by $z_{it} | y_{it} = 0$ if $y_{it} = 0$ and is Poisson with mean λ_{it} if $y_{it} = 1$. We can then use covariates to model $\log(\lambda_{it})$ and $\text{logit}(p_{it})$ with the two regression models assumed to be independent spatial processes with autocorrelation and space-time multivariate normal errors. Diffuse prior distributions were used for all the regression parameters, and posterior distributions were obtained. The models were

fitted using Markov chain Monte Carlo (MCMC) using **WinBUGS** software. The authors gave a directed graph comparing the ZIP and hurdle models.

Hurdle models are non-mixture models that completely specify and separate the binary distribution from the count distribution. Applying this idea to the previous model, we now assume z_{it} is Poisson ($\lambda_{it} + 1$) when $y_{it} = 1$, which the authors called the P1B hurdle model. One difference between the hurdle and ZIP models is that zeros play a different role; in the hurdle model, they come solely from the detection (Bernoulli) process.

Dorazio's Model

Dorazio (2014) described a hierarchical model with one component specifying the spatial distribution (true but unknown) within a bounded geographic region A (also its area) and a second component specifying models for two kinds of observations: (1) detections of individuals encountered during opportunistic surveys (i.e., presence-only observations) and (2) detections of individuals encountered during planned survey locations selected using a prescribed sampling design. The popular method of opportunistic or emergent sampling is discussed in Sect. 4.16.

Dorazio assumed that the spatial distribution of activity centers followed a Poisson process, as in the previous section, with parameter $\lambda(\mathbf{c})$, where $\mathbf{c} = (c_1, c_2)'$ denotes a location (point) in A . The number N of individuals in A , N_A , is Poisson(μ_A), where $\mu_A = \int_A \lambda(\mathbf{c})d\mathbf{c}$ is the mean intensity over A . Furthermore, if A is partitioned into a set of disjoint subregions, say $A_1 \cup \dots \cup A_K = A$, then $N_{A_k} \sim \text{Poisson}(\mu_{A_k})$, where $\mu_{A_k} = \int_{A_k} \lambda(\mathbf{c})d\mathbf{c}$. If n is the total (unknown) number of individuals in A , then

$$\Pr(N_A = n) = \exp(-\mu_A)\mu_A^n/n!. \quad (4.5)$$

Also, conditional on n , the joint density function of the n activity centers is

$$f(\mathbf{c}_1, \dots, \mathbf{c}_n \mid n) = \prod_{i=1}^n \lambda(\mathbf{c}_i)/\mu_A. \quad (4.6)$$

Combining (4.5) and (4.6), we have

$$g(\mathbf{c}_1, \dots, \mathbf{c}_n, n) = \frac{\exp[-\mu_A]}{n!} \prod_{i=1}^n \lambda(\mathbf{c}_i).$$

We can then model $\log(\lambda(\mathbf{c})) = \boldsymbol{\beta}'\mathbf{x}(\mathbf{c})$, where the regressors $\mathbf{x}(\mathbf{c})$ are location-specific covariates. Diffuse prior distributions were used for all the regression parameters, and posterior distributions were obtained. The models were fitted using Markov chain Monte Carlo (MCMC) in **WinBUGS** software. Dorazio (2014) discussed some problems relating to the identifiability of the parameters. Spatial models are considered in detail in Chap. 14.

In the case of planned surveys, a number of scenarios are possible such as double-observer surveys, removal surveys, and replicated point-count surveys, discussed in later chapters, and such surveys can be combined with opportunistic surveys, as discussed below. Dorazio concluded that “bias in estimates of SDMs induced by detection errors or survey bias can be reduced or eliminated by modeling presence-only data in conjunction with counts observed in planned surveys.”

4.16 Opportunistic and Citizen Surveys

Up till now, we have only alluded to opportunistic surveys, and we now look at these more closely. Informal monitoring efforts are often the only data available upon which to base population assessments and have been shown to be a useful supplement to traditional analyses (e.g., Cohn, 2008; Dennis et al., 2017, butterflies). They can cover long periods of time and large areas, thus providing information that is usually lacking in monitoring programs. Welvaert and Caley (2016) suggested distinguishing between crowdsourcing, unstructured citizen science, and structured citizen science. A somewhat different use of opportunistic data is when it was from common dolphin sighting data from two ferry routes across three regions collected by citizen scientists on ferries between April and October in 2006–2017 (Robbins et al., 2020). Hunter sightings can also provide citizen science data (e.g., Rich et al., 2013: wolves). General comments on citizen science are made below in Sect. 4.16.6.

4.16.1 Two-Stage Model

Dennis et al. (2013) proposed a two-stage model for counts of invertebrate species that made more efficient use of the data than previous analyses while accounting for missing values. In addition to creating more robust trend estimates, the new method allowed all volunteer records to contribute to the indices and thus incorporates data from more populations within the geographical range of a species. Monitoring invertebrates presents a number of technical challenges, such as sampling frequency to cover seasonal patterns and the specialized expertise required for identification. They listed monitoring schemes and research applications for butterflies and other seasonal insect taxa as follows.

Butterfly population data in the UK are principally gathered through an intensive, wide-scale monitoring system of weekly transect walks that form the UK Butterfly Monitoring Scheme (UKBMS). Since its inception in 1976, a large network of recorders has contributed to the UKBMS, making around a quarter of a million weekly visits to almost 2000 different sites and counting over 16 million butterflies. Transects were used, and some weeks of the transect season were missed due to unsuitable weather conditions or recorder unavailability, for example, due to illness or holidays, and hence, fewer than 26 counts per year are typically made at each site.

Under certain criteria, it was found that on average across the species monitored by the UKBMS, 38% of transect visits made did not contribute to population indices, which represents a substantial quantity of data not utilized. Current models for the estimation of missing counts were extended to allow for all incomplete series of recordings and annual variation in seasonal pattern, to make more efficient use of the data collected. A sketch of the method follows.

Let y_{sjt} represent the count of a species at site $s = 1, \dots, S$ in year $j = 1, \dots, J$ on day t . A generalized additive model (GAM) for a Poisson distribution has its expectation modeled by

$$E[y_{sjt}] = \mu_{sjt} = \exp[\alpha_s + \beta_j + \gamma_j(t)],$$

where α_s and β_j represent effects for the s th site and the j th year, respectively, and $\gamma_j(t)$ allows for the seasonal pattern, which can vary between years, but not over sites. Missing values can also be estimated using the above equation. A site index, M_{sj} for year j , can be calculated as the sum of the expected counts for that season, namely, $M_{sj} = \sum_t \mu_{sj}(t)$. The paper is referred to for further details. The method was extended by Dennis et al. (2016a,b) using a stopover model of Matechou et al. (2014), and a generalized abundance index was introduced.

Multiple Data Sets

With respect to multiple data sets, previously mentioned, Lele and Allen (2006) gave a model for incorporating expert opinion to improve the usefulness of the data, while Zulian et al. (2021) concluded that combining different data sets into one joint-likelihood model produced better estimates than any single data set used on its own. García et al. (2021) referred to a special citizen working protocol called “Adopt a plant” which is a collaborative network that is currently monitoring 332 populations of 204 plant taxa (threatened, of community interest, common, rare and habitat indicators) across a heterogeneous landscape in NE Spain. Their aim was to produce standardized indices of abundance change and other early-warning signals of concern or risk of population collapse. Callaghan et al. (2022) considered the effort needed to estimate species diversity from citizen science data. Species diversity is discussed in more detail in Sects. 5.5.11 and 5.5.12.

4.16.2 Pro and Cons

There is a great deal of literature looking at the pros and cons of citizen science. Opportunistic citizen science data may suffer from changes in field efforts over time (observation bias), from incomplete and selective recording by observers (reporting bias), and from geographical bias. These, in addition to detection bias, may lead to spurious trends. For example, Gardiner et al. (2012) demonstrated an overestimation of species richness and diversity values in lady beetles based on citizen science data compared to verified data. This was attributed to under-reporting of common species

combined with over-reporting of rare native species (including false positives). Also, anecdotal occurrence evidence, especially for rare or elusive species, can be very misleading (cf. McKelvey et al., 2008 for three examples).

Weller et al. (2014) evaluated whether an opportunistic collection of surveys recorded over a 22-year period could be used to assess the population trend of Townsend's big-eared bats (*Corynorhinus townsendii*) at Lava Beds National Monument in Northern California. They used records of counts of hibernating bats conducted during 1991–2012 to estimate the number of bats in 52 individual caves as well as cumulatively. Seventeen of 22 caves surveyed in 4 or more years had an increasing trend in the number of hibernating bats. They estimated the cumulative annual growth rate over the period to be 1.79%. A log-linear model was fitted to data from each cave that was surveyed four or more times during the period and for which 50% or more of the counts were non-zero.

Mair and Ruete (2016) noted that volunteer recorders do not select survey sites randomly; they may be influenced by accessibility, proximity to their home, and the known species richness of a site, which makes it more or less attractive to survey. We can add trends in recording intensity and differential recording rates among species. Also, there is a tendency to focus on particular taxonomic groups, so that they may not record every species they see, through either a lack of identification skills or a lack of interest in very common species. This can lead to various biases such as spatial biases, with some areas receiving little or no survey effort. This spatial biasedness was also found by Tiago et al. (2017)) with regard to the use of a Portuguese citizen science database.

Some methods are suggested for checking up on the validity of citizen science data. For example, Snäll et al. (2013) listed several references to some methods and mentioned the importance of having evaluation data, which is not always available. However, they described using data on environmental variables that affect the dynamics of the species of interest to set up an evaluation model. Mair and Ruete (2016) used taxonomic “ignorance” scores in Sweden, which are spatially explicit indices of the bias and lack of sampling effort across a study region. Raw citizen science observation data of a reference taxonomic group is used to allow the identification of the areas where there is least (or conversely most) confidence that absences of observations in the data set are true species absences. An ignorance score map could therefore inform users of undersampled areas to be targeted for surveys or be applied in species distribution modeling as a confidence or bias layer. Bayesian methods were used to model the ignorance scores.

4.16.3 Accounting for False Positives

Pillay et al. (2014) considered methods for improving estimates of occupancy by accounting for false positives, which can occur with public surveys. They fitted occupancy models that simultaneously account for false positives (misidentification) and negatives (imperfect detection) to data collected from a large-scale key infor-

mant survey for 30 species of large vertebrates and tested their performance against standard occupancy models that account only for false negatives. They concluded that generally in most cases, false positives need to be taken into account in using public survey data sets such as key informant interviews, volunteer surveys, citizen science programs, historical archives, and acoustic surveys. On the other hand, Higa et al. (2015) found that “citizen data with spatially biased sampling effort can be appropriately utilized for large-scale biodiversity distribution modeling with the use of occupancy models, which encourages data collection by citizen scientists.” The use of occupancy models for citizen science was strongly recommended by Van Strien et al. (2013) as a source of information to track distribution trends in many groups of species.

4.16.4 Disasters

Citizen science can be particularly useful in dealing with extensive disasters covering large areas such as extreme weather, and extensive, severe wildfires, and wildfire-induced smoke. For example, Yang et al. (2021) considered migratory bird deaths as the result of forest fires and climate change in the USA using regression models with various covariates. Russell et al. (2015) considered the impact of a hurricane on Swedish birds.

4.16.5 Bird Applications

With regard to birds, there is a huge network of data from citizen science organized by a package called eBird (Sullivan et al., 2009). Developing and shaping this network over time, eBird has created a near-real-time avian data resource producing millions of observations per year. One extensive application is the Swedish Species Gateway, established in 2000, which includes millions of bird observations by volunteers collected by presence-only observations. The Gateway also has platforms for other groups of organisms such as insects and vascular plants (Snäll et al., 2010). Wei et al. (2016) carried out an extensive review to help designers and managers in linking citizen science with urban ecology for birds and butterflies. These species were used because of known ecological knowledge, and they have an appeal to the layperson. Steen et al. (2019) compared data from eBird against models built from four independent, systematically collected data sets. They developed species distribution models for 96 species using both data sources and compared their predictive abilities. Also tested was whether culling eBird data by applying stringent data filters on survey effort or observer expertise improved predictions. How best to cull was open for further research.

It should be noted that anecdotal occurrence evidence, especially for rare or elusive species, can be very misleading (McKelvey et al., 2008 for three examples).

However, any form of information on rare species is worth considering (e.g., even birding trip reports, Camacho, 2016). Kamp et al. (2016) compared unstructured citizen science with structured surveys and found that the former failed to detect population declines of common birds in Denmark. They concluded that “citizen science data collected using unstructured methods may be useful for biodiversity monitoring for species or in areas where dedicated, structured survey data are not available, but that various sources of bias need to be considered in the interpretation of population trend estimates.”

4.16.6 General Comments

For general comments on citizen science, see Dickinson et al. (2012, description only), Hochachka et al. (2012, some models mentioned), Jiguet et al. (2012, some mention of designs), and Callaghan et al. (2020, urban environments). Chase and Levine (2016) reviewed 52 citizen science programs to provide a framework of resource characteristics critical to the design and assessment of citizen science programs that monitor natural resources. Lewandowski and Specht (2015) conducted a systematic search of the peer-reviewed literature for studies that explicitly assessed the quality of organismal data collected by volunteers and found that majority of the papers in their review showed that volunteers collect, or have the potential to collect, useful, high-quality data. However, a few papers showed that volunteers might not always be capable of meeting the needs of a survey scheme. There are a large number of research articles on citizen science, and this method has been used extensively for some time, as well as making use of other collection methods, e.g., line transects, mentioned in later chapters.

As noted by Van Strien et al. (2013), opportunistic citizen data has been described as observations of species collected without a standardized field protocol and without an explicit sampling design. It can be a challenge to achieve reliable estimates of distribution trends from them, because they may suffer from changes in field efforts over time (observation bias), from incomplete and selective recording by observers (reporting bias), and from geographical bias. These, in addition to detection bias, may lead to spurious trends.

The model they used was a standard one incorporating two hierarchically coupled submodels for occupancy and detection (conditional on occupancy), the latter being conditional on the occupancy submodel. The occupancy submodel contained two parameters, one pertaining to the probability of persistence (ϕ) and one to the probability of colonization (γ). A logistic regression model with covariates was used for p_{ijt} , the probability of detecting a given species at site i during visit j in year t . A Bayesian approach was used. Two main findings were as follows: (1) the distribution trends provided by opportunistic and monitoring data were well matched, and (2) opportunistic butterfly and dragonfly data can be successfully used for inferences on distribution trends, provided that they are analyzed with occupancy models to treat observation bias and other biases affecting detection.

4.16.7 Citizen Presence-Only Data

With citizen science, presence-only data is usually all there is. Although mentioned in Sect. 4.12, we now focus on this topic with regard to research on citizen science. Following Dorazio (2014), we begin by assuming that N_A individuals in area A are Poisson(μ_A), where $\mu_A = \int_A \lambda(\mathbf{c}) d\mathbf{c}$, with $\log \lambda(\mathbf{c}) = \boldsymbol{\beta}' \mathbf{x}(\mathbf{c})$. Also, each individual is assumed to be detected independently by an observer with probability $p(\mathbf{c})$ that depends only on the location \mathbf{c} of an individual. Therefore, $p(\mathbf{c})$ includes both the observer's detection ability and choice of survey location. For example, an observer may not survey individuals at location \mathbf{c} because the location is inaccessible. We can then model $\text{logit}(p(\mathbf{c})) = \boldsymbol{\alpha}' \mathbf{w}(\mathbf{c})$, which is assumed to have an intercept α_0 and where the covariates can be linked to an observer's ability or choice of survey location. It is assumed that individuals are detected at their activity center, which could be problematic (e.g., highly mobile animals). If M_A is the number of individuals present and detected, then

$$\Pr(M_A = m) = \exp[-v_A] (v_A)^m / m!, \quad (4.7)$$

where $v(A) = \int_A \lambda(\mathbf{c}) p(\mathbf{c}) d\mathbf{c}$, which will need to be computed numerically, and M_A is a Poisson process resulting from an independent thinning of the original process N_A . We have the joint distribution

$$h(\mathbf{c}_1, \dots, \mathbf{c}_m, m) = \frac{\exp[-v_A]}{m!} \prod_{i=1}^m \lambda(\mathbf{c}_i) p(\mathbf{c}_i), \quad (4.8)$$

where the first m of n locations are assumed to correspond to those of detected individuals. Here, (4.8) can be used as a likelihood $L(\boldsymbol{\beta}, \boldsymbol{\alpha})$ for estimating $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$.

Since there are no non-detection records, Ruete et al. (2017) constructed artificial data on non-detections as follows. They first considered each unique observer reporting at least one species at a site on a specific day to constitute a replicate visit within that day. Then, for each visit j , in day d , year t , and site i , any observation of the focal species was considered as a detection if the species was reported during the visit ($y_{jdti} = 1$) and as a non-detection if it was not reported ($y_{jdti} = 0$). A non-detection then corresponded to the focal species not being reported by an observer reporting at least one other species at the wetland in that day. This procedure was repeated for all study species. Observations were recorded as “missing value” for days and sites without visits (i.e., when no observations were reported from the site in that day). They used species list length (SLL) as a proxy for effort. For computational reasons, they restricted the maximum number of visits to 40 per day and site, prioritizing visits with the longest species lists. SLLs ranged from 1 to 45 species (wetland birds).

We have the modeling process

$$y_{jdti} \sim \text{Bernoulli}(u_{dti} p_{jdti}),$$

where $u_{d,i} = 1$ if present and 0 otherwise is the presence status and $p_{j,d,i}$ is the detection probability of the species, given present, with subscripts previously defined. We also add

$$u_{d,i} \sim \text{Bernoulli}(\psi_{d,i}), \\ \psi_{d,i} = u_{d-1,i}\varphi_{d-1,i} + (1 - u_{d-1,i})\gamma_{d-1,i}.$$

This means that whether site i occupied in day $d - 1$ is still occupied in day d is determined by the persistence probability (φ), whereas whether site i that is unoccupied in day $d - 1$ is occupied in day d depends on the colonization probability (γ). Probit regressions were given for $\varphi_{d-1,i}$ and $\gamma_{d-1,i}$.

Reich et al. (2018) investigated the use of opportunistic data as a guide to designing more efficient ecological surveys through a novel objective function that minimizes misclassification rate (i.e., false positives and false negatives) of the estimated occupancy maps. Here, the opportunistic data provides an initial occupancy estimate as a prior for a Bayesian model and an efficient posterior approximation that accounts for spatial dependence, covariate effects, and imperfect detection in an exchange algorithm to search for the optimal set of sampling locations to minimize misclassification rate. It is assumed that there are no false-positive observations. Their model is as follows.

We begin with the usual model of S possible sites, with \mathbf{s}_i the location of the i th site for which there are n_i sampling occasions. Suppose M of these sites are selected and sampled on T occasions so that $S - M$ sites will automatically have $n_i = 0$. The number of these where the species is observed at site i is y_i ($i = 1, \dots, M$), and, including the n_i zeros (and therefore the corresponding y_i zeros), we set $\mathbf{y} = (y_1, \dots, y_S)$. The y_i are assumed to be independently and identically distributed (iid), namely,

$$y_i | z_i \sim \text{Binomial}(n_i, p z_i), \quad (4.9)$$

where p is the probability of detection given present and $z_i = 1$ is 1 with probability ψ that the species occupies site i and 0 otherwise. The aim is to estimate $\mathbf{z} = (z_1, \dots, z_S)'$, where the z_i are iid satisfying

$$z_i | \theta_i \sim \text{Bernoulli}(g(\theta_i), | \mathbf{y},) \quad i = 1, \dots, S. \quad (4.10)$$

Here, θ_i ($i = 1, \dots, S$) is a latent random effect, and g is an inverse link function such as the inverse of the logistic function. A Bayesian approach is now used where it is assumed that $\boldsymbol{\theta} = (\theta_1, \dots, \theta_S)'$ has a multivariate normal distribution

$$\boldsymbol{\theta} \sim N(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}), \quad (4.11)$$

where \mathbf{X} is a known $S \times r$ matrix of covariates that uses the initial estimate of occupancy based on the auxiliary data, $\boldsymbol{\beta}$ is unknown, and $\boldsymbol{\Sigma}$ is assumed a known

$S \times S$ matrix. Finally, they used the priors $\beta \sim N(\gamma, \Lambda)$ and $p \sim \text{Beta}(a, b)$, which can be combined with Eqs. (4.9), (4.10), and (4.11).

The authors then focused on optimizing the design $\mathcal{D} = \{n_1, \dots, n_S\}$, i.e., the number of sampling occasions at each of the S locations under consideration. The optimization depends on the spatial configuration of the S locations, the matrix of covariates \mathbf{X} , and the true value of the unknown parameters $\phi_0 = \{p_0, \beta_0\}$, say. For a given data set \mathbf{y} , denote the posterior occupancy probability for location i as $\pi_i = \Pr(z_i = 1 \mid \mathbf{y}, \mathcal{D})$. They quantified the accuracy of the species distribution map using the Brier score

$$\mathcal{C}(\mathbf{y}, \mathbf{z}, \mathcal{D}) = \frac{1}{S} \sum_{i=1}^S (z_i - \pi_i)^2 = \frac{1}{S} \sum_{i|z_i=0} \pi_i^2 + \frac{1}{S} \sum_{i|z_i=1} (1 - \pi_i)^2, \quad (4.12)$$

where z_i is the true occupancy status. They chose the Brier score because it balances false-positive probabilities (i.e., π_i for sites with $z_i = 0$) and false-negative probabilities (i.e., $1 - \pi_i$ for sites with $z_i = 1$). A smaller average Brier score corresponds to a better design, and so they sought the design that minimizes the expected Brier score

$$\nu(\mathcal{D}) = \mathbb{E}[\mathcal{C}(\mathbf{y}, \mathbf{z}, \mathcal{D}) \mid \phi_0],$$

where the expectation is with respect to (\mathbf{y}, \mathbf{z}) given ϕ_0 .

Because they wanted to explore a large number of designs which all require expensive computations, they used an approximation of the posterior distribution for quicker evaluation. Their search algorithm for the optimal design required efficient posterior evaluation, and they proposed a three-step approximation:

- (1) Approximate the detection probability p in a way that is independent of θ .
- (2) Estimate the posterior of θ given p using a Gaussian approximation.
- (3) Approximate π_i by numerically integrating over θ .

The authors considered each of the above steps and then computed the optimal design by approximating it using Monte Carlo sampling over (\mathbf{z}, \mathbf{y}) , given the model parameters $\phi_0 = (p_0, \beta_0)$. Using simulation, they showed that the two-stage design leads to lower misclassification rates than several ad hoc designs.

Weller et al. (2014) evaluated whether an opportunistic collection of surveys recorded over a 22-year period could be used to assess the population trend of Townsend's big-eared bats (*Corynorhinus townsendii*) at Lava Beds National Monument in Northern California. They used records of counts of hibernating bats conducted during 1991–2012 to estimate the number of bats in 52 individual caves as well as cumulatively. Seventeen of 22 caves surveyed in 4 or more years had an increasing trend in the number of hibernating bats. They estimated the cumulative annual growth rate over the period to be 1.79%. A log-linear model was fitted to data from each cave that was surveyed four or more times during the period and for which 50% or more of the counts were non-zero.

In concluding this section, we note that, in the case of planned surveys, a number of scenarios are possible such as double-observer surveys, removal surveys, and replicated point-count surveys, discussed in later chapters, and such surveys can be combined with opportunistic surveys.

4.17 Multistate Methods

Occupancy models can be extended to allow for the possibility that individuals being detected are in more than one state, for example, sites used for resting vs foraging, sites with and without diseased or parasitized animals, and classifying occupied sites by whether young animals are produced at the sites. State-space methods have been used successfully for open populations (Seber and Schofield, 2019: chapter 13). Nichols et al. (2007), using data on the California spotted owl from S sites visited on T occasions, considered three true states, namely, unoccupied (state = 0), occupied with no production of young (state = 1), and occupied with successful reproduction (state = 2). Observations at site visits can be scored as 0 (non-detection; possible true states are 0, 1, and 2), 1 (detection with no evidence of production of young; possible true states are 1 and 2), or 2 (detection with evidence of successful reproduction; true state is 2). This yields a vector of observation data (0, 1, or 2) for each visited site. For example, detection history $h_i = 01021$ for site i indicates non-detection at sampling visits 1 and 3, detection with no evidence of young at visits 2 and 5, and detection with evidence of successful reproduction at visit 4. Let

$$\begin{aligned}
 \psi_i^{(1)} &= \Pr(\text{site } i \text{ is occupied regardless of reproductive state}) \\
 &= \Pr(\text{true state} = 1 \text{ or } 2) \\
 \psi_i^{(2)} &= \Pr(\text{young occurred, given that the site is occupied}) \\
 &= \Pr(\text{true state} = 2 \mid \text{true state} = 1 \text{ or } 2) \\
 p_{it}^{(1)} &= \Pr(\text{occupancy detected for site } i, \text{ visit } t \mid \text{true state} = 1) \\
 &= \Pr(\text{detection} \mid \text{true state} = 1) \\
 p_{it}^{(2)} &= \Pr(\text{occupancy detected for site } i, \text{ visit } t \mid \text{true state} = 2) \\
 &= \Pr(\text{detection} \mid \text{true state} = 2) \\
 \delta_{it} &= \Pr(\text{evidence of successful reproduction, given occupancy at site } i, \\
 &\quad \text{visit } t \text{ with successful reproduction}) \\
 \psi_i^{(1*2)} &= \psi_i^{(1)} \psi_i^{(2)}, \text{ probability site } i \text{ is occupied by successful breeders}
 \end{aligned}$$

If the state assignment is not characterized by uncertainty, then $\delta_{it} = 1$ and proceed with the subsequent modeling. For detection history $h_i = 1021$, we have

$$\Pr(h_i = 1021) = \psi_i^{(1)} \psi_i^{(2)} p_{i1}^{(2)} (1 - \delta_{i1}) (1 - p_{i2}^{(2)}) p_{i3}^{(2)} \delta_{i3} p_{i4}^{(2)} (1 - \delta_{i4}).$$

The site was known to be occupied, and it was known that successful reproduction occurred (observed at visit 3). The probability associated with this event is $\psi_i^{(1)} \psi_i^{(2)}$. The species was detected at the site at period 1 (probability $p_{i1}^{(2)}$), but no evidence of successful reproduction was detected at that time (probability $(1 - \delta_{i1})$). The species was not detected at period 2 ($1 - p_{i2}^{(2)}$). At period 3, the species was detected ($p_{i3}^{(2)}$) and evidence of successful reproduction found (δ_{i3}). Finally, at period 4, the species was detected ($p_{i4}^{(2)}$), but no evidence of successful reproduction was found ($1 - \delta_{i4}$). This is the simplest form of detection history, because it includes a 2, so there is no uncertainty about true occupancy status or reproductive state. For the situation where there is no 2, consider

$$\begin{aligned} \Pr(h_i = 0101) &= \psi_i^{(1)} [(1 - \psi_i^{(2)}) (1 - p_{i1}^{(1)}) p_{i2}^{(1)} (1 - p_{i3}^{(1)}) p_{i4}^{(1)} \\ &\quad + \psi_i^{(2)} (1 - p_{i1}^{(2)}) p_{i2}^{(2)} (1 - \delta_{i2}) (1 - p_{i3}^{(2)}) p_{i4}^{(2)} (1 - \delta_{i4})]. \end{aligned}$$

The site was known to be occupied ($\psi_i^{(1)}$), but evidence of successful reproduction was never observed, so uncertainty exists about the true reproductive state. Because of this uncertainty, the detection history model must account for the fact that the true reproductive state could have been either 1 (first additive term within brackets) or 2 (second additive term). If the pair at the site was in true reproductive state 1 ($1 - \psi_i^{(2)}$), then occupancy was detected at periods 2 ($p_{i2}^{(1)}$) and 4 ($p_{i4}^{(1)}$), but not at periods 1 ($1 - p_{i1}^{(1)}$) or 3 ($1 - p_{i3}^{(1)}$). There are no δ_{it} parameters in this first term, because the true reproductive state in this part of the expression is 1 and, by assumption, is not prone to misclassification. However, if the true reproductive state was 2 ($\psi_i^{(2)}$), then detection for visits 2 and 4 is accompanied by misclassification [$p_{i2}^{(2)} (1 - \delta_{i2})$, $p_{i4}^{(2)} (1 - \delta_{i4})$]. Non-detection at periods 1 and 3 is modeled using the appropriate detection parameters [$(1 - p_{i1}^{(2)})$, $(1 - p_{i3}^{(2)})$].

Finally, the detection history with the greatest degree of uncertainty is

$$\begin{aligned} \Pr(h_i = 0000) &= (1 - \psi_i^{(1)}) + \psi_i^{(1)} (1 - \psi_i^{(2)}) \prod_{t=1}^4 (1 - p_{it}^{(1)}) \\ &\quad + \psi_i^{(1)} \psi_i^{(2)} \prod_{t=1}^4 (1 - p_{it}^{(2)}) \end{aligned}$$

The first of the additive terms corresponds to the possibility that the site was unoccupied ($1 - \psi_i^{(1)}$). The second of the additive terms corresponds to the possibility

that the site was occupied but that successful reproduction did not occur at the site ($\psi_i^{(1)}(1 - \psi_i^{(2)})$). The true reproductive state was thus 1 for the second term, and the species went undetected in all four visits. The third additive term corresponds to the final possibility that the site was occupied and that successful reproduction occurred ($\psi_i^{(1)}\psi_i^{(2)}$) and that the species went undetected during all four visits.

Missing observations (e.g., all sites are not visited the same number of times) can be dealt with by omitting any modeling of that portion of the detection history. For example, if the final fourth visit was not made at site i ,

$$\Pr(h_i = 102 \cdot) = \psi_i^{(1)}\psi_i^{(2)} p_{i1}^{(2)}(1 - \delta_{i1})(1 - p_{i2}^{(2)})p_{i3}^{(2)}\delta_{i3}.$$

We can now obtain a likelihood function of the ψ , p , and δ parameters proportional to the product of the site-specific probabilities $\prod_{i=1}^S \Pr(h_i)$. If there are too many parameters to obtain maximum likelihood estimates, covariates can be introduced as has been done in previous complex models. However, as the number of parameters goes up, so does the need for the amount of data. With appropriate modeling, different species can be modeled as different states.

Another example of a state-based method is given by Jennelle et al. (2007), with an emphasis on estimating prevalence.

4.17.1 Undetected State

MacKenzie et al. (2009) also considered modeling species occurrence with multiple states and imperfect detection. However, in addition to the possibility of incomplete detectability of a species, there is also the possibility that a particular state of an individual seen (e.g., breeding status) is not detected. For example, consider the following example: 0, unoccupied or species absence; 1, occupied and not breeding; and 2, occupied and breeding. Then, if unit i is surveyed twice per season over three seasons, the resulting detection history (h_i) for that unit could be denoted as, for example, $\mathbf{h}_i = 10\ 00\ 21$. Let $\psi^{(m)}$ be the probability that a species is in occupancy state m , ($m = 0, 1, 2$) which will sum to 1. We therefore have the state probability vector

$$\boldsymbol{\psi}' = (1 - \psi^{(1)} - \psi^{(2)}, \psi^{(1)}, \psi^{(2)}).$$

Next, let $p_j^{(\ell m)}$ be the probability of observing unit i in occupancy state ℓ during survey j given the true occupancy state is m ($\sum_{\ell=0}^m p_j^{(\ell m)} = 1$); then we have the following state detection probabilities forming a detection probability matrix \mathbf{P}_j . To

this we can add a probability state vector for a particular history, e.g., $\mathbf{h}_i = 010$ for a single season and three surveys giving

$$\mathbf{p}_{010} = \begin{bmatrix} 0 \\ (1 - p_1^{(11)})p_2^{(11)}(1 - p_3^{(11)}) \\ (1 - p_1^{(12)} - p_1^{(22)})p_2^{(12)}(1 - p_3^{(12)} - p_3^{(22)}) \end{bmatrix}.$$

The authors gave two other matrix examples. The probability of being observed is $\text{Pr}(\text{observed} | \text{present})\text{Pr}(\text{present})$, and for the single-season example $\mathbf{h}_i = 1201$, we have the unconditional probability

$$\text{Pr}(\mathbf{h}_i = 1201) = \psi^{(2)} p_1^{(12)} p_2^{(22)} (1 - p_3^{(12)} - p_3^{(22)}) p_4^{(12)}.$$

They compared their method with two other approaches and considered the extension to m states and multiple seasons. Assuming detection histories are independent, the likelihood function finally takes the form, for s samples, of

$$L(\boldsymbol{\theta} | \mathbf{h}_1, \dots, \mathbf{h}_s) = \prod_{i=1}^s \text{Pr}(\mathbf{h}_i | \boldsymbol{\theta}),$$

where $\boldsymbol{\theta}$ denotes all the parameters in the model. The method was applied to California spotted owls and green frogs in Maryland, using the software **WinBUGS**. Further details are available from MacKenzie et al. (2009) (Table 4.3).

Gould et al. (2019) used a similar multistate model to improve the understanding of amphibian breeding dynamics in the Greater Yellowstone Area, USA. Several hundred wetlands were visited with two observers surveying each wetland independently on a single visit using visual observation and dip-netting to detect amphibian breeding evidence (egg masses, larvae, etc.). Replicate surveys were separated by approximately 15 min to prevent communication between observers until after survey completion. The states were the wetland unoccupied and dry (state 0), the wetland is wet but does not support breeding by the target amphibian (state 1), and an occupied state (state 2) is conditional on water presence and represents breeding by the target amphibian (breeding occupancy). Using the conditional parameterization of the multistate model, the initial state probability vector is defined as $\phi_0 = (1 - \psi, \psi(1 - p_2), \psi p_2)',$ where ψ is the probability of a wetland

Table 4.3 State detection probabilities

| True state | Observed | State | |
|------------|-------------------------------|--------------|--------------|
| | | 0 | 1 |
| 0 | 1 | 0 | 0 |
| 1 | $1 - p_j^{(11)}$ | $p_j^{(11)}$ | 0 |
| 2 | $1 - p_j^{(12)} - p_j^{(22)}$ | $p_j^{(12)}$ | $p_j^{(22)}$ |

being wet in the first year of the study and p_2 is the probability of being in state 2 (occupied with breeding), given that a site is wet. Detection probabilities are conditional on the true state of the site. The authors gave just an outline of their model. Kendall (2009) gave some examples of state-space modeling, particularly when an animal is detected but there is uncertainty about which state it occupies.

4.17.2 Dynamic State Models

As an introduction, we first consider models where multiple seasons are used to see how populations change over time (cf. Miller et al., 2013; Miller & Grant, 2015). These belong to the topic of dynamic occupancy models, which we briefly introduce to cover some of the ideas considered in the following sections.

From Kéry et al. (2013), the site dynamics can be modeled by a two-state Markov chain, which is essentially a hidden Markov model (HMM). State transitions are governed by the probabilities of colonization γ and local extinction ε and depend only on the occupancy status at the previous time step. This means that

$$\psi_{t+1} = \psi_t(1 - \varepsilon_t) + (1 - \psi_t)\gamma_t.$$

The above probability is made up of the probability that the site was occupied at t and the species did not become locally extinct in the interval between t and $t + 1$, plus the probability that the site was empty at t and the species colonized it between t and $t + 1$. The Markovian dependence accommodates the autocorrelation in occupancy probability, as in an AR(1) autoregressive time series. The authors considered a simple model with two seasons and three replicate surveys per site and season. Let ψ_1 be the occupancy probability in the first season, along with parameter p , the detection probability given present. The probability of observing in site i a detection history $h_i = 110\,000$ is then, assuming $\psi_t = \psi$ and $\varepsilon_t = \varepsilon$,

$$\Pr(h_i) = \psi_1 p^2 (1 - p)[\varepsilon + (1 - \varepsilon)(1 - p)^3].$$

Here, we have the probability that the site was initially occupied and the species was detected in two replicate surveys and missed in one and either the site became empty in the second season or it remained occupied and the species was missed in all surveys (MacKenzie et al., 2003). Note that the multiple-season occupancy data can be analyzed as independent separate single-season model, where independence assumes $\varepsilon = 1 - \gamma$.

The above model was extended by Gibson-Reinemer et al. (2016). Let y_{ijt} take the value 1 if a species is detected at site i , survey j during year t ($t = 1, \dots, T$), given present, with probability p_{ijt} , and z_{it} is the usual indicator variable for presence (occupancy) with $\psi_t = \Pr(z_{it} = 1)$. Then $y_{ijt} | z_{it} \sim \text{Bernoulli}(z_{it} p_{ijt})$.

The initial occupancy state was modeled as $z_{i1} \sim \text{Bernoulli}(\psi_{i1})$. Occupancy probabilities in subsequent years were

$$\psi_t = \psi_{t-1}\phi_{t-1} + (1 - \psi_{t-1})\gamma_{t-1},$$

for $t = 2, \dots, T$. Also,

$$z_{i,t+1} | z_{it} \sim \text{Bernoulli}(\pi_{i,t+1}),$$

where

$$\pi_{i,t+1} = z_{it}\phi_t + (1 - z_{it})\gamma_t.$$

Here, $\phi_t = \Pr(z_{i,t+1} = 1 | z_t = 1)$ is the local survival probability that the site remains occupied from survey t to survey $t + 1$. $\text{Logit}(p_{ijt})$ and $\text{logit}(\psi_{11})$ were modeled as linear regressions involving covariates (environment and gear type) with regression parameters having normal priors $N(0, 0.01)$. The theory was applied to 52 species of large-river fishes.

Royle and Kéry (2007) provided a Bayesian state-space formulation. They mentioned several other parameters of interest. For example, the growth rate $\lambda_t = \psi_{t+1}/\psi_t$, while the “turnover” is the probability that an occupied quadrat picked at random is a newly occupied one, namely,

$$\tau_t = \Pr(z_{t-1} = 0 | z_t = 1) = \gamma_{t-1}(1 - \psi_{t-1})/\psi_t.$$

Gimenez et al. (2014) showed how occupancy models can be implemented in program **E-SURGE**, which was initially developed to analyze capture-recapture data in the HMM framework. Here, the state process captures the Markovian dynamic of the actual but latent states, while the observation process consists of observations that are made from these underlying states. Replacing individuals by sites provides the user with access to several features of **E-SURGE**. Examples are given of some scenarios along with simulation and analysis of some real data.

Occupancy Related to Habitat

Previously, we have looked at models relating occupancy and habitat. An important topic is determining habitat suitability for occupancy. Martin et al. (2010) used a dynamic site occupancy model to determine the suitability of waterholes for elephants. They derived estimates of annual transition probabilities among three dry season states for water holes: (1) unsuitable state (dry water holes with no elephants), (2) suitable state (water hole with water) with low abundance of elephants, and (3) suitable state with high abundance of elephants.

We now consider dynamic state models that relate occupancy to changes in the habitat described by MacKenzie et al. (2011). They used the following notation. Let

$$\begin{aligned}
 \pi^{(H)} &= \text{Pr}(\text{sampling unit is of habitat state } H \text{ in the first season}) \\
 \psi^{(H)} &= \text{Pr}(\text{species present at a unit of habitat state } H \text{ in the first season}) \\
 \eta_t^{(y_t, H_t, H_{t+1})} &= \text{Pr}(\text{habitat changes from state } H_t \text{ in season } t \text{ to state } H_{t+1} \\
 &\quad \text{in season } t + 1, \text{ given the species was either present } (y = 1) \\
 &\quad \text{or absent } (y = 0) \text{ from the sampling unit in season } t) \\
 \gamma_t^{(H_t, H_{t+1})} &= \text{Pr}(\text{species colonizes a unit between seasons } t \text{ and } t + 1 \\
 &\quad (\text{i.e., species absence to presence}), \text{ given habitat transitioned} \\
 &\quad \text{from state } H_t \text{ in season } t \text{ to state } H_{t+1} \text{ in season } t + 1) \\
 \varepsilon_t^{(H_t, H_{t+1})} &= \text{Pr}(\text{species goes locally extinct from a unit between seasons } t \\
 &\quad \text{and } t + 1 (\text{i.e., species absence to presence}), \text{ given the habitat} \\
 &\quad \text{transitioned from state } H_t \text{ in season } t \text{ to state } H_{t+1} \\
 &\quad \text{in season } t + 1) \\
 p_{tj}^{(H_t)} &= \text{Pr}(\text{detecting the species in survey } j \text{ of season } t, \text{ given the} \\
 &\quad \text{habitat state in season } t \text{ is } H)
 \end{aligned}$$

Suppose that during three sampling seasons, the habitat at a sampling unit is in state A in the first two seasons and state B in season 3 (i.e., $H = AAB$). Over the same period, the unit is surveyed twice per season (because of imperfect detection of the focal species) resulting in the following species detection history, $h = 10\ 00\ 11$, which denotes that the species was detected in the first survey of season 1, never detected in season 2, and detected twice in season 3. This means that the unit is in habitat state A in the first season, with the species being present and detected in the first survey, but not in the second survey ($\pi^{(A)}\psi^{(A)}p_{11}^{(A)}(1 - p_{12}^{(A)})$). Between seasons 1 and 2, that habitat remained in state A (i.e., habitat did not change), given the species was present in season 1 ($1 - \varepsilon_1^{(1,A,B)}$) and either (i) the species did not go locally extinct (so it was still present in season 2) but was not detected in either survey, and then between seasons 2 and 3, the habitat changed from state A to B and the species did not go locally extinct with probability

$$(1 - \varepsilon_1^{(A,A)})(1 - p_{21}^{(A)})(1 - p_{22}^{(A)})(\eta_2^{(1,A,B)})(1 - \varepsilon_2^{(A,B)});$$

or (ii) the species went locally extinct (so was absent and could not be detected in season 2), and then between seasons 2 and 3, the habitat changed from state A to B , given the species was absent in season 2 and the species colonized the unit

$(\varepsilon_1^{(A,A)} \eta_2^{(0,A,B)} \gamma_2^{(A,B)})$. Finally, given the unit was in habitat state B and occupied by the species in season 3, the species was detected in both surveys ($p_{31}^{(B)} p_{32}^{(B)}$). Therefore, the associated probability of observing this entire detection history can be expressed as

$$\begin{aligned} & \Pr(H = AAB, h = 100011) \\ &= \pi^{(A)} \psi^{(A)} p_{11}^{(A)} (1 - p_{12}^{(A)}) (1 - \eta_1^{(1,A,B)}) \\ &\quad \times (1 - \varepsilon_1^{(A,A)})(1 - p_{21}^{(A)})(1 - p_{22}^{(A)}) (\eta_2^{(1,A,B)})(1 - \varepsilon_2^{(A,B)}) + \varepsilon_1^{(A,A)} \eta_2^{(0,A,B)} \gamma_2^{(A,B)} \\ &\quad \times p_{31}^{(B)} p_{32}^{(B)}. \end{aligned}$$

During a sampling season, with only two habitat states and two levels of occupancy (presence/absence), a sampling unit can be in one of four mutually exclusive states: (i) habitat A and species absent, (ii) habitat A and species present, (iii) habitat B and species absent, and (iv) habitat B and species present. Therefore, between seasons, there are 16 possible transitions to move from any of the 4 possible states in season t to any of the 4 states in season $t + 1$. To start the process, in the first season, the probability of a sampling unit being in each of the four possible states can be conveniently expressed as the vector

$$\boldsymbol{\phi}_0 = (\pi^{(A)}(1 - \psi^{(A)}), \pi^{(A)}\psi^{(A)}, (1 - \pi^{(A)})(1 - \psi^{(B)}), (1 - \pi^{(A)})\psi^{(B)})',$$

where the states are as previously ordered. Between seasons, the probability of a unit transitioning from one state to another could be defined as a 4×4 transition probability matrix (Φ_t), where rows indicate the state of a unit at time t and columns indicate the state at $t + 1$. This is a large 4×4 matrix, and details are given by the authors.

Let \mathbf{h} denote a detection history that includes information on both habitat and species detection, and let $\mathbf{p}_t^{(\mathbf{h})}$ be a state-dependent, conditional detection probability vector. For example, suppose in season t , the habitat was state A and the detection history 011 was observed, then

$$\mathbf{p}_t^{(A,011)} = (0, (1 - p_{t1}^{(A)})p_{t2}^{(A)}p_{t3}^{(A)}, 0, 0)'.$$

The unit was known to be in state A and occupied, so the latter two elements must be zero (these elements are associated with the two occupancy states for habitat B), and the probability of getting at least one detection if the unit was habitat A but unoccupied by the species (the first element) must also be zero. There is only one non-zero element because when the species is detected and habitat state is known, there is only one possible true state for the unit in that season. However, when the species is never detected in a season, there are two possible states resulting in two

non-zero elements. For example, suppose a unit is habitat state B at time t , but the species was never detected. The conditional detection vector would be

$$\mathbf{p}_t^{(B,000)} = (0, 0, 1, (1 - p_{t1}^{(B)})(1 - p_{t2}^{(B)})(1 - p_{t3}^{(B)}))'.$$

If the site was unoccupied by the species (element 3), the probability of never seeing the species is 1, but if the species had been present (element 4), then it was not detected in any of the three surveys. The probability of observing a particular detection history at a unit for all T seasons can then be obtained using matrix multiplication which automatically accounts for the ambiguity caused by imperfect detection of the species, namely,

$$\Pr(\mathbf{h}_i \mid \boldsymbol{\theta}) = \boldsymbol{\phi}'_0 \left[\prod_{t=1}^{T-1} \mathbf{D}(\mathbf{p}_t^{(\mathbf{h})}) \boldsymbol{\Phi}_t \right] \mathbf{p}_T^{(\mathbf{h})},$$

where $\boldsymbol{\theta}$ is just a vector of all the parameters to be estimated and $\mathbf{D}(\mathbf{p}_t^{(\mathbf{h})})$ is a diagonal matrix with the elements of $\mathbf{p}_t^{(\mathbf{h})}$ on the main diagonal (top left to bottom right) and zero elsewhere. The first and last terms are outside of the product statement to ensure the matrix multiplication yields a single value. Assuming the detection histories are independent for each unit, the joint probability for the data (and the model likelihood) is

$$\Pr(\mathbf{h}_1, \dots, \mathbf{h}_s) = L(\boldsymbol{\theta} \mid \mathbf{h}_1, \dots, \mathbf{h}_s) = \prod_{i=1}^s \Pr(\mathbf{h}_i \mid \boldsymbol{\theta}),$$

where s is the total number of units surveyed during the study. The model likelihood could then be used to provide maximum likelihood estimates of the parameters or else used in a Bayesian framework to obtain posterior distributions of parameters. The same basic model structure can be recast into a data augmentation approach (Appendix A.5). The authors gave a number of extensions to the model.

Mosher et al. (2018) proposed a dynamic multistate modeling approach that is flexible enough to account for the detection structures that may be present in complex multistate systems, for example, where unobservable states exist. They considered a host-pathogen system as a general example to first illustrate how misleading inferences can result from failing to incorporate pathogen dynamics into the modeling process, especially when the pathogen is difficult or impossible to survey in the absence of a host species. They advocated the use of time-varying covariates to explain temporal trends in the data and the development of sampling techniques that match the biology of the system to eliminate unobservable states when possible.

4.18 Disease Modeling

In addition to multistate capture-recapture models, occupancy methods have been used for disease modeling (e.g., Murray et al., 2009; Kendall, 2009). McClintock et al. (2010), in a very comprehensive paper, discussed using occupancy methods incorporating a spatiotemporal model with repeat observations to investigate disease (state) prevalence, particularly when there are various levels of uncertainty attached to a diagnosis. Cooch et al. (2010) reviewed a variety of methods of modeling disease dynamics for wild populations including capture-recapture and occupancy in a state-space framework. Such methods are important for disease modeling in humans (e.g., Chaudhary et al., 2020; Section 12.2.4), as well as in wildlife because of the emergence and re-emergence of a number of pathogens causing disease in both humans and livestock.

4.18.1 Underlying Problems

There is uncertainty associated with state assignments due to non-detection or misclassification that needs to be modeled. For example, there are two types of error: firstly false negatives may arise due to a specific sample missing the pathogen or the failure of an assay to identify the pathogen in a positive sample, and secondly false positives due to falsely detecting infection, given absence of the pathogen. Lachish and Murray (2018) provided a helpful review to assist researchers and practitioners to navigate the pitfalls of uncertainty in wildlife disease ecology studies. Disease mapping is discussed by Bivand et al. (2013): Chapter 10, and Conn et al. (2012) reviewed and considered the estimation of force of infection, the instantaneous rate at which susceptible individuals acquire infection.

As an example of some of the issues involved, we refer to *Bd*, a wetland fungus affecting amphibians worldwide, that was considered by Murray et al. (2009) as one of the pathogens, as follows. Suppose three sites are sampled and in the simple case in which false positives are not possible, a detection history of 011 would indicate an unambiguous *Bd* state assignment for the site (wetland). Here, the pathogen was present in the wetland, but failed to be detected in the first of the three samples. As an example, the detection of the pathogen at site i might have the history (001) with probability

$$\Pr(H_i = 001) = \psi(1 - p_1)p_2p_3,$$

where ψ is the probability that *Bd* is positive and p_t is the probability that *Bd* is detected in the t th sample (i.e., the sample contains the pathogen and it is detected by the assay) and conditional on the site (wetland) i being infected. The parameter ψ can also be interpreted as the expected proportion (i.e., spatial prevalence) of *Bd*-positive wetlands within the population area. The ambiguous detection history 000

would have event probability

$$\Pr(H_i = 000) = \psi \prod_{t=1}^3 (1 - p_t) + (1 - \psi,)$$

where the additive terms account for the possibility that the wetland was *Bd* positive (but the fungus was not detected by sampling) or the wetland was truly *Bd* negative. The likelihood for the observed data is then

$$L(\psi, \{p_i\}) \propto \prod_{i=1}^S \Pr(H_i)$$

and can be used to obtain maximum likelihood estimates. The authors discussed the assumptions involved and considered how some of them could be relaxed. A time factor (year) and transition probabilities were also introduced, and examples of detection histories were given for this complex situation along with helpful flow diagrams. Other extensions included the use of logistic regression using covariates and incorporating other auxiliary information.

Miller et al. (2012c) also used *Bd* as an example and noted that in addition to infection prevalence, it is also important to estimate the distribution of infection intensities based on repeated sampling of individuals in the wild. Suppose n individuals are sampled, with a subset sampled at least twice and individual i sampled T_i times. Let $z_i = 1$ if the i th sampled individual is infected, with probability ψ . Each infected individual has an infection intensity x_i which may be measured with error. Let $y_{it} = 1$ if individual i from sample t has a detected infection with probability $p_i = \Pr(y_{it}) = 1 | z_i = 1, x_i)$ and a measured intensity of infection w_{it} . Then we have the usual equations

$$z_i \sim \text{Bernoulli}(\psi), \quad y_{it} \sim \text{Bernoulli}(z_i p_i),$$

as well as

$$x_i \sim N(\mu_x, \sigma_x^2), \quad w_{it} \sim N(x_i, \sigma_w^2).$$

The two sets of equations are linked using

$$\text{logit}(p_i) = \alpha + \beta x_i,$$

with normal priors $N(0, 0.001)$ for μ and β , uniform priors on $[0,1]$ for ψ and $\text{logit}(\alpha)$, and uniform priors on $[0,10]$ for σ_x and σ_w . Extensions and further details are given by the authors.

Chaudhary et al. (2020) extended the previous model to when there are S tissues (sites) within a host, where for a single tissue that may be infected by a pathogen, we again score 1 for disease present (state 1) and 0 for not present (state 2). They

considered both when the tissues are independent and when they are dependent, where infection in one tissue may effect what happens in another tissue. They applied their method to feral swine (*Sus scrofa*) where the tissues were from oral, nasal, and genital mucosa.

4.18.2 Multiple Data Sources

Hanks et al. (2011) considered how multiple data sources, varying in extent and accuracy, can be combined using a Bayesian hierarchical framework to improve the accuracy of large-scale prediction of forest disease incidence. They applied their method to study the incidence of eastern spruce dwarf mistletoe (*Arceuthobium pusillum*) in Minnesota black spruce (*Picea mariana*). It was estimated that 35–59% of black spruce stands in northern Minnesota are infested with dwarf mistletoe. They assumed that the response variable of interest was observed in two separate surveys, one of which, denoted here as D_A , is assumed to be more accurate than the other, denoted as D_L . They assumed that the observations in D_A are collected on a representative subset of the observation locations for which we have the less accurate large set of data, D_L , and on the same spatial scale. The D_L data came from the Minnesota Department of Natural Resources (DNR) survey.

Let y_i denote the presence ($y = 1$) or absence ($y = 0$) of mistletoe in a stand, as identified in the intensive survey, and w_i is the DNR inventory presence or absence in the same stand; then we have

$$\begin{aligned} w_i \mid \phi_i, \theta_i, y_i &\sim \text{Bernoulli}(\phi_i), \text{ if } y = 1, \\ &\sim \text{Bernoulli}(\eta_i), \text{ if } y = 0, \end{aligned}$$

where ϕ_i is the probability of the DNR survey finding mistletoe if it is present in the intensive survey ($1 - \phi$ is the probability of a false negative) and η_i is the probability of the DNR survey reporting mistletoe present if it is not present in the intensive survey (η_i is the probability of a false positive). Here, D_L corresponds to the DNR inventory (w), and D_A corresponds to the intensive survey (y). They used the probit link, denoted as Φ^{-1} , which is the inverse cumulative distribution function of the standard normal distribution and assumed

$$y_i \mid \boldsymbol{\beta} \sim \text{Bernoulli}(\theta_i), \Phi^{-1}(\theta_i) = \mathbf{x}'_i \boldsymbol{\beta},$$

where θ_i is the latent probability that stand i is infected and depends on the DNR stand characteristics in \mathbf{x}_i . The parameters ϕ_i and η_i could also vary with stand characteristics in the DNR inventory. These were modeled separately as $\Phi^{-1}(\phi_i) = \mathbf{x}'_i \boldsymbol{\beta}_\phi$ and $\Phi^{-1}(\eta_i) = \mathbf{x}'_i \boldsymbol{\beta}_\eta$.

To allow for spatial autocorrelation, additional covariates were added based on a neighborhood structure. Priors were specified for the regression parameters $\boldsymbol{\beta}$, $\boldsymbol{\beta}_\phi$,

and β_η . Posterior distributions were developed for the N sites for the observations from both the DNR inventory and the intensive survey, where N_1 is the number of sites where mistletoe was present in the intensive survey and N_0 is the number of sites where mistletoe was absent in the intensive survey. The next step was to find the probability of mistletoe presence in a stand surveyed by the DNR, but not by the intensive survey. All full conditional distributions of parameters in the model were found analytically, and necessary codes for computations using **R** were given in a supplement. Further details are given by Hanks et al. (2011), our aim here being to present the general idea of combining the two such types of data.

4.18.3 Dynamic Model

Bled et al. (2011) focused on the dynamics of an invasive spread and the possibility of building on the work of MacKenzie et al. (2003). Although our focus is on closed populations, it seems appropriate to briefly raise the topic here. Let $z_{it} = 1$ if cell i ($i = 1, \dots, M$) is occupied at time t ($t = 1, \dots, T$) and 0 otherwise. Then

$$z_{it} \mid z_{i,t-1} \sim \text{Bernoulli}(\mu_{it}),$$

and

$$\begin{aligned} \mu_{it} &= \Pr(z_{it} = 1 \mid z_{i,t-1}) \\ &= z_{i,t-1}\phi_{it} + (1 - z_{i,t-1})\gamma_{it}, \end{aligned}$$

where $\phi_{it} = \Pr(z_{i,t+1} \mid z_{i,t} = 1)$ is the probability of site i being occupied at time t and occupied at time $t + 1$ again and $\gamma_{it} = \Pr(z_{i,t+1} \mid z_{i,t} = 0)$ is the local colonization probability. The authors then split γ_{it} into two components, γ_{it} , which is now the probability of the first colonization, and θ_{it} , a recolonization parameter. Let $A_{it} = 1$ if the site has never been occupied or colonized before and 0 otherwise, so that

$$A_{it} = \prod_{k=1}^{t-1} (1 - z_{ik}) \text{ and } A_{iT} = 1 \text{ if } \sum_{t=1}^T z_{it} = 0.$$

This means that sites will have different colonization probabilities depending on whether $A_{it} = 1$ or $A_{it} = 0$, namely, γ_{it} and θ_{it} , respectively. The model now becomes

$$z_{i,t+1} \mid z_{i,t-1} \sim \text{Bernoulli}(\mu_{i,t+1}),$$

with

$$\mu_{i,t+1} = \phi_{it} z_{it} + \gamma_{it}(1 - z_{it})A_{it} + \theta_{it}(1 - z_{it})(1 - A_{it}).$$

Let N_i represent the set of cells that are neighbors of the cell i and n_i the actual number. Then define a spatiotemporal autocovariate D_{it} as

$$D_{it} = \sum_{j \in N_i} z_{jt} w_{ij},$$

where w_{ij} , which can be expressed in different ways, is a measure of the effect of neighboring site j on site i . The authors expressed logistic regressions of ϕ_{it} , γ_{it} , and θ_{it} as quadratic regressions in D_{it} .

The above ecological component of the model incorporates time and spatial dimensions, and we now consider the observation component. Let K_{it} be the number of visits or replications in the cell i at time t , let y_{it} denote the total number of stops at which an individual is detected during the visits, and let p_{it} be the probability of detection at a stop, *given* present. Then

$$y_{it} \sim \text{Binomial}(K_{it}, p_{it} z_{it}).$$

Some model adaptions are considered, and a Bayesian analysis, via Markov chain Monte Carlo, is implemented using **WinBUGS**.

Disease modeling for humans is also considered in Sect. 12.2.4.

4.19 Multiple Data Methods

We now consider a number of methods where different sources of data are used together. We have already referred to this extension with respect to Sect. 4.18.2 dealing with invasive species. Also, this has been mentioned with regard to citizen science in Sect. 4.16.1. Looking at other species, we note that the behavioral patterns and large territories of large carnivores like wolves make the animals challenging to monitor. Ausband et al. (2014) used combined data from hunter surveys, howling and sign surveys conducted at predicted wolf rendezvous sites, and locations of radio-collared wolves to model occupancy and estimate the number of gray wolf (*Canis lupus*) packs and individuals in Idaho during 2009 and 2010. They allowed for false positives, namely, misidentifications, and made three further assumptions: (1) average territory size is known, (2) average pack size is known, and (3) territories do not overlap. Occupancy estimation followed the methods presented by Miller et al. (2011), where one can estimate both false-negative and false-positive errors among uncertain detections.

Extensive mail surveys were carried out, and, from predicted rendezvous sites, the presence and abundance of wolf tracks, scat, hair, daybeds, and kills were

recorded. Where possible, genetic samples (i.e., scat and hair) were collected to verify species and determine the number of unique individuals at a site. Radio-telemetry data from radio-collared wolves was also collected. The 3 sources of data (hunter surveys, rendezvous site surveys, locations of radio-collared wolves) yielded a total of 14 sampling occasions per year, 12 from hunter surveys, 1 from rendezvous site surveys, and 1 from radio-collar locations. The method of Miller et al. (2011), including the use of covariates, was used to develop an occupancy model based on data from the three survey methods to produce annual estimates of number and distribution of wolf packs and wolf abundance, with associated measures of precision. Crum et al. (2017) used hunter survey data to estimate the occupancy probability of moose. The use of hunter surveys comes under the topic of citizen science.

Dey et al. (2017) considered the estimation of animal abundance and distribution over large regions using two sources of data, namely, data from reliable but expensive surveys conducted at smaller scales and cost-effective but less reliable data generated from surveys at wider scales. In particular, using Bayesian methods, they combined abundance estimates of tigers in wildlife reserves from intensive photographic capture-recapture methods and estimates of tiger habitat occupancy from indirect sign surveys, conducted over a wider region as follows.

Dividing the region into a large number of sites, assume that N_s , the number of individuals in site s , is Poisson(λ_s), and that the N_s are independently distributed. A reliable estimate $\hat{\lambda}_{s,ct}$ from camera trap data is available from sites $s = 1, \dots, k_1$. Additionally, for all the sites, including the ones with camera trap data, estimates of occupancy probabilities are available, namely, $\hat{\psi}_1, \dots, \hat{\psi}_k$, $k = k_1 + k_2$, are obtained from count data on animal signs such as tracks or scats detected along surveyed trails. Attention is restricted to the estimation of abundance for the k sites, with particular emphasis on the k_2 sites not part of the camera trap survey.

Since $\psi_s = \Pr(N_s > 0) = 1 - \Pr(N_s = 0) = 1 - \exp[-\lambda_s]$, we have $\lambda_s = -\log(1 - \psi_s)$ as an index of true animal abundance that accounts for the imperfect detection of signs. We can therefore consider the estimate $\hat{\psi}_{s,os}$, $s = 1, \dots, k$ for the occupancy survey as indicative estimates of abundances for all the sites. They can be used as an index wherever reliable estimates are unavailable. Let $v_s = \log(\lambda_s)$, $\hat{v}_{s,ct} = \log(\hat{\lambda}_{s,ct})$, $\hat{v}_{ct} = (\hat{v}_{1,ct}, \dots, \hat{v}_{k_1,ct})'$, and $\hat{\lambda}_{os} = (\hat{\lambda}_{1,os}, \dots, \hat{\lambda}_{k_1,os}, \dots, \hat{\lambda}_{k,os})'$. The last two vectors and other covariates (\mathbf{z}_s) give us the data. If \mathbf{x}_s is the vector of $\hat{\lambda}_{s,os}$ and \mathbf{z}_s (or their transforms), it is assumed that

$$\mathbf{v}_s = \boldsymbol{\beta}' \mathbf{x}_s + \boldsymbol{\phi}_s, \quad s = 1, \dots, k,$$

where $\boldsymbol{\phi}_s$ is the parameter that incorporates spatial correlation between neighboring sites. It is assumed that

$$\boldsymbol{\phi}_s \mid \boldsymbol{\phi}_j, j \neq s \sim N\left(\sum_{j \neq s} a_{sj} \phi_j, \sigma_s^2\right), \quad i = 1, \dots, k,$$

where the intended “neighborhood structure” is reflected by the choice of the a_{sj} ’s. This means that ϕ has a multivariate normal distribution with the usual conditionally autoregressive (CAR) structure. Various normal prior distributions for parameters are used, and the total abundance as defined by $A(\lambda) = \sum_{s=1}^k \lambda_s = \sum_{s=1}^k \exp[v_s]$ is estimated. Further details and expressions for the hierarchical Bayes estimate of total abundance and its posterior variance are given by Dey et al. (2017).

4.20 Fisheries and Marine Environments

Methods for using fishery models are considered throughout this book as, for example, using distance methods in Chaps. 7 and 8. A variety of methods can be used as different species often require special methods. For example, Hammond et al. (2013) used shipboard transects and aerial transects for cetaceans. Acoustical methods were found useful for species that sometimes disappeared below the surface. Large jellyfish can be difficult to study because they are highly aggregated, but video methods can be used (Graham et al., 2003). Aerial methods can also be used for surface aggregations of jellyfish (Houghton et al., 2006). General methods for jelly fish are described by Purcell (2009). However, our focus here is on occupancy and detectability.

By considering reef fish, Edgar et al. (2004) highlighted some of the problems such as diver bias and detectability. In the latter case, catches from fishing are mentioned as a fish may not be caught (detected) even though present. Here, catches (which may be numbers or biomass) are used, and the difference here is that if a fish is detected, it is “removed.” Methods for studying reef fish were considered by Caldwell et al. (2016). They note that a variety of visual census techniques are employed; however, results are often incomparable due to differential methodological performance. This will apply to other species.

Groce et al. (2012) investigated the effectiveness of translocation as a management strategy to conserve the threatened Arkansas darter *Etheostoma cragini* in the Arkansas River basin of southeastern Colorado. Their goal was to evaluate the success of the translocation and the habitat factors that occupancy at suitable (wet) sites that are most important to promote persistence and colonization and thus translocation success. Dipnetting surveys were conducted by a two-person crew, and a single-season multistate occupancy model incorporating covariates measured at three spatial scales was used. The observations at each site were classified as having one of three states: dry and thus unoccupied by darters (0), wet but no darters detected (1), or wet and darters detected (2). Data sets were created for each age group and were used to estimate the probability that site i contained suitable (wet) habitat (ψ_{1i}); the conditional probability that darters occurred at site i , given that it was wet (ψ_{2i}); and the probability of detecting darters at site i during survey t (p_{2it}), given that the site was wet and occupied (i.e., in true state 2). The unconditional probability that site i was wet and occupied was $\psi_{1*2i} = \psi_{1i}\psi_{2i}$. For each site i , a

detection history (h_i) for each age class of darters was compiled. Various covariates were considered.

Thorson and Ward (2013) focused on so-called delta generalized linear mixed models (delta-GLMMs) for fisheries that incorporate information about both the probability of catch being non-zero and the expected value for non-zero catches (catch rates). If c_i is the non-zero catch in weight for tow i , with strata s_i and year y_i , then $\Pr(C = c_i > 0 | s_i, y_i)$ is expressed as a logistic regression model with main effects and interactions for strata and year. The density function of C was modeled as $\text{Gamma}(c | \alpha, \beta_i)$, where $\alpha = 1/CV^2$ and $\beta_i = 1/(\mu_i CV^2)$. Here, CV is the estimated coefficient of variation for C given that $C > 0$, and μ_i (a function of a_i , the area swept by tow i) is the expected value of catch i for a non-zero tow. They then used a log-linear model for μ_i with main effects and interactions. The stratum \times year interactions were treated as random effects using a bivariate normal distribution, and prior distributions were used for all parameters. Four models were proposed and used to estimate an index of abundance for all 28 species considered.

The above model was adapted by Thorson and Ward (2014) to include a vessel-year effect. This was extended to a geostatistical model by Thorson et al. (2015) that specifies population densities at nearby sites are more similar than densities at geographically remote sites. Then followed a series of papers with various adaptions with lead author James Thorson listed in Thorson (2019), where a package called **VAST** (vector autoregressive spatiotemporal) develops a very general model.

Katsanevakis et al. (2012) gave a comprehensive and useful review of methods used for marine populations that involve imperfect detection, including occupancy methods and methods described in other chapters. In plot sampling, the size of the plot depends on the size of the species being sampled and the relevant sampling technique. Small plots, say no bigger than 1 m^2 , are generally sufficient for small-sized organisms. Strip transects for surveys of fish or large invertebrates by divers usually have a surface of some hundreds of square meters, while strips in trawl surveys may extend many thousands of square meters. Incomplete detectability will generally need to be considered with underwater surveys. Various devices and gear equipment for sampling are described. Also considered are distance methods, mark-recapture and removal methods, and catch-effort methods.

Baker et al. (2017) considered a variety of sampling methods modeling the occupancy of Arctic grayling (*Thymallus arcticus*) in the Canadian Barrenlands. They assessed the reliability and efficiency of alternative sampling strategies: (i) randomly distributed versus sequential adjacent spatial replicates, (ii) visual versus electrofishing surveys, and (iii) spatial versus temporal replicates. Sequential, adjacent spatial replicates produced spatially autocorrelated data.

4.21 Conclusion

Models for occupancy, where it is important to determine whether a particular species is present or not, have a long history. The literature is very extensive, as referred to in some reviews mentioned, and both frequentist and Bayesian methods

have been developed for all kinds of population, including those of animal signs. The methods are particularly important in dealing with the incomplete detection of rare species or species with a low detection rate and with detecting the presence of an invasive species.

In addition to the basic models used in the past based on single surveys, there have been a number of extensions to include multiple surveys, multiple seasons, multiple species, and multistate models. The use of so-called robust models with primary and secondary surveys is discussed. The aim is to either overcome or allow for violations of traditional model assumptions such as nonclosure, non-independent replicates, heterogeneous detection, and false positives. Key assumptions are listed, and appropriate tests for them are described in two places using goodness-of-fit tests and residual plots for covariate regressions, as well as tests for occupancy differences.

As well as estimating detection probabilities, abundance estimates can be made with some models. Design questions are considered in several places, such as determining the number of visits to a site and dealing with possible dependent visits. Covariates have also been used more frequently, and these help to reduce the number of unknown parameters.

A general collection of various models are presented including an introduction to spatial models with spatial autocorrelation, time-to-detection removal models, and staggered arrival and departure times. Adaptive-type methods using a two-phase design or a conditional design are given for rare species. Multi-scale models are considered as species distributions depend critically on the scales at which data is used for model building.

Types of observation errors are discussed as well as combining types of detection using multiple detection devices. DNA and eDNA (environmental DNA) as genetic tags are being used increasingly. Sometimes, there is incomplete data information such as having presence-only data, but extra environmental information can often be incorporated. In addition to using specialist observers, it can be cost-effective to use “opportunistic” data such as data from volunteers, called “citizen” science.

Occupancy models are extended to allow for the possibility that individuals are being detected in one of the several states (multistate models), including disease modeling. Undetected states are considered along with the use of dynamic state models. It is shown how multiple data sources can be combined. The chapter ends with a brief discussion on fisheries and marine environments.

Chapter 5

Species Methods



5.1 Introduction

This chapter is essentially a follow-on from the previous chapter, but with an additional level of complexity. One could simply add an additional subscript to allow for species type. This means that some of this chapter deals with general issues rather than more specific mathematical modeling. However, there are some additional topics that need consideration.

Firstly, each species may have a different detection distribution, and not just having different parameter values. Secondly, there is the problem of species misidentification, so-called positive detection error, especially if various species are difficult to distinguish, which may include the need for specialized input. Thirdly, there is an interest in knowing how many species there are in an area when there is incomplete detection, called species richness (Sect. 5.5), with the associated measures of biodiversity (Sect. 5.5.11). Fourthly, species may interact with one another, especially if one of the species is an invasive one (Sect. 5.4). This can involve multivariate modeling (Sect. 5.3). Notation is a problem, and we will sometimes change it, especially with multiple species, so that original articles can be viewed.

Many individual species are on the decline, and many are on the endangered list, such as large carnivores, for instance. Various methods have been developed to estimate population numbers such as individual counts, distance sampling, mark-recapture, call-in surveys, camera surveys, and track counts. The latter are the most commonly used, for example, with lions using N-mixture models (Belant et al., 2016).

A basic hierarchical model for several species is as follows. Let

$$z_{ij} \sim \text{Bernoulli}(\psi_{ij}),$$

where z_{ij} is the latent occupancy state of species i at site j (occupied if $z_{ij} = 1$ and unoccupied if $z_{ij} = 0$). Here, ψ_{ij} is the occupancy (presence) probability of site j by species i and may be expressed as a logistic regression of the form

$$\text{logit}(\psi_{ij}) = \beta_i + \boldsymbol{\beta}'_1 \mathbf{x}_j,$$

where for species i , \mathbf{x}_j is a vector of covariates (e.g., habitat variables). Let p_{ijt} be the probability of detection for species i , given present, for an observation used at site j for visit t , and let y_{ijt} be a binary indicator of detection ($y_{ijt} = 1$) or non-detection ($y_{ijt} = 0$) for site j of species i for visit t . We can then assume that

$$y_{ijt} \sim \text{Bernoulli}(\psi_{ij} p_{ijt}),$$

where $\text{logit}(p_{ijt})$ can also be modeled as a regression of covariates, with random effects sometimes added. The regression parameters can be given prior distributions.

Another layer of structure can be added if one uses a superpopulation approach with data augmentation (Appendix A.5). Instead of a Poisson distribution controlling the abundance process, we can use a zero-inflated Poisson, a negative binomial, or another suitable distribution. Various versions of this approach are available. For example, Yamaura et al. (2011) supposed that there are N_{ij} individuals of species i on site j and individuals are detected independent of one another with probability r_{ij} . If y_{ij} is the total number detected in v_j visits for site j , and p_{ij} is the probability that at least one individual is detected, then $p_{ij} = 1 - (1 - r_{ij})^{N_{ij}}$. It can then be assumed that $y_{ij} \sim \text{Binomial}(v_j, p_{ij})$, where v_j is the number of visits to site j . If N_{ij} is Poisson(λ_{ij}), linear covariate models can be fitted to $\log(\lambda_{ij})$ and $\text{logit}(r_{ij})$. Bayesian analysis can then be carried out using data augmentation.

Of particular interest are rare and elusive species, which will have special mention from time to time, especially when it comes to experimental design. For example, with single-species studies and rare species, it is better to sample more sites fewer times, whereas for common species, it is better to sample fewer sites on more occasions.

5.2 Species Distribution Models (SDMs)

The geographic distribution over which a single species occurs is one of its most fundamental and defining traits for species methods. In Chap. 4, we referred to modeling a species distribution, which we now consider further under the title of species distribution models (SDMs) before considering multiple species models.

5.2.1 *Background and Reviews*

There is a very large literature on the subject (Franklin, 2013; Zurell, Franklin et al., 2020a) because of its importance, and there are a number of earlier reviews of Guisan and Thuiller (2005, mainly descriptive) and Elith and Leathwick (2009), as well as later ones described below. The latter authors described an SDM “as a model that relates species distribution data (occurrence or abundance at known locations) with information on the environmental and/or spatial characteristics of those locations.” Such models are applied across terrestrial, freshwater, and marine environments, at widely varying spatial and temporal scales. Some important aspects considered are scale size, climate variables, hierarchical models, geographic versus environmental space, explanation versus prediction, regression models, and modeling presence data.

There are a large number of modeling algorithms to predict species ranges from point locality data. Wisz et al. (2008) compared the effect of sample size on the performance of 12 algorithms using independent presence-absence data from multiple species and regions. Zurell, Franklin et al. (2020a) pushed for a better standardization and documentation of modeling protocols, which were currently limited. They proposed a protocol, an extension of an older protocol ODD, called the ODMAP (Overview, Data, Model, Assessment, and Prediction), as its components reflect the main steps involved in building SDMs and other empirically based biodiversity models. It provides a useful and extensive guide for researchers. Fletcher (2018) provided a practical guide for combining data to model species distributions and gave a survey of 353 articles combining 2 or more sources of data under the categories pooling, independent models, auxiliary data, informed priors, and integrated models.

Guillera-Arroita et al. (2015) reviewed the problem of matching data and models. They noted that species distribution methods have been applied to the management of threatening species, to controlling threatening processes, to predict impacts of climate change, to understand phylogeographic patterns, and to manage landscapes and biological invasions. A recent extensive review of species distributions was given by Martin-Schwarze et al. (2018) who focused more on the statistical issues and methods. They particularly emphasized the need to incorporate spatial autocorrelation because of the proximity of observations, which has generally been ignored in the past. Spatial data almost always arises from a spatiotemporal generating process observed over a fixed interval of time. For example, Hefley et al. (2017a) used a generalized linear mixed model (GLMM) with random effects to understand individual-level and location-specific risk factors associated with chronic wasting disease in white-tailed deer from Wisconsin, USA. Spatial models are discussed further in Chap. 14.

Another extensive review that particularly focused on comparing the predictive performance of 33 species distribution models at species and community levels was given by Norberg et al. (2019). They focused on the following aspects: whether or not(a) they use parametric or semi-parametric; (b) they account for

interactions among environmental covariates when estimating species' responses to the environment; (c) they assess shared responses by species to the environment; (d) they explicitly include species co-occurrences not related directly to environmental variables; (e) they explicitly account for spatial structure; (f) they applied the statistical inference framework using shrinkage when estimating the response of each species to its environment; and (g) they used the statistical framework to account for parameter uncertainty when generating the predictions.

5.2.2 *Species and the Environment*

Species modeling also applies to relating distributions to environmental conditions and to detection and disease mapping. Species changes can track climate changes, especially in the ocean (Lenoir et al., 2020). Predation together with environment can also have an effect on distribution, as indicated by Puerta et al. (2015). They studied two species of cephalopod (octopus and squid) and their interaction with chlorophyll a (Chla) and sea surface temperature. They also looked at trophic (prey density) conditions, along with processes governing their distributions with contrasting life histories across the western Mediterranean Sea. Due to the high proportion of zeros in the observations of both species (~50%), they used delta generalized additive models (GAM) to estimate the cephalopod distributions. The Delta-GAM included two different submodels. First, presence-absence data are modeled using a binomial logit GAM and second where only positive density values are used in a Gaussian GAM with a log link function. Details are given in their paper.

5.2.3 *Model Building*

In the past, there has been a tendency to overlook imperfect detection, which we now know can have a major effect on estimation processes (Lahoz-Monfort et al., 2014), but is now incorporated into model building. Also, in the past, there has been a tendency to connect static observations of the occurrence (presence and absence) or abundance of a species with abiotic variables. However, both occurrence and abundance of a species can change over time and across space through colonization. This has led to a move toward dynamic models involving temporal and spatial changes, as well as allowing for joint dependencies among co-occurring species, as we shall see later below. The size of the scale (grain size) used for plots can affect some aspects of species distribution modeling (Guisan et al., 2007; Song et al., 2013).

Dynamic local interactions among species in a community can affect abundance and then scale up to affect distributions at broader spatial scale. Incorporating a time series of observations can improve species distribution models, and Bayesian

hierarchical models have also become popular along with new computing methods. Ordinal data are often used with species numbers such as 0 (absent), 1 (sparse), 2 (light), 3 (moderate), and 4 (heavy). Schliep et al. (2018) gave a helpful table of examples using ordinal data in ecology fitted using methods considered later, as well as a review of some aspects of species modeling.

5.2.4 Five Challenges

Araújo and Guisan (2006) mentioned five challenges for single-species distribution modeling, namely, (1) clarification of the niche concept, (2) improved designs for sampling data for building models, (3) improved parameterization, (4) improved model selection and predictor contribution, and (5) improved model evaluation. Their paper should be consulted for details. Another challenge is the presence of spatial autocorrelation, and methods for accounting for it are given in a review by Dormann et al. (2007) and considered further below. Veloz (2009) considered the effect of spatially autocorrelated sampling on measures of accuracy for presence-only niche models.

5.2.5 Bayesian Models

As only briefly referred to so far, Bayesian models have been used for species distributions, and they occur in the following sections. Latimer et al. (2006), beginning with a simple logistic regression with environmental covariates, added in spatial random effects and a conditional autoregressive (CAR) model. Using a hierarchical model, landscape alteration due to human influence can also be incorporated. All prior distributions for the parameters are vague or uninformative so that their posterior distributions are driven by the data. Computational methods are also described.

Gelfand et al. (2005) and Gelfand et al. (2006) developed an extensive hierarchical model with a focus on biodiversity. However, there is too much detail to reproduce the theory here. A multilevel model is used with modeling potential presence, transformed presence (both available and suitable) given potential absence, and observed presence/absence given suitability and availability. Broadly speaking, the method uses a binary process leading to the mapping of the model-estimated probability of presence for each species across the domain of interest and allows for the region having been partly transformed because of human involvement. Using a Bayesian framework, a two-stage, spatially explicit, hierarchical logistic regression using covariate information was developed. Along with the specification of all the posterior full conditional distributions of the unknown parameters in the model, computations can be carried out using Gibbs sampling.

Mattsson et al. (2013) applied a hierarchical Bayesian approach using point counts to model the relationships and interaction between bird species occupancy and six classes of local land cover variables, accounting for spatial autocorrelation, species similarities, and partial observability. They considered four hypotheses relating to spatial distribution (e.g., is it random?) and used a random effect for species.

Complications can arise in the modeling of a species distribution due to imperfect survey data such as observer error, gaps in the sampling, missing data, the spatial mobility of the species, and the fact that data may have been collected over long periods of time. There is also the problem of incorporating uncertainty from spatial climate variables into SDMs (Stoklosa et al., 2015).

Martin-Schwarze et al. (2018) gave a helpful review of three groups of various algorithms that are available for fitting a variety of models that can include temporal, spatial, and environmental variables. For example, an observation could be $y(\mathbf{s}, t)$, where \mathbf{s} are space coordinates, observed at n locations, and time t is measured at T time points. If y_{it} is the observation at location \mathbf{s}_i at time t ($t = 1, \dots, T$) with mean μ_{it} , then each μ_{it} can be linked to a structured additive predictor η_{it} through a link function $g(\cdot)$ so that we can model

$$\eta_{it} = g(\mu_{it}) = \alpha + \sum_{m=1}^M \beta_m x_{mit} + \sum_{k=1}^K f_k(z_{kit}) + u_{it}.$$

Here, the β coefficients quantify the linear effect of M covariates, f_k represents some latent model applied to covariates, and u_{it} represents one of the several possible spatiotemporal structures for the model, of which four are described. Also discussed were preferential (opportunistic) sampling, possible spatial misalignment (e.g., presence/absence data collected at points different from the spatial points where covariate data are available), nonstationarity, and the problem of zero inflation. They emphasized the use of the integrated nested Laplace approximation (INLA) methodology (Rue et al., 2009) and software (<http://www.r-inla.org>) as an alternative to Markov chain Monte Carlo (MCMC) methods, the main reason being the speed of calculation.

The authors noted that Bayesian inference proves to be a good option to deal with spatial hierarchical models because it allows both the observed data and model parameters to be random variables. As seen elsewhere in this book, Bayes' theorem is the tool that combines prior information with the likelihood to yield the posterior distributions. Also, SDMs are a very good example of a hierarchical structure that can be expressed as a hierarchical Bayesian model (Hefley & Hooten, 2016).

Martin-Schwarze et al. (2018) considered using Bayesian inference combined with N-mixture models to simultaneously account for both imperfect detection and spatial autocorrelation as follows. For site occupancy, let z_i be a random variable describing habitat suitability at site \mathbf{s}_i taking the values 1 or 0 (yes or no, respectively), depending on the habitat suitability, with π_i the probability that $z_i = 1$, and there are several visits at time $t = 1, \dots, T$. Let y_{it} be a random

variable representing the presence of a given species at site i and time t . The species is observed at site i ($\sum_t y_{it} \geq 1$) only if the habitat is suitable ($z_i = 1$). The species is unobserved at site i ($\sum_t y_{it} = 0$) if the habitat is not suitable ($z_i = 0$) or if the habitat is suitable ($z_i = 1$) but the probability α_{it} of detecting the species at site s_i and time t is lower than 1. Then, for the ecological process,

$$z_i \sim \text{Bernoulli}(\pi_i) \text{ and } \text{logit}(\pi_i) = \beta_0 + \sum_{m=1}^{M_1} \beta_m x_{mi}^{(1)} + w_i,$$

and for the detection process,

$$y_{it} \mid z_i \sim \text{Bernoulli}(z_i \alpha_{it}) \text{ and } \text{logit}(\alpha_{it}) = \gamma_0 + \sum_{m=1}^{M_2} \gamma_m x_{mit}^{(2)},$$

where we have parameters β_i and γ_i and x -covariates. The N-mixture model has a similar model structure. For the ecological process,

$$N_i \sim \text{Poisson}(\lambda_i) \text{ and } \log(\lambda_i) = \beta_0 + \sum_{m=1}^{M_1} \beta_m x_{mi}^{(1)} + w_i,$$

and for the detection process,

$$y_{it} \mid N_i \sim \text{Bernoulli}(N_i \alpha_{it}) \text{ and } \text{logit}(\alpha_{it}) = \gamma_0 + \sum_{m=1}^{M_2} \gamma_m x_{mit}^{(2)}.$$

The authors finished their article by referring to the problem of excess zeros, which can occur with mixture models (Sect. 4.15).

5.2.6 Community Species

We are aware of the ongoing loss of species throughout the world and the need for wide-scale studies (Winfree et al., 2015, who include information on some computer packages). There has been proliferation of species distribution models, and it is important to learn how well they perform at a community level. The species-area relationship, the species rank in occurrence, the local species richness, and the frequency of species combinations (all discussed below) are biodiversity indicators that can reveal important parts of the underlying ecosystem structure. Sollmann et al. (2016) developed a hierarchical community distance sampling model that estimates both community-level and species-level parameters related to detection

and abundance. An important application of community interactions (considered in Sect. 5.4) is that of dealing with an invasive species.

As already mentioned, if we are interested in more than one species, we can simply add another subscript in Chap. 4 to denote the species. For example, Dorazio (2007) applied the method to frog species using logistic regression with weather covariates (temperature, wind, and humidity) and added an extra subscript for year to observe the dynamics of the population. Another study involving several species was given by Dorazio and Royle (2005a) who developed likelihood functions and provided an estimate of the number of species. They used an alternative model based on a reparameterization of the model for estimating species richness and accumulation that allowed it to be fitted by Markov chain Monte Carlo (MCMC) sampling without any numerical integration (Dorazio & Royle, 2005b).

One approach for dealing with several species was given by Chen et al. (2013) who considered the effect of incomplete detection on a multispecies model, where the connection between species is done through covariates. It was based on data from the Biodiversity Monitoring project in which vascular plants are surveyed twice in the same year along a 2.5-km transect in 451 1-km² quadrats. They chose a random sample of 100 species from the 1700 species to determine general detection levels. Results from four life forms (LFs) grass, forb, shrub and tree were compared along with the species. Let z_{ij} be the latent occurrence state of species i ($i = 1, \dots, 100$) at quadrat j ($j = 1, \dots, 451$) such that $z_{ij} = 1$ denotes presence and $z_{ij} = 0$ denotes absence. If ψ_{ij} is the corresponding probability of presence and $y_{ijk} = 1$ with probability p_{ijk} if the i th species on quadrat j at survey k is observed given present, then we can assume that

$$y_{ijk} \mid z_{ij} \sim \text{Bernoulli}(z_{ij} p_{ijk}).$$

The authors then expressed $\text{logit}(\psi_{ij})$ and $\text{logit}(p_{ijk})$ as linear models using various covariates such as the LF of species, elevation, survey date, and their interactions. They used vague prior distributions for the regression parameters and applied Bayesian methods using WinBUGS with details in an appendix.

Sanderlin et al. (2014) used a Bayesian hierarchical model to consider design and cost trade-offs for community avian design monitoring programs. Using simulation, they considered optimizing a design for maximum accuracy of species richness, detection probability, and occupancy probability of rare species, given a fixed budget, and for a single season. It was assumed that all of the avian community species were part of the same super-community, an important assumption for data augmentation used later. Let $w_i = 1$ with probability Ω if species i ($i = 1, \dots, S$) was available for sampling during the season or $w_i = 0$ otherwise. Let $z_{ij} = 1$ if a species i is present at site j ($j = 1, \dots, J$) with probability ψ_{ij} , so that

$$z_{ij} \mid \psi_{ij}, w_i \sim \text{Bernoulli}(\psi_{ij} w_i),$$

and it is assumed that $\text{logit}(\psi_{ij}) = a_{i0} + a_{i1} h_j$, where h_j is a habitat covariate.

Let $x_{ijk} = 1$ if species i was detected at site j during sampling occasion k ($k = 1, \dots, K$), and let $y_{ij} = \sum_{k=1}^K x_{ijk}$ ($0 \leq y_{ij} \leq J$). We now have $y_{ij} | z_{ij}, P_i, J \sim \text{Binomial}(J, P_i z_{ij})$, where P_i is the probability of detecting species i , given present. Also, the authors assumed that $\text{logit}(P_i) = b_{i0}$ with covariate b_{i0} , a species-specific normal random effect. Similarly, detection parameter estimates for each species were random variables governed by a community-level parameter. Heterogeneity among species was modeled using a covariance term between species intercepts of occurrence (a_{i0}) and detection probability (b_{i0}). Stratified random sampling by habitat was used for sample site selection. The cost function used was $C = C_0 + C_1 S + C_2 JS$, where C_0 was the initial start-up cost, C_1 was the additional establishment cost per site, and C_2 was the additional cost to sample each site per sampling occasion. The aim was to minimize a root mean square error subject to constraints on it and on the total cost C . Further details of the simulation are given by the authors Sanderlin et al. (2014).

Golding et al. (2017) considered a dependent double-observer model for estimating multispecies abundance. The dependent double-observer method is a survey method that reduces false-positive observations using removal-based methodology to calculate detectability. Using two observers, the primary observer dictates all individuals observed during a survey to the secondary observer, who notes the identity and location of the individuals observed by the primary observer, as well as those missed by the primary observer.

5.3 Multivariate Species Models

5.3.1 Introduction

When there is more than one species, there can be interest in species interactions and cohabitation (e.g. Fidino et al., 2019). Species can vary significantly with respect to detection and habitat types (Ruiz-Gutiérrez & Zipkin, 2011). Clark et al. (2014) and Harris (2015) demonstrated that treating species separately and then combining (“stacking”) them might lead to large errors in forest and bird studies. Guisan and Rahbek (2011) also raised some criticisms in the context of presence/absence data. Zurell, Zimmermann et al. (2020b) compared the accuracy of assemblage predictions derived from stacked and JSMDs (joint species’ distribution models). With multispecies models, one species can provide information for predicting the distribution of other species.

Clark et al. (2017) expressed the view that species models need to involve multivariate distributions because of mutual dependence and interaction among species. They developed a generalized joint attribute model (GJAM) that readily applies to data that are combinations of presence-absence, ordinal, continuous, discrete, composition, zero-inflated, and censored data. For example, species groups not easily counted may be recorded in ordinal categories such as “rare,” “moderate,” and “abundant.” Presence-absence of a predator, pathogen, or mutualist might be

recorded, along with attributes such as body condition, infection status, and herbivore damage. The presence of other species and covariates involving environmental and other factors can play an important role in the mix of data.

The authors referred to two problems. One problem with such varied data is that transformed data, such as used in nonlinear link functions, is different from the original data. This means that covariances are then different from covariances between original species abundance. Another problem arises when abundance on many scales is replaced by presence-absence data or by changing the data in other ways. The authors modeled each observation by two vectors, \mathbf{x} , a vector of predictors, and \mathbf{y} , a vector of responses with the latter modeled in continuous and discrete spaces. Partitioning, judicious censoring, and attention given to zeros were also used with further details given by the authors.

Presenting another viewpoint, Guillera-Arroita (2017) discussed the very general problem of modeling communities in the presence of imperfect detection, pointing out some of the difficulties and asking the question of whether accounting for detection was really worth it, with examples from the literature.

5.3.2 *Various Multivariate Models*

Latimer et al. (2009) looked at Bayesian hierarchical modeling for a JSDM (joint species distribution model) using large spatial point data sets. Their aim was to simultaneously model the distributions of a set of four major invasive species in a spatially explicit multivariate model. D'Amen et al. (2017), in their general descriptive review, briefly described the concepts and theories that span the different drivers of species assembly and then summarized some of the existing methods. In their extensive review, Norberg et al. (2019) commented that “methodological advances in multiple-species distribution modeling have lagged behind, but are recently experiencing a rapid expansion.” We now consider some of these joint species distribution models.

Warton et al. (2015) provided a useful review of some of the joint species methods that are all extensions of the generalized linear model (GLM) and also gave information on computer packages that are available. They noted that there is a need to include multivariate random effects in a mixed model to capture correlation in abundance across taxa and this leads to multivariate generalized linear mixed models (GLMMs). Such a model is especially useful when the number of taxa is small compared to the number of samples and can be fitted using standard mixed modeling software. It can be usefully combined with a latent variable model (LVM), which introduces some unobserved (“latent”) predictors to each sample. The latent variables induce correlation between taxa, and their number controls model complexity, such that it is possible to fit joint models across many taxa. They outlined the following models.

For the generalized linear mixed model (GLMM), let y_{ij} be the abundance for sample species i ($i = 1, \dots, I$) and sample j ($j = 1, \dots, J$), with expected value

μ_{ij} , which can be modeled as

$$g(\mu_{ij}) = \beta_{0i} + \alpha_j + \boldsymbol{\beta}'_i \mathbf{x}_j + u_{ij}.$$

Here, $g(\cdot)$ is a link function, \mathbf{x}_j are covariates (predictors), and u_{ij} are random effects that can be modeled as a multivariate normal distribution. If we define $\mathbf{u}_j = (u_{1j}, \dots, u_{Ij})'$, we can then model

$$y_{ij} \mid \mathbf{u}_j \sim f(\mu_{ij}, \boldsymbol{\phi}_i), \text{ and } \mathbf{u}_j \sim N(\mathbf{0}, \boldsymbol{\Sigma}).$$

Here, $f(\cdot)$ is some distribution, which is typically normal for continuous data, but binomial for presence/absence data, and Poisson or negative binomial for counts, with some parameter $\boldsymbol{\phi}_i$. The variance-covariance matrix of random effects $\boldsymbol{\Sigma}$ controls the correlation between taxa, though its specification can be problematic if the number of taxa J is large.

A flexible way to incorporate correlation is to use a latent variable model (LVM) model, which is a function of unmeasured predictors (or “latent variables”), \mathbf{w}_j , as well as measured predictors so that we now have

$$g(\mu_{ij}) = \beta_{0i} + \alpha_j + \boldsymbol{\beta}'_i \mathbf{x}_j + u_{ij}, \text{ and } u_{ij} = \boldsymbol{\lambda}'_i \mathbf{w}_j.$$

Distributions are the same as before except that $\boldsymbol{\Sigma} = \boldsymbol{\Lambda} \boldsymbol{\Lambda}'$, where $\boldsymbol{\Lambda}$ is the full matrix of factor loadings, with the $\boldsymbol{\lambda}_i$ as its columns. We then assume

$$y_{ij} \mid \mathbf{w}_j \sim f(\mu_{ij}, \boldsymbol{\phi}_i), \text{ and } \mathbf{w}_j \sim N(\mathbf{0}, \mathbf{I}).$$

The authors discussed the pros and cons of the two models and some other strategies. If incomplete detection needs to be added to the above model, which is usually the case, Beissinger et al. (2016) pointed out that an extra layer needed to be added to accommodate imperfect detection using measurements derived from repeated surveys of species over a period when the population is closed. For example, if y_{ijk} is the number of species i detected at site j for replicate k , then $y_{ijk} \sim \text{Binomial}(N_{ij}, p_{ijk})$, where p_{ijk} is the probability of being detected and N_{ij} , the true abundance of species i at site j , is now a latent variable with individuals having a probability of being in the latent state “presence.”

Harris (2015) commented that previous species distribution models tended to assume either (i) that that species’ occurrence probabilities are uncorrelated or (ii) that species respond linearly to preselected environmental variables. He introduced the stochastic feedforward neural network, called *MistNet*, which makes neither assumption and can account for non-independent co-occurrence patterns driven by unobserved environmental heterogeneity. It can also learn nonlinear functions relating species’ occurrence probabilities to environmental predictors.

Ovaskainen et al. (2010) introduced a multivariate logistic regression to model species co-occurrence. This was followed by Ovaskainen et al. (2016a) and

Ovaskainen et al. (2016b), who presented a hierarchical joint species distribution modeling framework that attributes variation in species occurrence and co-occurrence to the influences of environmental variables and species-to-species associations. Here, $z_{ij} = 1$ if species i is detected at site j , and z_{ij} is expressed as a probit regression model with covariates and random effects. They followed Ovaskainen and Soininen (2011) in borrowing information across species by using multivariate normal distributions for the random effects and the covariates in a Bayesian framework. Pollock et al. (2014) used a similar method based on probit regression.

Pacifici et al. (2017) considered the problem of using both standardized and non-standardized (lesser-quality) data type, which endeavored to offer a potential solution to the trade-off between data quality and quantity. They extended previous work in this subject by allowing for explicit spatial autocorrelation in occurrence and also detection error using a multivariate conditional autoregressive (MVCAR) model. Starting with a standard “single” data source, they developed were three models that integrate a second data source, usually of lesser quality in a less direct manner resulting in more robust performance when the auxiliary data is of lesser quality, namely, “shared,” “correlation,” and “covariates.” The shared model performed the best, but the other two also performed well.

Ovaskainen et al. (2017b) proposed hierarchical modeling of species communities (HMSC) as a general, flexible framework for modern analysis of community data and referred to recent advances in joint species distribution models. Using their notation, they proposed four kinds of data matrices: \mathbf{Y} , typical community data that includes observations on the occurrence of species in a set of temporal and/or spatial replicates, referred to as “occurrence” data; \mathbf{X} , environmental data consisting of a set of measured covariates; \mathbf{T} , data on species-specific traits such as morphological traits (e.g., body size) and physiological traits (e.g., tolerance to salinity); and \mathbf{C} , functional traits (e.g., such as feeding type or to the actual position of the species within the surrounding food web). We have $\mathbf{Y} = (y_{ij})$, where y_{ij} refers to occurrence data (presence-absence, count, or biomass) for species j at site i , where y_{ij} has some distribution $f(L_{ij}, \sigma_j^2)$, and where L_{ij} is a linear predictor of a generalized linear model (e.g., log-linear, logistic, or probit). Here, f depends on the type of data and σ_j^2 , and $L_{ij} = L_{ij}^F + L_{ij}^R$ with fixed (F) and random (R) parts, where

$$L_{ij}^F = \sum_k \beta_{ik} x_{jk},$$

and $(x_{jk}) = \mathbf{X}$. If $\boldsymbol{\beta}_j = (\beta_1, \beta_2, \dots)'$, then it is assumed that the $\boldsymbol{\beta}_j$ have multivariate normal distributions. The authors gave just a broad brush approach without specific details.

Tikhonov et al. (2020) extended HMSC by exploiting two spatial statistics techniques that facilitate the analysis of large spatial data sets, namely, the Gaussian predictive process and the nearest neighbor Gaussian process. They also devised an efficient Gibbs posterior sampling algorithm for Bayesian model fitting that allows

one to analyze community data sets consisting of hundreds of species sampled from up to hundreds of thousands of spatial units.

Hierarchical multispecies occupancy models (MSOMs) are also described in Sect. 5.5.12 where they are indirectly linked to biodiversity, along with multispecies abundance models (MSAMs) for estimating occupancy and abundance.

Ordinal Data

Schliep et al. (2018) introduced a multivariate generalized linear model with probit link function (as described by Pollock et al., 2014) as a “baseline” joint species distribution model for ordinal data and then added in more layers of structure. Let $y_{jt}^{(i)} \in \{0, 1, \dots, L - 1\}$ denote the observable ordinal abundance for species i ($i = 1, \dots, m$) at plot j ($j = 1, \dots, J$) and time t ($t = 1, \dots, T$). Here, L is the number of ordinal categories, with $L = 2$ for binary occurrence data because binary data are a special case of ordinal data with only two categories (presence and absence). The relationship between $y_{jt}^{(i)}$ and the underlying (latent) continuous abundance $z_{jt}^{(i)}$ is given by

$$y_{jt}^{(i)} = \begin{cases} 0, & z_{jt}^{(i)} < \lambda_1^{(i)}, \\ 1, & \lambda_1^{(i)} \leq z_{jt}^{(i)} < \lambda_2^{(i)}, \\ \vdots & \\ L - 1, & \lambda_{L-1}^{(i)} \leq z_{jt}^{(i)}. \end{cases}$$

Let $\boldsymbol{\lambda}^{(i)} = (\lambda_1^{(i)}, \dots, \lambda_{L-1}^{(i)})'$, the vector of threshold parameters with $\lambda_1^{(i)} = 0$, and $\lambda_\ell^{(i)} \leq \lambda_{\ell+1}^{(i)}$ for $\ell = 1, \dots, L - 2$ and all i . It is assumed that L is the same for all species, though this can be modified. Let $\mathbf{z}_{jt} = (z_{jt}^{(1)}, \dots, z_{jt}^{(m)})'$ and let $\mathbf{k}_{jt} = (k_{jt}^{(1)}, \dots, k_{jt}^{(m)})'$, where $k_{jt}^{(i)}$ is the latent process of abundance for plot j at time t and species i . It is then assumed that the elements of \mathbf{z}_{jt} are conditionally independent given \mathbf{k}_{jt} and \mathbf{z}_{jt} has a multivariate normal distribution $N(\mathbf{k}_{jt}, \mathbf{I}_m)$, where \mathbf{I}_m is an $m \times m$ identity matrix for parameter identifiability under the probit model. Using climatic and abiotic plot-specific covariates, the authors used the model

$$\mathbf{k}_{jt} = \mathbf{B}\mathbf{x}_{jt} + \boldsymbol{\eta}_{jt},$$

where \mathbf{B} is an $m \times p$ matrix of species-specific coefficients describing the species-environment relationship and \mathbf{x}_{jt} is a vector containing an intercept and $p - 1$ time- and plot-specific covariates. The spatial random effects, $\boldsymbol{\eta}_{jt}$, are modeled using a linear model of coregionalization (Gelfand et al., 2004), which the authors described mathematically.

A temporal effect can be added using the model

$$\mathbf{k}_{jt} = \boldsymbol{\alpha}_t + \mathbf{B}\mathbf{x}_{jt} + \boldsymbol{\eta}_{jt},$$

where the elements of α_t denote the species-specific random effects for time t . Some of the parameters are set equal to zero to achieve identifiability. The next addition can be

$$\mathbf{k}_{jt} = \alpha_t + \mathbf{Bx}_{jt} + \mathbf{Rk}_{j,t-1} + \eta_{jt},$$

with the addition of an autoregressive term consisting of an $m \times m$ matrix $\mathbf{R} = (\rho_{ab})$ that allows inter- and intraspecific processes to affect abundance. The final addition is the inclusion of replicated observations within a location.

The authors applied the above methods to modeling the abundance for two invasive insects, hemlock woolly adelgid (*Adelges tsugae*) and elongate hemlock scale (*Fiorinia externa*), that share the host tree eastern hemlock (*Tsuga canadensis*) in the Eastern USA. Details regarding the prior distributions and the Metropolis-within-Gibbs Markov chain Monte Carlo sampling algorithm are given in their Supporting Information, Appendices S1 and S2. The **R** function *Multivariate.Ordinal.Spatial.Model* was used to fit the model and extensions.

5.3.3 Occupancy Models

We now discuss a selection of occupancy models for several interacting species, which have become more popular, especially with supporting covariates. A two-species model was developed by MacKenzie et al. (2004) to investigate species co-occurrence patterns. Changing the notation from that above, they considered site i and two species A and B with 1's and 0's for those observed and not observed, respectively. There are four possible states with probabilities: (1) occupied by both species A and B ($\psi_i^{(AB)}$), (2) occupied by species A ($\psi_i^{(A)}$) and only by species A ($\psi_i^{(A)} - \psi_i^{(AB)}$), (3) occupied by species B ($\psi_i^{(B)}$) and only by species B ($\psi_i^{(B)} - \psi_i^{(AB)}$), or (4) occupied by neither species ($1 - \psi_i^{(A)} - \psi_i^{(B)} + \psi_i^{(AB)}$). The probability of location i being in each of the four states is given by the vector

$$\phi_i = (\psi_i^{(AB)}, \psi_i^{(A)} - \psi_i^{(AB)}, \psi_i^{(B)} - \psi_i^{(AB)}, 1 - \psi_i^{(A)} - \psi_i^{(B)} + \psi_i^{(AB)})',$$

with elements summing to 1. We now list the various probabilities of detection as follows:

$$p_{ij}^{(A)} = \text{Pr}(\text{detection for species } A, \text{ given only } A \text{ present})$$

$$p_{ij}^{(B)} = \text{Pr}(\text{detection for species } B, \text{ given only } B \text{ present})$$

$$r_{ij}^{(AB)} = \text{Pr}(\text{detecting both species, given both present})$$

$$r_{ij}^{(Ab)} = \text{Pr}(\text{detection for species } A \text{ but not } B, \text{ given both present})$$

$$\begin{aligned}
r_{ij}^{(ab)} &= \Pr(\text{detection for species } B \text{ but not } A, \text{ given both present}) \\
r_{ij}^{(ab)} &= \Pr(\text{detection for neither species, given both present}) \\
&= 1 - r_{ij}^{(AB)} - r_{ij}^{(Ab)} - r_{ij}^{(aB)}
\end{aligned}$$

The sequence of detections and non-detections at a site for each species may be recorded as history with 1 for detection and 0 for non-detection. For example, for three surveys, we have as an example for site i the history $x_i^{(A)} = 101$ (species A detected at only the first and third occasions) and $x_i^{(B)} = 000$ (species B not detected at all). Then given both events, we have as examples

$$\begin{aligned}
\Pr(x_i^{(A)} = 101, x_i^{(B)} = 000 \mid \text{both present}) &= r_{i1}^{(Ab)} r_{i2}^{(ab)} r_{i3}^{(Ab)} \\
\Pr(x_i^{(A)} = 101, x_i^{(B)} = 000 \mid \text{only } A \text{ present}) &= p_{i1}^{(A)} (1 - p_{i2}^{(A)}) p_{i3}^{(A)}.
\end{aligned}$$

The probability of observing this combination of histories for all other occupancy states (occupied by species B only and occupied by neither species) is 0, as both states prohibit species A from being at the site, yet species A was actually observed there. We can therefore define a vector $\mathbf{p}_i(x_i^{(A)}, x_i^{(B)})$ representing the probability of observing the detection histories conditional upon each state. For instance, using the above example,

$$\mathbf{p}_i(101, 000) = \begin{pmatrix} r_{i1}^{(Ab)} r_{i2}^{(ab)} r_{i3}^{(Ab)} \\ p_{i1}^{(A)} (1 - p_{i2}^{(A)}) p_{i3}^{(A)} \\ 0 \\ 0 \end{pmatrix}$$

The unconditional probability of observing the two histories is

$$\Pr(x_i^{(A)}, x_i^{(B)}) = \boldsymbol{\phi}'_i \mathbf{p}_i(x_i^{(A)}, x_i^{(B)}).$$

Given S sites, the likelihood function is

$$L = \prod_{i=1}^S \Pr(x_i^{(A)}, x_i^{(B)}),$$

which can be maximized numerically. The authors gave several examples of other detection histories. Examples of hypotheses that can be tested were given such as $\psi^{(AB)} = \psi^{(A)}\psi^{(B)}$ for species independence. The authors introduced a species

interaction factor (SIF) γ that is the ratio of how likely the two species are to co-occur compared to what would be expected under a hypothesis of independence, namely,

$$\gamma = \frac{\psi^{(AB)}}{\psi^{(A)}\psi^{(B)}}.$$

Here, $\gamma = 1$ with complete independence, whereas less than 1 suggests species avoidance, and greater than 1 suggests contagion, or a tendency to co-occur more frequently than expected under independence. The reader is referred to their paper for further details of the likelihood function, various models considered, and computations. They suggested measuring the degree of independence by γ and noted that there were some problems in expressing the estimate of γ in terms of covariates.

Bailey et al. (2009) used a similar approach to investigate the expansion of the barred owls into the territory of the northern spotted owl to determine the interaction between the two species and possible changes in vocalization behavior and detection probabilities of the spotted owl. Rossman et al. (2016) also studied barred owls using dynamic N-occupancy models.

Richmond et al. (2010) introduced a new parameterization for the two-species occupancy model considered above called the “conditional two-species occupancy model” that successfully incorporated covariates. They applied it to two species of secretive wetland birds called rails to assess their interaction, where one species, say A , dominates the other species B . The method estimated the probability of occupancy for a subordinate species conditional upon the presence of a dominant species (e.g., prey-predator relationship) using a different parameterization that allows for a better incorporation of covariates. As before, $\psi_i^{(A)}$ is the probability that species A is present. They also considered $\psi_i^{(B|A)}$, the probability of presence for species B conditional on species A being present, and $\psi_i^{(B|a)}$ the probability of presence for species B conditional on species A being absent. The vector ϕ_i now becomes

$$\phi_i = (\psi_i^{(A)}\psi_i^{(B|A)}, \psi_i^{(A)}(1 - \psi_i^{(B|A)}), (1 - \psi_i^{(A)})\psi_i^{(B|a)}, (1 - \psi_i^{(A)})(1 - \psi_i^{(B|a)}))'$$

and γ , the species interaction factor, becomes

$$\phi = \frac{\psi^{(A)}\psi^{(B|A)}}{\psi^{(A)}(\psi^{(A)}\psi^{(B|A)} + (1 - \psi^{(A)})\psi^{(B|a)})}.$$

Again, independence occurs if $\phi = 1$ as before. The paper proceeds in a similar fashion to MacKenzie and Bailey (2004).

Waddle et al. (2010) considered the same problem using $\psi^{(B)}$, $\psi^{(A|B)}$, and $\psi^{(A|b)}$, where b denotes B absent, as above, where site number i is suppressed. Let $z^{(A)}$ and $z^{(B)}$ be the usual indicators of presence with Bernoulli distributions taking

values 0 or 1. Now, the joint probability distribution takes the form $\Pr(z^{(A)}, z^{(B)}) = \Pr(z^{(A)}|z^{(B)})\Pr(z^{(B)})$, so that

$$\begin{aligned} z^{(B)} | \psi^{(B)} &\sim \text{Bernoulli}(\psi^{(B)}) \\ z^{(A)} | z^{(B)}, \psi^{(A|B)}, \psi^{(A|b)} &\sim \text{Bernoulli}\left(z^{(B)}\psi^{(A|B)} + (1 - z^{(B)})\psi^{(A|b)}\right). \end{aligned}$$

If species A does not depend on whether species B is present or absent, then $\psi^{(A|B)} = \psi^{(A|b)}$, which can be tested.

We now bring in site i and replicate t ($t = 1, \dots, T$). Let $\mathbf{y}_i^{(C)} = (y_{i1}^{(C)}, \dots, y_{iT}^{(C)})'$, where $y_{it}^{(C)} = 1$ if an individual of species C ($C = A$ or B) is observed at site i and replicate t , so that $\mathbf{y}_i^{(C)}$ is a sequence of 1's and 0's. Let

$$\begin{aligned} p_{it}^{(A|B)} &= \Pr(y_{it}^{(A)} = 1 | z_i^{(A)} = 1, z_i^{(B)} = 1) \\ p_{it}^{(A|b)} &= \Pr(y_{it}^{(A)} = 1 | z_i^{(A)} = 1, z_i^{(B)} = 0). \end{aligned}$$

We have that

$$y_{it}^{(A)} | z_i^{(A)}, z_i^{(B)}, p_{it}^{(A|B)}, p_{it}^{(A|b)} \sim \text{Bernoulli}(z_i^{(A)}\{z_i^{(B)}p_{it}^{(A|B)} + (1 - z_i^{(B)})p_{it}^{(A|b)}\})$$

with logistic regressions for $\psi_i^{(A)}$, with covariates the presence of B , temperature, and habitat; for $\psi_i^{(B)}$, with habitat and longitude covariates; and for $p_{ij}^{(A)}$, with covariates the presence of B and temperature. Maximum likelihood estimates were obtained for tree frog species in Southern Florida, USA.

Rota et al. (2016) introduced a multispecies single-season occupancy model for two or more interacting species. Here, a set of n sites is randomly selected from a population of interest, and they modeled detections ($y_{ist} = 1$) and non-detections ($y_{ist} = 0$) of species s ($s = 1, \dots, S$) at site i during survey t ($t = 1, \dots, J_i$), conditional on the presence of species s ($z_{si} = 1$) so that

$$y_{sit} | z_{si} \sim \text{Bernoulli}(z_{si} p_{sit}),$$

where p_{sit} is the probability of detection given present, modeled using covariates from $\text{logit}(p_{sit}) = \mathbf{w}'_{sit} \boldsymbol{\alpha}_s$. If $\mathbf{z}_i = (z_{1i}, \dots, z_{Si})'$, the latent occupancy state of S species at site i is modeled as a multivariate Bernoulli random variable (cf. (A.3)), a multinomial distribution with index 1, namely,

$$\mathbf{z}_i \sim \text{MVB}(\psi_i),$$

where \mathbf{z}_i is an S -dimensional vector of 1's and 0's denoting the latent occupancy state of all S species. Here, ψ_i is a 2^S -dimensional vector denoting the probabilities

of all S sequences of 1's and 0's \mathbf{z}_i can attain such that $\sum_{a=1}^{2^S} \psi_{ai} = 1$. For two species, $S = 2$ and

$$\mathbf{z}_i \sim \text{Bernoulli}(\psi_{11}, \psi_{10}, \psi_{01}, \psi_{00}),$$

with probability function

$$f(\mathbf{z}_i \mid \psi_{11}, \psi_{10}, \psi_{01}, \psi_{00}) = \psi_{11}^{z_1 z_2} \psi_{10}^{z_1(1-z_2)} \psi_{01}^{(1-z_1)z_2} \psi_{00}^{(1-z_1)(1-z_2)}.$$

Then, for a log-linear model, the log of this function is

$$\begin{aligned} & \log(\psi_{00}) + z_1 \log\left(\frac{\psi_{10}}{\psi_{00}}\right) + z_2 \log\left(\frac{\psi_{01}}{\psi_{00}}\right) + z_1 z_2 \log\left(\frac{\psi_{11} \psi_{00}}{\psi_{01} \psi_{10}}\right) \\ &= \log(\psi_{00}) + f_1 + f_2 + f_{12}. \end{aligned}$$

The following quantities can then be obtained:

- (1) The probability of occurrence for species 1 is $\Pr(z = 1) = \psi_{11} + \psi_{10}$.
- (2) The covariance between species 1 and 2 is $\text{cov}[z_1, z_2] = \psi_{11}\psi_{00} - \psi_{10}\psi_{01}$.
- (3) The probability of occurrence for species 1, conditional on the presence or absence of species 2, is $\Pr(z_i = 1 \mid z_2) = \psi_{1z_2}/(\psi_{1z_2} + \psi_{0z_2})$.

The ψ_{rs} 's are expressed as exponential functions of the f 's, and the f 's are then expressed in terms of linear functions of covariates. Summing up, we see that the multispecies occupancy model developed above assumed detection and occurrence probabilities are modeled as a function of covariates via a multinomial logit link function. Alternatively, this model can be specified as multinomial probit regression, which allows the exploitation of data augmentation techniques and Gibbs sampling.

Rota et al. (2016) proposed a general two-species model to account for both interacting species and spatial dependence. They assumed that detection ($y_{ijk} = 1$) or non-detection ($y_{ijk} = 0$) is recorded for species i at site j and survey k . Also, if $z_{ij} = 1$ if species i occupies site j and 0 otherwise, then it is assumed that

$$y_{ijk} \mid z_{ij} \sim \text{Bernoulli}(z_{ij} \Phi(\mathbf{w}'_{jk} \boldsymbol{\alpha}_i)),$$

where $\Phi(\cdot)$ represents the cumulative distribution function of a standard normal distribution (i.e., a probit link), $\boldsymbol{\alpha}_i$ is a d -dimensional regression parameter, and \mathbf{w}_{jk} is a d -dimensional vector of covariates. It is also assumed that

$$z_{ij} \sim \text{Bernoulli}(\Phi(\mathbf{x}'_j \boldsymbol{\beta}_i + \eta_{ij})),$$

where the η_{ij} are spatial random effects modeled using kriging and multivariate Gaussian distributions. Bayesian methods are used, and the paper is referred to for details. Rota et al. (2016) extended the method to two or more species using a multivariate Bernoulli distribution.

Kellner et al. (2022) modeled species interaction extending the model by Rota et al. (2016) above using continuous time with 22 different parameters for a two-species model and a Markov-modulated Poisson process for detection. Four occupancy states were used for the two species, (10), (01), (11), and (00), where 1 denotes present and 0 absence.

5.3.4 Comparing Models

Four different joint species' distribution models were defined by Wilkinson et al. (2012) that are used to predict unsampled areas or new environmental conditions. The models corresponded to different aspects of species distribution and community assemblage processes (with general probabilities in brackets and species A , B , and C). The four models are marginal predictions that are environment-only predictions akin to predictions from single-species models ($\text{Pr}(A)$), joint predictions that simultaneously predict entire community assemblages ($\text{Pr}(A \cap B \cap C')$), conditional marginal ($\text{Pr}(A|B)$), and conditional joint ($\text{Pr}(A \cap B|C')$) predictions that are made at the species or assemblage level, conditional on the known occurrence state of one or more species at a site, where C' is the complement of event C . They referred to the core model (5.1) above. The marginal model simply treats the species individually, the joint species model accounts for both the environmental covariates and species co-occurrence, and the other two use known information from other species. They discussed the use of various measures (metrics) for evaluating a model.

Wilkinson et al. (2019) compared seven JSDMs that incorporate species presence-absence data. They began with the following core model. Let $z_{sj} = 1$ if the s th species at site j is present and 0 otherwise ($s = 1, \dots, S$, $j = 1, \dots, J$), and let $y_{sj} = 1$ if species s is observed at site j . It is assumed that

$$\begin{aligned} z_{sj} &= \mu_{is} + e_{sj} \\ \mu_{sj} &= \beta_j' \mathbf{x}_s \\ \mathbf{e}_j &\sim N(\mathbf{0}, \mathbf{R}), \end{aligned} \tag{5.1}$$

where the \mathbf{x}_j are environmental covariates and \mathbf{R} is the $J \times J$ correlation matrix that captures the dependencies in co-occurrence among species after accounting for covariates in the model. Its diagonal elements are unity.

The seven models were:

- (1) MPR: Multivariate probit regression (Golding et al., 2015)
- (2) MLR: Multivariate logistic regression (Ovaskainen et al., 2010)
- (3) HPR: Hierarchical multivariate probit regression (Pollock et al., 2014)
- (4) LPR: Multivariate probit regression with latent factors (Hui, 2016)
- (5) DPR: Multivariate generalized regression model with dimension reduction (Clark et al., 2017)

- (6) Two versions of a hierarchical multivariate probit regression with latent factors (Ovaskainen et al., 2016b) consisting of spatially explicit (HLR-S) and non-spatially explicit (HLR-NS) versions

Using a desktop computer, the authors found that only four of the above models ran to completion within the time limit set for all the data sets, namely, MPR, LPR, DPR, and HLR-NS. Also, MPR and HLR-NS were markedly faster than the others. The authors concluded that the core statistical differences between the different models can be broadly defined by two methodological choices: “the inclusion of a hierarchical structure on the regression coefficients, and the use of latent factors to account for shared species responses to hypothetical, unmeasured environmental variables impacting species co-occurrence patterns.”

5.3.5 Spatial Models

Thorson et al. (2015) emphasized the importance of spatial models for estimating the strength of density dependence. Thorson et al. (2016) combined a dynamic model with a joint species distribution, using geostatistical methods, to include spatial autocorrelation for spatial similarity when estimating one or more “factors” (cf. Thorson et al., 2015). This led to a new spatial dynamic factor analysis model using species ordination.

Thorson and Barnett (2017) compared multispecies models with similar single-species models using a vector autoregressive spatiotemporal model that simultaneously estimates spatiotemporal variation in density for multiple species. Using three fish groups that have similar spatial distributions, they found that “The multispecies model was more parsimonious and had better predictive performance than fitting a single-species model to each taxon individually,...the single-species models showed similar abundance trends and distribution shifts to those of the multispecies model, with slightly smaller standard errors.”

Spatial autoregression can be incorporated into a logistic regression for the probability of presence (occupancy) using a spatial covariate as in Webb et al. (2014), for example. Miller et al. (2019) considered data integration methods to estimate species’ distributions.

5.3.6 Computing

Considering models available as R packages, Zhang et al. (2018) systematically evaluated the performances of five JSDMs in predicting the occurrence and biomass of multiple species. They particularly focused on diverse characteristics of sampling data, including the type of response variables, the number of sampling sites, and the number of species included in models. The five models were “BayesComm”

(Golding & Harris, 2015), “MistNet” (Harris, 2015), “Boral” (Hui, 2016), “HMSC” (Ovaskainen et al., 2017b), and “Gjam” (Clark et al., 2017).

Hui (2016) developed an **R** computer package called **BORAL** available on CRAN for analyzing multivariate abundance data using Bayesian Markov chain Monte Carlo methods and the **JAGS** program. A key feature of the package is the ability to incorporate latent variables as a parsimonious method of modeling between species correlation.

5.3.7 *Species Correlations*

Wagner et al. (2020) mentioned that a primary goal in fisheries science is to predict how species distributions and abundances respond to current and future habitat conditions and to use those predictions to inform fisheries management. In reviewing some of the recent literature, they said that in modeling for species interactions in freshwater fisheries, the main focus in the past has been on isolating the effects of two or three species in the community. For example, two-species occupancy models have been used to elucidate how the presence of one species modulates the response of another species to watershed stressors, and the presence of one species has been used to predict the occurrence of another focal species.

The authors reminded us that a JSMD (joint species distribution model) relies on a species correlation matrix to account for possible dependencies among species (i.e., potential biotic interactions) at a given location that are not accounted for by other fixed effects (environmental predictors). Also, JSMDs are particularly useful for the analyses of fish community data, as fisheries data are often collected across space and over time and thereby may exhibit spatial and (or) temporal autocorrelation. They used the notation of Wilkinson et al. (2019) described above in Eq. (5.1) in Sect. 5.3.4.

The correlation matrix **R** hopefully identifies those species that co-occur more or less frequently than predicted, given the environmental covariates. Residual dependencies also allow for predicting the occurrence of each species conditionally on the occurrence of other species in the community. All covariates were standardized to have mean zero with standard deviation one, and default priors specified in the Gjam package mentioned above were used for all parameters. The odds ratios for species i and i' at site j were calculated. If the ratio was greater than 1, there is positive dependence between species, and if the ratio was less than 1, there is negative dependence. Computational details and applications are given by Wagner et al. (2020).

With regard to correlation matrices, Gelfand and Shirota (2019) commented that JSMDs attempt to capture species dependence through an associated correlation matrix arising from a set of latent multivariate normal variables, but these associations offer little insight into dependence behavior between species at sites. They focused on presence/absence data using joint species modeling, which incorporates spatial dependence between sites. For pairs of species, they emphasized that the

induced odds ratios (along with the joint probabilities of occurrence) provided much clearer understanding of joint presence/absence behavior. Details are given in their lengthy paper.

5.4 Species Interactions

Although multivariate species distributions have been discussed above, it is of particular interest to look further at how much species interact with one another, as interactions can have a profound effect on species distributions and species diversity. Understanding these interactions is helpful in predicting how communities will respond to climate change or habitat modification, but they can be tricky to interpret. For example, the negative correlation between two competing species can be reversed in the presence of a third species that outcompetes both of them. Urbanization can also affect how species interact (Lewis et al., 2015: bobcats and pumas). Care is therefore needed in interpreting biotic interactions, and Dormann et al. (2018) listed ten questions that should be asked. An important aspect of modeling interactions is how standard multivariate autoregressive (MAR) time-series models determine population dynamics (Ovaskainen et al., 2017a).

Dorazio and Connor (2014) gave an earlier summary of some methods of investigating species interaction and described some of the difficulties involved. Often incomplete detectability is ignored. They were interested in whether the effects of environmental (habitat) covariates or interactions between species were more influential in determining the abundances of individual species. They introduced a model to estimate the abundances of potentially interacting species encountered while conducting point-count surveys at a set of ecologically relevant locations, as in a metacommunity of species, for example. In the model, which focused on forest birds during the breeding season, it was assumed that abundances of species with similar traits (e.g., body size) are potentially correlated and that these correlations, when present, may exist among all species or only among functionally related species (such as members of the same foraging guild). When it comes to counting several species at the same time, technology can be helpful. For example, Acevedo et al. (2009) compared various methods of machine learning for the automated classification of bird and amphibian calls.

5.4.1 Two Interactive Species

Popovic et al. (2019) mentioned three important drivers of co-occurrence: (1) two species may co-occur because they both respond to the same (or similar) environmental variables; (2) two species may both respond to the presence or abundance of another species (a “mediator species”), for example, they may both be hunted by the same predator but have no direct association between them; and (3)

two species might have a direct association, such as facilitation, seed dispersal, or pollination. They can also compete for food and habitat. It is therefore important to consider conditional dependence relationships, which describes how pairs of species are related, after controlling for all the other species in the data set.

Popovic et al. (2018) demonstrated the use of Gaussian copula graphical models (GCGMs) to uncover conditional relationships among species from abundance data and hence untangle the impact of mediator species on the co-occurrence patterns between pairs of species. Copula models are so named because they couple a multivariate distribution (e.g., multivariate Gaussian) with any set of marginal distributions (e.g., Poisson, binomial, negative binomial, and multinomial distribution for ordinal data). Their simulations showed that GCGMs perform as well or better than competing methods in most scenarios, especially when the data are ordinal or biomass or when there is known environmental heterogeneity.

Zurell et al. (2018) proposed to test the capacity of joint species distribution models (JSDMs) to detect species interactions from co-occurrence data, at different scales of data aggregation. They used a point process model to simulate the co-occurrence of species pairs at steady state. They cautioned against interpreting residual correlations from JSDMs as interaction strength and against comparing these across different species and communities.

Thurman et al. (2019) endeavored to test the link between species interactions and species co-occurrence in a trophic network using what was effectively known as networks for pond-breeding amphibian species. They compared the accuracy of four statistical methods in inferring these known species interactions from spatial associations. Using the best method, they found that co-occurrence methods are generally inaccurate when estimating trophic interactions. The strength and sign of inferred interactions were dependent upon the spatial scale of observation, and predator presence influenced the detectability of competitive interactions among prey species.

Tobler et al. (2019) included both measurement errors and species correlation in their multivariate model. They developed two presence-absence JSDMs for modeling pairwise species correlations while accommodating imperfect detection: one using a latent variable and the other using a multivariate probit approach. A slight adaption of the usual multivariate model was used. One of the difficulties in dealing with such models is controlling the number of variables and parameters.

A related topic is disease control and coinfection patterns, and it is particularly important with invasive species to preempt their arrival on the scene. Singh et al. (2017) discussed predicting which species will become invasive in each country or region before they arrive in order to devise and implement measures for minimizing the costs of biological invasions. They tabled ten criteria for evaluating species biosecurity risks and focused on an assessment procedure that allowed many species to be assessed relatively quickly with readily available information. This led to providing risk scores that will be valuable to prioritize which species warrant more detailed assessments.

Azaele et al. (2010) used the package **MAXENT** to quantify species effective couplings using a pairwise model. In considering possible interactions, there is the

question of what scale do biotic interactions occur among species. Godsoe and Harmon (2012) noted that some authors have argued that most biotic interactions occur over small scales (tens to hundreds of meters), while abiotic factors vary over a much broader scales. As such, SDMs (species distribution models) at large scales represent an effect of the abiotic environment on a species.

5.4.2 *Markov Networks*

Harris (2016) introduced Markov networks (Markov random fields) from statistical mechanics to predict the direct and indirect consequences of any possible species interactions. He found that Markov networks consistently outperformed six other existing approaches that included the use of various correlations. They also correctly isolated direct interactions between species pairs even when indirect interactions or abiotic factors largely overpowered them. The model is described briefly as follows.

A Markov network defines the relative probability of observing a given vector of species-level presences (1's) and absences (0's), \mathbf{y} , at a site for n species as

$$\Pr(\mathbf{y} | \boldsymbol{\alpha}, \boldsymbol{\beta}) \propto \exp \left[\sum_{i=1}^n \alpha_i y_i + \sum_{i,j} \beta_{ij} y_i y_j, \right]$$

where the second sum is over all $1/2(n(n - 1))$ pairs of n species. Here, α_i controls the prevalence of species i , while β_{ij} refers to the joint contribution of species i and j and relates to the probability that they will be found together, after controlling for the other species in the network. Combining all the data into a likelihood and normalizing the probabilities, the α and β parameters can be estimated using maximum likelihood. Harris used a small example for two species to demonstrate how a network is set up using 2×2 tables of data showing the presence/absence of two species.

Clark et al. (2018) used multispecies occurrence data and the previous method to estimate interspecific interactions while controlling for indirect interactions. They focused on determining interactions between co-occurring species using the Markov networks and a logistic model for observing species i , given the presence-absence of species j , using various covariates. So-called regularization was used (LASSO) due to its ability to simultaneously perform variable selection with the least important variables being more likely to be penalized at lower regularization values and model shrinkage to avoid overfitting. Covariates needed to be appropriately scaled.

Mayfield and Stouffer (2017) noted that higher-order interactions may be needed to model species interactions. Although used initially for annual plants, the following model can be used more generally, namely,

$$F_{ji} = \lambda_i \exp[D_{ji} | \mathcal{N}] \exp[H_{ji} | \mathcal{N}],$$

where F_{ji} is the fecundity of a focal individual j of species i in the presence of a specified set $\{\mathcal{N}\}$ of neighbors. Here, λ_i is the intrinsic fecundity of individuals of species i at a given site (e.g., their fecundity when no competitors are present locally), and the two terms in the exponentials capture the direct (D_{ji}) and higher-order (H_{ji}) effects (i.e., cumulative impacts of interactions) of all neighboring species, respectively. These last two expressions are modeled, and their paper is referred to for details.

5.4.3 Environmental Effects

Dorazio and Connor (2014) were motivated by a desire to determine whether the effects of environmental (habitat) covariates or the role of interactions between species were more influential in determining the abundances of individual species before moving on to species' counts as follows.

Let N_{ij} denote a random variable for the number of individuals of species i ($i = 1, \dots, I$) that are present and available to be observed at site j ($j = 1, \dots, J$). It is assumed that the expected abundance of individuals of species i at location j , say λ_{ij} , is constant during surveys of each location and that N_{ij} is Poisson(λ_{ij}). Also, it is assumed that $\log(\lambda_{ij}) = \mathbf{b}_i' \mathbf{x}_j$, where $\mathbf{x}_j = (1, x_{1j}, \dots, x_{qj})'$ are q covariates and $\mathbf{b}_i = (b_{0i}, \dots, b_{qi})'$ includes an intercept. If the values of each covariate are centered at zero, the intercept parameter b_{0i} denotes the log scale, expected abundance of species i at the average value of the covariates.

To allow for potential correlations in abundance among species, the authors specified dependence among the I intercepts using the I -dimensional multivariate normal distribution:

$$(b_{01}, \dots, b_{0I})' \sim N(\beta_0 \mathbf{1}_I, \sigma_{b_0}^2 \mathbf{R}),$$

where $\mathbf{1}_I$ is an I -dimensional vector of ones, β_0 and $\sigma_{b_0}^2$ parameterize the mean and variance among intercepts, and \mathbf{R} is their $I \times I$ correlation matrix. To ensure that $\mathbf{R} = (r_{ij})$ is positive definite, let $r_{ij} = \exp[-d_{ij}/\phi]$ specify the correlation between species i and j as a function of a positive scalar-valued parameter ϕ and a measure of dissimilarity d_{ij} in traits of species i and j . The authors proposed the Euclidean distance $d_{ij} = \| \mathbf{v}_i - \mathbf{v}_j \|$ between trait vectors \mathbf{v}_i and \mathbf{v}_j , though other distance

measures can be used (cf. Seber, 1984, Section 5.5.1, for properties). Finally, it is assumed that for the m th covariate ($m = 1, \dots, q$),

$$(b_{m1}, \dots, b_{mI})' \sim N(\beta_m \mathbf{1}_I, \sigma_{b_m}^2 \mathbf{I}),$$

where \mathbf{I} is an $I \times I$ identity matrix. Two special cases are of interest: first, \mathbf{R} is block diagonal when there are guilds (any group of species that exploit the same resources or that exploit different resources in related ways), and the guilds are independent; and second, $\mathbf{R} = \mathbf{I}$ if species are independent.

Mattsson et al. (2013) considered interactions between mobile species and land cover characteristics. Let $z_{ij} = 1$ if species i is present at site j , namely,

$$z_{ij} \sim \text{Bernoulli}(\psi_{ij})$$

with observation model and K visits

$$y_{ijk} \mid z_{ij} = 1 \sim \text{Bernoulli}(p_{ijk}),$$

where p_{ijk} is the probability of detection for species i at site j (conditional on presence) during k visits, with closure assumed. To account for spatial autocorrelation, they included an autologistic variable that is specified as follows:

$$P_{ij} = \frac{\sum_r z_{ir}}{R_j},$$

Here, P_{ij} is the proportion of sites neighboring j that are occupied by species i estimated by dividing the total number of presences (in each of the r neighboring cells) of species i by R_j , the total number of sites that are neighbors to site j . Thus, all neighboring sites were given equal weight. They then used

$$\text{logit}(\psi_{ij}) = \alpha_{1i} + \alpha_{2i} P_{ij} + \alpha_{3i} x_1 + \dots + \alpha_{7i} x_5$$

with five land cover covariates and

$$\text{logit}(p_{ijk}) = \beta_i,$$

where the α 's and β are random effects and vague Bayesian priors were used for their parameters. Computational details are given by the authors as well as references about the role of spatial autocorrelation in the analysis of species distribution data.

5.4.4 Species' Counts

Looking now at species' counts and detection probabilities, we continue with Dorazio and Connor (2014). Let K_j independent surveys be taken at the j th site such that each species' abundance remains constant during the period of sampling. Let y_{ijk} be the number of individuals of species i detected during the K_j th surveys of site j . Then $y_{ijk} | N_{ij} \sim \text{Binomial}(N_{ij}, p_{ijk})$, where p_{ijk} denotes the conditional probability of detecting an individual of species i during the k th survey of site j . In addition, because surveys are conducted independently at each site, we assume that each site's observations are conditionally independent, that is,

$$\Pr(y_{ij1}, \dots, y_{iK_j} | N_{ij}, \{p_{ijk}\}) = \prod_{k=1}^{K_j} \Pr(y_{ijk} | N_{ij}, p_{ijk}).$$

Various models can be used for p_{ijk} such as a logistic model with covariates, namely,

$$\text{logit}(p_{ijk}) = \alpha_{0i} + \boldsymbol{\alpha}' \mathbf{w}_i,$$

where the α_{0i} are independently and identically distributed as $N(\alpha_0, \sigma_{\alpha_0}^2)$. The authors noted that with this formulation, latent (unobserved) sources of heterogeneity in detection among species are specified exchangeably using the normal distribution, whereas systematic (observed) sources of heterogeneity in detection are specified by the regression parameters $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_q)'$. They used Markov chain Monte Carlo methods to estimate the parameters previously described as well as the variance parameters. For an example, they analyzed species- and location-specific counts of 73 avian species observed while sampling 46 tracts of forest in Southeastern Connecticut (Askins et al., 1987) and included a number of appendices relating to the data and computations. Several kinds of information about individual species were used for different components of their model. One important limitation of their model is that species abundances can only be positively correlated owing to the assumed structure of the correlation matrix \mathbf{R} .

The above methodology was developed further by Dorazio et al. (2015). They believed that models of species-specific counts may be more useful for estimating the effects of biological interactions than models of presence-absence data, with particular emphasis on N-mixture models as they accounted for detection errors. Let N_{ij} be the number of individuals present and available of species i ($i = 1, \dots, I$) at site survey j , and define $\mathbf{N}_j = (N_{1j}, \dots, N_{Ij})'$. It is assumed that \mathbf{N}_j has a multivariate Poisson lognormal distribution with covariates, namely,

$$f(\mathbf{N}_j | \boldsymbol{\varepsilon}_j) = \prod_{i=1}^I f(N_{ij} | \lambda_{ij}), \quad \lambda_{ij} = \exp[\mu_{ij} + \varepsilon_{ij}],$$

and

$$g(\boldsymbol{\varepsilon}_j) = (2\pi)^{-I/2} |\boldsymbol{\Sigma}|^{-1/2} \exp[\boldsymbol{\varepsilon}'_j \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon}_j / 2],$$

where f is the Poisson distribution with mean λ_{ij} and $g(\boldsymbol{\varepsilon}_j)$ is the multivariate normal distribution of the residual vector $\boldsymbol{\varepsilon}_j = (\varepsilon_{1j}, \dots, \varepsilon_{Ij})$ with mean $\mathbf{0}$ and variance-covariance matrix $\boldsymbol{\Sigma}$. The multivariate Poisson lognormal is a mixture of Poisson distributions that allows the specification of habitat effects and species interactions. We have $\log(\lambda_{ij}) = \mu_{ij} + \varepsilon_{ij}$, and we can use $\mu_{ij} = \boldsymbol{\beta}'_i \mathbf{x}_{ij}$ with covariates \mathbf{x}_{ij} . Also, the multivariate normal distribution implies that the conditional mean of the residual abundance of one species can be expressed as a linear combination of the residual abundances of all the other species.

Having developed the underlying unknown model structure, they added in the data observations. Let y_{ijk} be the number of individuals of species i detected on the j th site from the k th replicate ($k = 1, \dots, K_j$). Define $\mathbf{y}_{ij} = (y_{ij1}, \dots, y_{ijK_j})'$, where it is assumed that

$$\mathbf{y}_{ij} \mid N_{ij} \sim \text{Multinomial}(N_{ij}, \pi_{ij1}, \dots, \pi_{ijK_j}),$$

where

$$\pi_{ijk} = \binom{K_j}{k} p_{ij}^k (1 - p_{ij})^{K_j - k},$$

and p_{ij} is the probability of detecting a single individual of species i during an interval of observation associated with site j . We can then model $\text{logit}(p_{ij}) = \boldsymbol{\alpha}'_i \mathbf{w}_{ij}$ with \mathbf{w}_{ij} relating to site- and species-specific covariates. In the authors' application to 73 species, it was found that using a model selection procedure, the single species N-mixture model, was appropriate for 57 of the 73 species encountered in their surveys. Another 8 of the remaining 16 had counts too low for a multispecies model, and the remaining 8 were analyzed using the multispecies N-mixture model.

Lany et al. (2017) used a dynamic log-linear model to model species' counts from sample sites and allowing for species interaction. Sites were surveyed annually for certain marine species during the summer months using a transect-quadrat method, and various covariates denoted by \mathbf{x} were measured. For their first model, let m_{ijk} be the observed abundance of species i ($i = 1, \dots, I$) at site j and time k . Setting $n_{ijk} = \log(m_{ijk})$, a model allowing for species interactions takes the form

$$n_{ijk} - n_{ij,k-1} = r_i (1 - \sum_{r=1}^I \alpha_{ir} n_{rj,k-1}) / k_i + \beta_{i1} x_{1jk} + \beta_{i2} x_{2jk} + \tau_{ijk}.$$

Here, r_i is the intrinsic growth rate for species i , k_i is the natural logarithm of the carrying capacity of species i , α_{ir} is the interaction coefficient quantifying the effect of species r on species i , and x_{1jk} and x_{2jk} represent abiotic covariates specific to

site j and time k . Also, β_{i1} and β_{i2} are species-specific coefficients describing the effect of the abiotic covariates on the growth rate of species i . The residual error τ_{ijk} was modeled with a multivariate normal distribution with mean vector $\mathbf{0}$ and variance-covariance matrix Σ . A second slightly modified model was also used. Inference was carried in a Bayesian framework using the statistical software **JAGS** (Plummer, 2003) and **R**. Their paper should be consulted for further details. We now consider a state-space model that is being more widely used.

State-Space Model

Miller et al. (2012) used a state-space model to determine how disturbance and species interactions affected the co-occurrence probabilities of an endangered toad and recently introduced non-native predators in stream breeding habitats. During each year, a site belonged to one of the five discrete states: (1) unsuitable; (2) suitable, toads and predators not present; (3) suitable, toads present and predators not present; (4) suitable, toads not present and predators present; and (5) suitable, toads and predators present. The proportion of sites in each of the states in year t was given by state vector $\psi = (\psi_1, \dots, \psi_5)'$, and the changes in the states of sites were modeled as a discrete Markov process where the state of a site in year $t + 1$ was solely a function of the state in year t . This means that in each year, a site either remains in its current state or transitions to any other state with a probability conditional on the starting state.

Probabilities are defined relating to input to sites (colonization, γ 's), output from sites (extinction, ε 's), and probability of detection (given various conditions, p 's), in all 16 parameters. Tables are given for, firstly, transition probabilities from time t to time $t + 1$ expressed as a function of extinction and colonization probabilities for habitat suitability, toad occupancy, and predator occupancy and for, secondly, the probabilities of observing a site being in state (columns) given the true state (rows). As the notation and probabilities are complicated, it is best to refer to the original paper.

5.4.5 Consumer Resource Dynamical Models

Goodsoe and Harmon (2012) took a completely different approach to the species interaction process and used consumer resource dynamics based on differential equations. If N_i is the abundance of species i , it can be modeled as

$$\frac{dN_i}{dt} = N_i(f_{i1}a_{i1}R_1 + f_{i2}a_{i2}R_2 - d_i),$$

for $i = 2$ species. Here, each species increases in abundance as it consumes resources R_1 and R_2 . Also, we have for $j = 1, 2$

$$\frac{dR_j}{dt} = c(S_j - R_j) - \sum_{i=1}^2 f_{ij}N_iR_j$$

Without going into details, the various constants mentioned refer to aspects of feeding, population growth, and mortality. By solving $dN_1/dt = 0$ for the equilibrium case, presence probabilities can be obtained to provide information on the suitability of an environment for the two species together when coexistence is stable.

5.4.6 *Using Absences*

Species distribution models generally attempt to provide detailed species' distributions that endeavor to relate presence to environmental predictions and habitat suitability. In the above methods, we have focused on just presence-only data and ignored absence data. Clearly confirmed absences can be very difficult to obtain, especially for mobile species, and failing to detect a species does not guarantee the species is absent from a site. In some situations, reliable data on absences is not readily available for some species, for example, from "atlases, museum and herbarium records, species lists, incidental observation databases and radio-tracking studies" (Pearce & Boyce, 2006). Also, higher levels of sampling effort are required to ensure the reliability of absence data compared with presence data. One method currently used is to add randomly chosen "pseudo-absences" to the data such that it can be analyzed using logistic regression, an approach which has "weaknesses in model specification, in interpretation, and in implementation" (Warton & Shepherd, 2010).

Hirzel et al. (2001) compared two approaches, a generalized linear model (GLM) using covariates that combined presence and absence data and an ecological niche factor analysis (ENFA) that is a new multivariate analysis method that used just presence-only data. Using simulation by creating a virtual habitat and an ecological niche, they were able to investigate the predictive power of both methods. They found that the power depended on the situation: for example, when absence data are reliable, GLM could get extra power by using this information, but in other situations, it could be misled by false absences. The ENFA appeared to be very robust to data quality and quantity, while the GLM was moderately sensitive to data quality but not to data quantity. ENFA and GLM were not sensitive to sample size, as both analyses produce only slightly better results with 1200 points than with 300 points. ENFA is particularly suitable and efficient when the quality of data is either poor (the absence data are unreliable) or unknown. The GLM offers slightly better results when the available presence/absence data are sufficiently good. Virtual species simulation is a useful technique. Brotons et al. (2004) carried out a similar comparison of both GLM and EFNA with regard to accuracy of predictions and concluded that if absence data is available, methods using this information should be preferably used in most situations.

Elith et al. (2006) noted that there are a large number of past methods for modeling species' distributions and they vary in how they model different variables, allow for interactions, and predict geographic patterns of occurrence. They mentioned

Table 5.1 Confusion matrix

| | | Actual | |
|-----------|---|--------|---|
| | | + | - |
| Predicted | + | a | b |
| | - | c | d |

that, historically, attention turned to adapting presence-absence methods (i.e., those that model a binomial response) to model presence-only data, which used samples of the background environment (random points throughout the study area) or of areas designated as “non-use” or “pseudo-absence.” One way of testing the validity of a distribution in the past has been to use one part of the data to fit the model and then use it to try and predict the remaining data, so-called “training” and “testing” data, though this method can have problems (Chatfield, 1995; Fielding & Bell, 1997).

Forbes (1995) considered various measures for comparing the performance of classification algorithms obtained from the so-called confusion matrix given by Table 5.1. He recommended that a suitable measure should meet four requirements and obey six additional constraints. In particular, it should measure agreement and not association. Table 5.1 was used by Fielding and Bell (1997) with frequencies in four categories, where + denotes presence and – denotes absence. If $n = a + b + c + d$, then the prevalence is $(a + c)/n$, the correct classification proportion is $(a + d)/n$, the false-positive proportion is $b/(b + d)$, the false-negative proportion is $c/(a + c)$, and the odds ratio is $(ad)/(cb)$. Several other measures were defined by Forbes (1995) such as

$$\kappa = \frac{a + d - [(a + c)(a + b) + (b + d)(c + d)]/n}{n - [(a + c)(a + b) + (b + d)(c + d)]/n}.$$

Landis and Koch (1977) have suggested the following ranges of agreement for the κ statistic: poor ($\kappa < 0.4$), good ($0.4 \leq \kappa \leq 0.75$), and excellent ($\kappa > 0.75$). If $e = (c + d) \log(c + d)$, we have the normalized mutual information from Forbes (1995), namely,

$$NMI = \frac{-a \log(a) - b \log(b) - c \log(c) - d \log(d) + (a + b) \log(a + b) + e}{n \log(n) - (a + c) \log(a + c) - (b + d) \log(b + d)}.$$

5.4.7 Time-to-Detection Models

Garrard et al. (2013) extended the TTD (time-to-detection) model of Garrard et al. (2008), considered in Sect. 4.3.5, to include multiple species as follows (with a change in notation) and focusing on plants. They used an exponential-binomial mixture distribution with parameter (intensity) λ adjusted for true absences giving

the density function for a single site and single species, namely,

$$f(t \mid \lambda) = \begin{cases} \lambda \exp[-\lambda], & t > T \\ \exp[-\lambda], & t \leq T, \end{cases}$$

where t is the observed time to detection and T is the duration of the survey. Let i denote the species ($i = 1, \dots, I$), j denote the site or quadrat ($j = 1, \dots, J$), and k denote the observer or survey number ($k = 1, \dots, K$). We define ψ_{ij} to be the occupancy probability of site j by species i , $z_{ij} = 1$ if species i is present at site j and 0 otherwise, and p_{ijk} is the probability of detecting species i , at site j , and for survey k . Then

$$z_{ij} \sim \text{Bernoulli}(\psi_{ij}).$$

Let t_{ijk} be the time to the first detection of species i at site j , and observer (survey) k , with corresponding probability p_{ijk} . The detection probability until time t_{ijk} is $1 - \exp[-\lambda_{ijk}t_{ijk}]$. The likelihood for species i , site j , and all k observers ($k = 1, \dots, K$) is

$$\begin{aligned} L(t_{ij1}, \dots, t_{ijk} \mid \psi_{ij}, \lambda_{ijk}) \\ = \psi_{ij} \prod_k (\lambda_{ijk} \exp[-\lambda_{ijk}t_{ijk}])^{z_{ijk}} (\exp[-\lambda_{ijk}t_{ijk}])^{1-z_{ijk}}, \sum_k z_{ijk} \geq 1, \\ = \psi_{ij} \prod_k (\exp[-\lambda_{ijk}t_{ijk}]) + 1 - \psi_{ij}, \sum_k z_{ijk} = 0. \end{aligned}$$

The first line accounts for the case where species i is seen by at least one of the K observers at site j , and the second line accounts for the case when there were no detections. Covariates were used to model $\log(\psi_{ijk})$ with species traits playing a predominant role. Vague prior distributions were specified to ensure posterior estimates were driven by the data. In particular, ψ_{ij} was assumed to be constant and was assigned a uniform prior distribution on (0,1), while random effects for site, observer, and species were drawn from a normal distribution with a mean of 0 and standard deviation specified as uniform between 0 and 100. Prior distributions for the regression coefficients on detectability variables were specified as normal distributions with a mean of 0 and a standard deviation of 1000.

Bornand et al. (2014), assuming no false positives, compared the above TTD model using two observers, with the traditional one based on replicate visits, and found that the TTD model was as accurate, but its data was easier to obtain. They also assumed that t_{ijk} was exponentially distributed with parameter λ_{ijk} , and whenever a species is not detected after 8 min, the observation was censored at T . Similarly, for an absent species, the observation of time to detection is also censored at T . Given presence (and a change in their notation), let $y_{ijk} = 1$ if species i is

detected at site j at time k and 0 otherwise. Then, for the traditional model,

$$y_{ijk} \mid z_{ij} \sim \text{Bernoulli}(z_{ij} p_{ijk}),$$

where p_{ijk} is the probability of detecting species i , at site j , and for replicate k . The two models were then analyzed by applying Bayesian methods with vague priors using the software **JAGS 3.3.0** called from **R** through package **R2jags**.

The authors commented that although TTD provided reliable estimates of detectability after a single visit, multiple visits may still be desirable in particular situations, at least in a fraction of the surveyed sites. Three advantages of multiple visits are helping to detect false positives due to species misidentifications, considering temporary emigration in organisms with movement, and allowing for temporary unavailability to detection due to species phenology. To deal with the issue of temporary unavailability, TTD models could be extended to use the extra information from visits at different seasons. The authors recommended that some of the resources saved by the cost-efficient estimation of detectability through TTD design be allocated to repeat visits by another observer, among seasons or even within a season. As we have previously seen, there can be a three-level hierarchical model with occupancy, availability (or presence), and detection, and a TTD model can be used on multiple occasions during the growing season.

5.5 Species Richness

In addition to estimating the number of a given species in an area, there is the related and extensive topic of estimating the number of different species present. This topic has already been mentioned in passing, and in this section, we have a closer look at the topic. Some papers cover both topics, so there is some overlap. What is typical of species richness is that we have a few common species and many uncommon or rare species, which can make sampling difficult. An early review on estimating species richness is given by Chao (2005). A problem with estimating species richness is that when using counts from a plot survey, the number of observed species increases with time spent and number of replicate visits. Some graphical examples of this for butterflies were given by Hardersen and Corezzola (2014). A plot of species number versus number of samples (sampling effort) is also referred to as a rarefaction or species accumulation curve (Sect. 5.5.3). A recent review of species richness was given by Henderson (2021, Chapter 13). Camera methods have also been used to estimate species richness (e.g., Tobler et al., 2015). An R package *Rich* to analyze species richness was given by Rossi (2011).

To set the scene, let P be the probability of detecting a particular species (which may be random), let S be the actual number of different species, and let S_{obs} be the observed number of detected species. Then $E[S_{obs}] = SE[P]$, and, given an

estimate \widehat{P} , a simple estimate of S is given by

$$\widehat{S} = S_{obs}/\widehat{P}.$$

However, this approach would need to be modified to allow for species variation of the probability P of being detected, given present, and in the probability of being present, both which will generally vary with the type of species. This means that species counts can be used as an index of true species richness only if $E[P]$ remains constant over species. Also, we shall see below that we can include repeat (replicate) visits to a site.

A state-space approach can be used. Zeller et al. (2011) used multistate models to estimate the probability of presence in each of the three states, state 0 (undetected), state 1 (rare to moderately present), and state 2 (frequently present). Habitat covariates were used for a jaguar study, and ψ_j , the usual probability of presence, was interpreted as the probability of habitat use.

One issue that comes up with estimating species numbers is the effect of plot size in plot sampling that was considered by Peterka et al. (2020), who studied the effect of variable plot size on fen. They noted that one of the fundamental but unresolved questions is whether to combine plots of different sizes in vegetation surveys, as the number of species detected tends to increase with plot size. The authors found that species-area curves displayed shifts in plant species richness along increasing plot size across two study regions and four vegetation types. They concluded that: “In fens, plot sizes of at least $1 m^2$ describe sufficiently the broad-scale pattern in specialists’ diversity as well as the main environmental gradients. The range of plot sizes of $1–25 m^2$ may be safely merged in broad-scale analyses of fen vegetation without introducing substantial bias, at least when compared with other possible uncertainty sources.” For each habitat, semi-log species-area curves of specialists and other species were fitted using generalized additive models (GAM). Non-metric multidimensional scaling (NMDS) was used to compare the resulting patterns with Procrustes analysis (cf. Seber, 1984, Sections 5.5 and 5.6).

We shall consider below a number of methods for determining species richness including both parametric and nonparametric methods. But first we look at two modeling aspects: using replicate visits and the method of data augmentation.

5.5.1 Replicate Visits

When there are repeated visits to a site, capture-recapture methods can be used to estimate species richness. The probability of “capture” p_t of a given species for visit t is now the probability of being observed, given present. Kéry and Schmid (2004) discussed the following model. For a given species i , say, at site j , with three repeat visits, the probability of a site history 101 would have probability $p_1(1 - p_2)p_3$. If $p^* = 1 - (1 - p_1)(1 - p_2)(1 - p_3)$ is the probability that a species present is detected

on at least one visit, then S can be estimated by S_{obs}/\hat{p}^* . Three factors can affect p : time t (i.e., p varies between visits), behavior b (p differs for the first detection and redetection of a species on a site), and species heterogeneity h (every species has a specific p). If just one of these factors is present or more than one, then the models of Chap. 12 can be used. For example, with just heterogeneity, model M_h can be used. An example of the method applied to a large avian monitoring program was given by Kéry and Schmid (2006), and it was also applied to butterflies by Kéry and Plattner (2007) using line transects.

Dorazio et al. (2006) developed the following model. Let x_{ij} denote the number of times that species i ($i = 1, \dots, S$) is detected in T visits to site j ($j = 1, \dots, J$). Let $z_{ij} = 1$ if species i is present at site j with probability ψ_{ij} and 0 otherwise, and let p_{ij} denote the probability of detection of species i , given that it is present at site j . In addition, we assume that if species i is present at site j ($z_{ij} = 1$), then the number of detections is assumed to have a binomial distribution

$$f(x_{ij} | z_{ij}, p_{ij}) = \left[\binom{T}{x_{ij}} p_{ij}^{x_{ij}} (1 - p_{ij})^{T-x_{ij}} \right]^{z_{ij}}.$$

Removing z_{ij} , we get

$$f(x_{ij} | \psi_{ij}, p_{ij}) = \psi_{ij} \binom{T}{x_{ij}} p_{ij}^{x_{ij}} (1 - p_{ij})^{T-x_{ij}} + (1 - \psi_{ij}) I(x_{ij} = 0),$$

where $I(\cdot)$ is the usual indicator function, which equals 1 when its argument is true and 0 otherwise. The authors then assumed that $\text{logit}(\psi_{ij})$ and $\text{logit}(p_{ij})$ are linear regression models of random species-level and site-level effects. The random effects were modeled using normal distributions. They then considered predicting the relationship between the expected number of species that occur in some prescribed region as a function of the area of that region. Kéry and Royle (2009) used a similar model using covariates such as forest and elevation and date and effort. Because of the large number of parameters and random effects, data augmentation (Sect. 5.5.2) was used along with the **WinBUGS** package. The model was applied to the national Swiss breeding bird survey. A related model was developed by Zipkin et al. (2009) with a focus on forest fragmentation using a hierarchical approach.

A further extension was given by Dorazio et al. (2010) to a dynamic model incorporating a metacommunity consisting of a collection of local communities as follows. Let $w_i = 1$ if species i is a member of the metacommunity of species exposed to sampling during any of T sampling periods and 0 otherwise when the occupancy state of that species then equals zero for every time period. Using z again, let z_{ij1} be the usual indicator variable for presence for species i at location j during time period 1, and assume

$$z_{ij1} | w_i, \psi_{ij1} \sim \text{Bernoulli}(w_i \psi_{ij1}),$$

where $\psi_{ij1} = \Pr(z_{ij1} = 1 \mid w_i = 1)$ denotes the probability that species i is present at location j during time period 1, given that this species is a member of the metacommunity. A Markov process was assumed where occupancy states in time periods 2 through T depended on the occupancy states one period earlier as follows:

$$z_{ij,t+1} \mid w_i, z_{ijt}, \psi_{ijt}, \gamma_{ijt} \sim \text{Bernoulli} (w_i[\phi_{ijt}z_{ijt} + \gamma_{ijt}(1 - z_{ijt})]),$$

for $t = 1, \dots, T - 1$, where

$$\gamma_{ijt} = \Pr(z_{ij,t+1} = 1 \mid z_{ijt} = 0, w_i = 1)$$

denotes the conditional probability that the j th location will become occupied by species i during period $t + 1$ (local colonization). Here,

$$\phi_{ijt} = \Pr(z_{ij,t+1} = 1 \mid z_{ijt} = 1, w_i = 1)$$

denotes the conditional probability that the j th location will remain occupied by species i during period $t + 1$, given that this species is a member of the metacommunity and is present at that location during the previous period (called local survival or persistence). The probability of local extinction is $\varepsilon_{ijt} = 1 - \phi_{ijt}$. Also, the ψ_{ijt} are essentially derived parameters because of the recurrence relation

$$\psi_{ij,t+1} = \phi_{ijt}\psi_{ijt} + \gamma_{ijt}(1 - \psi_{ijt}), \quad t = 1, \dots, T - 1,$$

where $\psi_{ij,t+1} = \Pr(z_{ij,t+1} = 1 \mid w_i = 1)$. The authors then used the logistic regressions

$$\begin{aligned} \text{logit}(\psi_{ij1}) &= b_{0i} + b_{1i}x_{j1} \\ \text{logit}(\gamma_{ijt}) &= c_{0it} + c_{1it}x_{jt} \\ \text{logit}(\phi_{ijt}) &= d_{0it} + d_{1it}x_{jt}, \end{aligned}$$

for $t = 1, \dots, T - 1$. The same covariate x is assumed to affect species occurrence, colonization, and survival probabilities; however, different covariates can be used if there are other factors.

Looking now at the data and ignoring community membership for the moment, we can, for each species i , sample location j , and time period t , observe an encounter history $\mathbf{y}_{ijt} = (y_{ijt,1}, \dots, y_{ijt,K})'$ composed now of K independent binary observations that indicate whether species i is detected ($y = 1$) or not detected ($y = 0$). For example, $\mathbf{y} = (01010)'$ indicates a species that is detected twice (during the second and fourth observations) in $K = 5$ replicate observations. We assume

$$y_{ijt,k} \mid z_{ijt}, p_{ijt,k} \sim \text{Bernoulli}(z_{ijt}p_{ijt,k}),$$

where $p_{ijt,k} = \Pr(y_{ijt,k} = 1 \mid z_{ijt} = 1)$ denotes the conditional probability of detecting the i th species during the k th observation at site j and time period t , given that the species is present. A logistic model can be used again, namely,

$$\text{logit}(p_{ijt,k}) = a_{0it} + a_{1it}u_{jt},$$

where u_{jt} is a covariate observation.

Various species numbers can be estimated, making use of some predicted z_{ijt} . For example, the number of species present at sample location j during period t is $z_{jt} = \sum_{i=1}^S z_{ijt}$. If J is the number of sample (site) locations, the number of sample locations within this region where the i th species is present during period t is $z_{it} = \sum_{j=1}^J z_{ijt}$. These quantities (one for each species) can then be used to compute the number of species present in the region during period t , namely, $\sum_{i=1}^S I(z_{it} > 0)$, where I is the usual indicator function that takes the value 1 when its argument is true and 0 otherwise. We can then move to the community and use data augmentation described below, where M is the known number species in the augmented data set giving us $S = \sum_{i=1}^M w_i$. Computational details of model fitting are given in the appendix of Dorazio et al. (2010).

White et al. (2020) developed a similar model for fisheries. We have

$$z_{ij} \sim \text{Bernoulli}(\psi_{ij}), \quad y_{ijt} \sim \text{Bernoulli}(\psi_{ij} p_{ijt}),$$

and $\text{logit}(\psi_{ij})$ and $\text{logit}(p_{ijt})$ are expressed as linear models in covariates. The regression parameters are given normal prior distributions. The potential correlation between abundance and detection was modeled using a bivariate normal distribution. Species richness was estimated using data augmentation (described below), and attention was given to rare species.

5.5.2 Data Augmentation

Yamaura et al. (2016) used the method of data augmentation (A.5) to account for unobserved species in a community abundance model. Here, we add an arbitrary number $S_+ - S$ of hypothetical, unobserved species to the detection histories of observed species and analyze the augmented set of S_+ species (called the “super-community”), where S_+ is chosen to be much larger than S . Let N_{ij} be the abundance of species i at site j , p_i be the probability that species i is detected, and n_{ijk} be the number of the detected individuals (i.e., the count) of species i ($i = 1, \dots, S$), at site j on visit k (the authors used y_{ijk}). Then, assuming independent observations with identical detection probabilities, we have

$$n_{ijk} \sim \text{Binomial}(N_{ij}, p_i).$$

Also, if we use the Poisson distribution, we have $N_{ij} \sim \text{Poisson}(\lambda_{ij})$, and the authors used the covariate model

$$\log(\lambda_{ij}) = \beta_{0i} + \boldsymbol{\beta}'_i \mathbf{x}_j,$$

where \mathbf{x}_j are site-level covariates and the β parameters were given prior distributions.

Let w_i be a binary, partially observed indicator variable that takes the value of 1 if a species in the super-community is a member of the community of S species that are exposed to sampling and 0 otherwise. It is known to be 1 for all species that are observed at least once, but its value must be estimated for the augmented species. We assume that the w_i are mutually independent Bernoulli random variables with an inclusion parameter P so that

$$w_i \sim \text{Bernoulli}(P),$$

$$S \sim \text{Binomial}(S_+, P),$$

and $S = \sum_{i=1}^{S_+} w_i$. In the analyses, the authors used a sufficiently large value of S_+ such that all the mass of the posterior distribution of P is well away from 1. Here, the detection histories are zeros for an augmented species and for an existing species not present or present but not observed, giving zero inflations (see Sect. 1.5). Then,

$$n_{ijk} | w_i \sim \text{Binomial}(N_{ij} w_i, p_i),$$

so that for species that are not exposed to sampling (with $w_i = 0$), the counts are necessarily zero. The authors also estimated site-specific species richness and total abundance of communities as derived parameters using the posterior samples of the latent variables N_{ij} and w_i . One can similarly estimate the number of species occurring at any site. The model was assessed using simulations and fitted using the R package R 3.0.2 (R Core Team) and JAGS 3.4.0 (Plummer, 2003), via the package R2jags0.03-11 (Su & Yajima, 2013). See also Chapter 11 in Kéry and Royle (2016) for more information and an R function that can be used to simulate community abundance data.

5.5.3 Parametric Methods

Chao and Chiu (2016a) reviewed methods of species richness. Initially, the focus was on finding the relationship between species richness and locality through the *species accumulation curve* (or collector's curve, species-cover curve). This is a plot of the accumulated number of species (y-axis) found with respect to the number of units of effort expended (x-axis), where effort can be of a continuous type (area, trap time, volumes) or a discrete type (such as number of individuals,

sampling occasions, quadrats, plots, traps, number of nets, and transect lines). Some clarification and early history of this type of curve were given by Chiarucci et al. (2008).

The curve, as a function of effort (or sample size), monotonically increases and typically approaches an asymptote, which can be used as an estimate of the total number of species. The idea here is that a very large sample could give a good estimate of species richness. The shape of the curve is affected by changes in species richness, overall abundance and diversity, and the proportion of rare species. Eleven nonlinear regression models were fitted by Thompson et al. (2003) to the curves based on pit-trapping data for reptiles in Western Australia to see which one fitted best.

Several curve-fitting approaches have been used such as the traditional one of fitting a parametric curve to the observed species frequencies using one of the several contenders. This leads to a prediction of the graph asymptote and an estimate of species richness. There are two problems with this model-based approach. Different models can fit the data well, but can give rise to very different estimates. Variances of the resulting asymptotes cannot be evaluated without imposing further assumptions. Unfortunately, extensive numerical procedures are typically required to find the species richness estimates under these models. However, this approach does not directly use information on the frequencies of common and rare species, but rather only uses presence data to forecast the shape and asymptote of the rising curve. Also, the parametric approach does not permit meaningful comparisons of assemblages with different distribution functions.

We now look at further methods of estimating species richness and follow the general approach of Chao and Chiu (2016a,b), considered in detail below. Here, the general term “sampling unit” refers to a quadrat, occasion, site, transect line, team, occasion, fixed period of time, fixed number of traps, investigator, and so on. To set the scene, we begin with a population of N individuals and S distinct species and let $N_i > 0$ be the number of individuals of the i th species ($i = 1, \dots, S$), giving $N = \sum_{i=1}^S N_i$. The relative abundance of species i is $p_i = N_i/N$, so that $\sum_{i=1}^S p_i = 1$. Two sampling data structures are first considered, “abundance” data and “incidence” data. In many biological studies (e.g., bird, insect, mammal, and plant), it is often the case that an individual is observed or encountered one at a time and classified as to its species identity, leading to incidence data. We first focus on these two types of data.

Abundance Data

It is assumed that a random sample of n individuals (called the reference sample) is taken from a population of S species and a total of S_{obs} species are observed. This type of data can be obtained by two different sampling schemes. (i) The first is the discrete-type sampling in which the sampling unit is an individual and we sample a fixed number of n individuals in a study area. Here, sample size n is fixed by design, and each species can be represented by at most n individuals. (ii) The second is the continuous-type sampling in which sampling efforts are measured on a continuous scale such as time, area, or water volume. For example, we sample a fixed area or for a fixed amount of time in a study site.

For discrete-type sampling, let $x_i > 0$ be the number of individuals of the i th species observed in the sample, and let f_k (abundance frequency counts), $k = 0, 1, \dots, n$, be the number of species represented by exactly k individuals in the reference sample. Then $n = \sum_{i=1}^S x_i = \sum_{k \geq 1} k f_k$, and $S_{obs} = \sum_{k \geq 1} f_k$. Here, f_0 is the number of undetected species in the sample, which is present in the population of N individuals and S species, but not detected in the sample of n individuals and S_{obs} species. This means that $S = S_{obs} + f_0$. Estimation is based on a set of sampling units in which the incidence of each species is recorded in each sampling unit instead of its abundance. If the sampling unit is the individual, let θ_i be the probability of detecting an individual of species i . Then the species detection probability ϕ_i for the i th species in any observation is, for $i = 1, \dots, S$ and $p_i = N_i/N$,

$$\begin{aligned}\phi_i &= \frac{N_i \theta_i}{\sum_{k=1}^S N_k \theta_k} \\ &= \frac{p_i \theta_i}{\sum_{k=1}^S p_k \theta_k}.\end{aligned}$$

The x_i now have the multinomial probability function

$$f(x_1, \dots, x_S) = \frac{n!}{\prod_{i=1}^S x_i!} \prod_{i=1}^S \phi_i^{x_i}, \quad (5.2)$$

with marginal binomial distributions. In discrete-type sampling, it is assumed that the sampling procedure itself does not substantially alter the species detection probabilities ϕ_i . This assumption is fulfilled if individuals are sampled with replacement so that any individual can be repeatedly observed.

For continuous-type sampling, it can be assumed that the population is surveyed through continuous sampling efforts and that the total amount of effort is increased from 0 to A units. One approach is to assume that individuals of the i th species arrive in a sample according to a Poisson process with mean species occurrence or detection rate $A\lambda_i$, where λ_i represents the mean rate per unit of effort. We then get the product Poisson distribution

$$f(x_1, \dots, x_S) = \prod_{i=1}^S (A\lambda_i)^{x_i} \frac{\exp[-A\lambda_i]}{x_i!}, \quad (5.3)$$

and, if we condition on n , we get the multinomial distribution

$$f(x_1, \dots, x_S | n) = \frac{n!}{\prod_{i=1}^S x_i!} \prod_{i=1}^S \left(\frac{\lambda_i}{\sum_{k=1}^S \lambda_i} \right)^{x_i}. \quad (5.4)$$

If we assume that λ_i is proportional to the product of species abundance N_i and individual detectability θ_i , then this conditional multinomial distribution is identical to (5.2). In general, we can assume that λ has a mixing density function $f(\lambda; \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ denotes a vector of parameters for the species detection rates $(\lambda_1, \dots, \lambda_S)$ in the above equations.

Let $p_{\boldsymbol{\theta}}(k)$ ($k = 0, 1, \dots$) be the probability that any species is observed k times in the sample. Then from Eq. (5.3), we have

$$p_{\boldsymbol{\theta}}(k) = \int_0^\infty (A\lambda)^k \frac{\exp[-A\lambda]}{k!} f(\lambda; \boldsymbol{\theta}), k = 0, 1, \dots, \quad (5.5)$$

and $E[f_k] = Sp_{\boldsymbol{\theta}}(k)$. Suppose that each species can be classified into the following disjoint categories: undetected, detected once, detected twice, and so on. Then the likelihood function for S and $\boldsymbol{\theta}$ from all species can be written as

$$L(S, \boldsymbol{\theta}) = \frac{S!}{(S - S_{obs})! \prod_{k \geq 1} f_k!} [p_{\boldsymbol{\theta}}(0)]^{S - S_{obs}} \prod_{k \geq 1} [p_{\boldsymbol{\theta}}(k)]^{f_k}, \quad (5.6)$$

which can be used to obtain maximum likelihood estimators of S and $\boldsymbol{\theta}$ and asymptotic variances. We now have the factorization

$$L(S, \boldsymbol{\theta}) = L_b(S, \boldsymbol{\theta}) L_c(\boldsymbol{\theta}),$$

where we have the binomial distribution

$$L_b(S, \boldsymbol{\theta}) = \binom{S}{S_{obs}} (1 - p_{\boldsymbol{\theta}}(0))^{S_{obs}} (p_{\boldsymbol{\theta}}(0))^{N - S_{obs}} \quad (5.7)$$

and

$$L_c(\boldsymbol{\theta}) = \frac{S_{obs}}{\prod_{k \geq 1} f_k!} \prod_{k \geq 1} \left(\frac{p_{\boldsymbol{\theta}}(k)}{1 - p_{\boldsymbol{\theta}}(0)} \right)^{f_k}. \quad (5.8)$$

If $\hat{\boldsymbol{\theta}}$ maximizes likelihood (5.8), then a conditional maximum likelihood estimate for S obtained from (5.7) is

$$\hat{S}_{cond} = S_{obs} / [1 - p_{\hat{\boldsymbol{\theta}}}(0)].$$

Fisher et al. (1943) adopted a general gamma distribution with $\boldsymbol{\theta} = (\tau, \beta)'$ and density function

$$f(\lambda; \tau, \beta) = \beta^{-\tau} \lambda^{\tau-1} \exp[-\lambda/\beta] / \Gamma(\tau),$$

which leads to the negative binomial distribution

$$p_{\theta}(k) = \frac{\Gamma(k + \tau)}{\Gamma(k + 1)\Gamma(\tau)} \left(\frac{\beta}{1 + \beta} \right)^k \left(\frac{1}{1 + \beta} \right)^{\tau}, \quad k = 0, 1, \dots$$

Various parametric distributions have been proposed, and, as noted above, the problem is that two models with different parametric functions or distributions may fit the data equally well, but yield widely different estimates. In addition, these approaches do not perform well in comparisons with empirical or simulated data sets.

Incidence Data

Here, estimation is based on a set of sampling units in which the incidence of each species is recorded in each sampling unit (e.g., a trap, net, quadrat, plot or timed survey). It is these sampling units, and not the individual organisms, that are actually sampled randomly and independently. The reference sample for incidence data consists of a set of T sampling units. Let $w_{ij} = 1$ if species i is detected in the j th sampling unit and 0 if it is not detected, and let $\mathbf{W} = (w_{ij})$ denote the $S \times T$ matrix. Then, analogous to x_i in the individual-based frequency vector, let $y_i = \sum_{j=1}^T w_{ij}$ denote the incidence-based frequency of species i ($i = 1, \dots, S$). As before, the total number of species observed in the reference sample is S_{obs} with only species with $y_i > 0$ contributing to S_{obs} .

Let g_k denote the number of species that are detected in exactly k sampling units, $k = 0, \dots, T$, that is g_k is the number of species each represented exactly $y_i = k$ times in the incidence matrix sample. Here, g_0 denotes the number of species among the S species not detected in any of the T sampling units. Then, similar to n for abundance data, we have $u = \sum_{k=1}^T kg_k = \sum_{i=1}^S y_i$. Here, $S = S_{obs} + g_0$. A parametric model can be used for incidence data, and there is an equivalence in terms of the parameters between the models for incidence data and abundance data spelt out by Chao and Chiu (2016a,b). Also, the theory behind capture-recapture heterogeneity-type models can be applied here.

5.5.4 Nonparametric Estimates: Abundance Data

As there are some problems with parametric models, briefly referred to above, we turn to nonparametric models to estimate species' richness that use the frequencies f_k .

Chao1 Estimators

Chao (1984, 1987) obtained a lower bound, essentially an asymptotic estimate, and variance estimate using just f_1 (singletons) and f_2 (doubletons) called the Chao1

estimator in the literature. This is given by

$$\widehat{S}_{Chao1} = \begin{cases} S_{obs} + [(n-1)/n][f_1^2/(2f_2)], & \text{if } f_2 > 0 \\ S_{obs} + [(n-1)/n][f_1(f_1-1)/2], & \text{if } f_2 = 0, \end{cases}$$

which is valid under both the multinomial and product Poisson models discussed above. If $f_2 > 0$, we have the variance estimator

$$\widehat{\text{var}}[\widehat{S}_{Chao1}] = f_2 \left[\frac{a}{2} \left(\frac{f_1}{f_2} \right)^2 + a^2 \left(\frac{f_1}{f_2} \right)^3 + \frac{a^2}{4} \left(\frac{f_1}{f_2} \right)^4 \right],$$

where $a = 1 - 1/n$. If $f_2 = 0$, the above variance formula is modified to

$$\widehat{\text{var}}[\widehat{S}_{Chao1}] = \frac{af_1(f_1-1)}{2} + \frac{a^2 f_1(2f_1-1)^2}{4} - \frac{a^2 f_1^4}{4\widehat{S}_{Chao1}}.$$

A confidence interval of species richness based on the Chao1 estimator can be constructed using an asymptotic variance and a log transformation (Chao, 1987; Chao & Shen, 2010) so that the lower bound of the interval is not less than S_{obs} . The Chao1 estimator, although a lower bound, works satisfactorily when an undetected species in the sample has approximately the same probability of being detected as a singleton. This condition is satisfied if the sample size is very large or the rare species are nearly homogeneous in terms of detection probabilities; in the latter case, other species could be highly heterogeneous. For the homogeneous case (i.e., all species detection probabilities or rates are equal), a bias-corrected estimator (referred to as Chao1-bc estimator) is

$$\widehat{S}_{Chao1-bc} = S_{obs} + \frac{n-1}{n} \frac{f_1(f_1-1)}{2(f_2+1)}.$$

A improved lower bound that makes use of f_3 and f_4 as well, called the iChao1 estimator (Chiu et al., 2014), is

$$\widehat{S}_{iChao1} = \widehat{S}_{Chao1} + \frac{n-3}{n} \frac{f_3}{4f_4} \times \max \left(f_1 - \frac{n-3}{n-1} \frac{f_2 f_3}{2f_4}, 0 \right).$$

The authors also provided an analytic variance estimator to construct the associated confidence intervals. Chao and Lin (2012) extended the Chao1 estimator to deal with data based on sampling without replacement (wor) (using N_i the number of species i)

$$f(x_1, \dots, x_S) = \binom{N_1}{x_1} \binom{N_2}{x_2} \cdots \binom{N_S}{x_S} / \binom{N}{n}$$

to get

$$\widehat{S}_{Chao1.wor} = S_{obs} + \frac{f_1^2}{2nf_2/(n-1) + nf_1/(N-n)}.$$

Coverage and Jackknife Estimators

Chao and Chiu (2016a,b) considered several other estimators such as the ACE (abundance-based coverage estimator) of species richness based on the concept of “sample coverage” and the modified estimator ACE-1, particularly for species-rich and highly heterogeneous populations. Also mentioned are jackknife estimators, where a first-order estimator is

$$\widehat{S}_{jk1} = S_{obs} + \frac{n-1}{n} f_1 \approx S_{obs} + f_1$$

and a second-order estimator is

$$\widehat{S}_{jk2} = S_{obs} + \frac{2n-3}{n} f_i - \frac{(n-2)^2}{n(n-1)} f_2 \approx S_{obs} + 2f_1 - f_2.$$

Extensive simulations conducted by Chiu et al. (2014) based on various species abundance models revealed that there is a limited range where jackknife estimators are close to the true species richness. However, the theoretical behavior is not predictable because the narrow range of good performance changes with each model and outside this range, the two jackknife estimators may have appreciable biases.

5.5.5 Nonparametric Estimates: Incidence Data

Corresponding to the Chao1-type estimators of species richness for abundance data, we have for incidence data the Chao2-type and ICE-type estimators. The estimation procedures are generally parallel so that we simply replace the sample size n and the capture frequencies f_k ($k = 1, \dots, n$) by the number of replicate samples T and the incidence-based frequencies g_k ($k = 1, \dots, T$) in the formulae involving Chao1-type estimators. We can use the same formulae above with \widehat{S}_{Chao2} instead of \widehat{S}_{Chao1} , T instead of n , and replace a by $b = (T-1)/T$. For example,

$$\widehat{S}_{Chao2} = \begin{cases} S_{obs} + [(T-1)/T][g_1^2/(2g_2)], & \text{if } g_2 > 0 \\ S_{obs} + [(T-1)/T][g_1(g_1-1)/2], & \text{if } g_2 = 0, \end{cases}$$

and Chiu et al. (2014)

$$\widehat{S}_{iChao2} = \widehat{S}_{Chao2} + \frac{(T-3)}{4T} \frac{g_3}{g_4} \times \max \left(g_1 - \frac{(T-3)}{2(T-1)T} \frac{g_2 g_3}{g_4}, 0 \right).$$

There is also an incidence-based coverage estimator ICE-type, and jackknife estimators follow from the previous sections using the same notation substitutions. For further comments and applications of the Chao estimator, see Chao and Colwell (2017).

Species Abundance

Norris and Pollock (1998) considered nonparametric maximum likelihood estimation (MLE) for Poisson species abundance models allowing for heterogeneity between species. Changing their notation, let x_i denote the (counted) abundance of species i ($i = 1, \dots, S$), and assume x_i is Poisson(λ_i), where the λ_i 's are from a common distribution $F(\cdot)$. Then, unconditional on λ_i , the x_i have a Poisson(F) mixture distribution. Let f_k denote the number of species having abundance k , i.e., $f_k = \sum_{i=1}^S I(x_i = k)$, where $I(\cdot)$ is the indicator function. All the f_k 's except f_0 are observable. Let \mathbf{f} be the vector of the f_k , let $t = \max_k(x_k)$, and let $S_{obs} = f_1 + \dots + f_t$ be the total of observed distinct species. Assuming independence of the S species' abundances, then the likelihood function is

$$L(S, F | \mathbf{f}) = \frac{S!}{(\prod_k f_k!) (S - S_{obs})!} \left\{ \prod_{k \neq 0} (P_{k,F})^{f_k} \right\} \left\{ (P_{0,F})^{S - S_{obs}} \right\},$$

where

$$P_{k,F} = \int_0^\infty \frac{\exp[-\lambda] \lambda^k}{k!} dF(\lambda).$$

The nonparametric MLE for (S, F) , say (S^*, F_{N^*}) , is found using an analogous EM algorithm and global convergence algorithm as that used for the binomial and censored geometric mixtures of Norris and Pollock (1996a); see their paper for details. Bootstrapping can be used here. Estimators are given for various diversity measures.

Wang and Lindsay (2005) considered several estimators of species numbers and referred to some problems with coverage methods. The nonparametric maximum likelihood estimator can have a severe instability problem, which can be alleviated by using an appropriate penalty. Using a similar Poisson model and method to that given above, they used a similar likelihood, but factorized it into the usual conditional multinomial and binomial distributions used in capture-recapture models for estimation (Sanathanan, 1972a, 1977). They subtracted a penalty function from the log-likelihood and maximized the result to obtain estimates. Bootstrap methods were used.

5.5.6 Rarefaction and Extrapolation Methods

Gotelli and Chao (2013) and Chao and Chiu (2016a,b) discussed methods based on the idea of rarefaction, referring to both abundance and incidence data, as well as coverage and extrapolation. Gotelli and Chao (2013) defined rarefaction “as a statistical interpolation method of rarefying or thinning a reference sample by drawing random subsets of individuals (or samples) in order to standardize the comparison of biological diversity on the basis of a common number of individuals or samples.” A rigorous comparison of rarefaction curves at a common number of individuals or a common number of sampling units then requires computation of confidence intervals for these curves. Chao et al. (2015a) applied the methods to phylogenetic diversity.

Chao and Jost (2012) noted that the sample richness of species might never stabilize as the sample size increases, especially if there are many rare species (e.g., in a community of tropical arthropods or orchids). As sample richesses of such communities depend strongly on sample size, one way of trying to control this dependence and create “a level playing field” is to use rarefaction that refers to down-sampling the larger samples until they are of the same size as the smallest sample. However, this does not necessarily make for a fair comparison, as samples standardized by size will usually have different degrees of completeness depending on the species abundance distributions of the communities being compared. As an example, the authors noted that a temperate-zone tree community with 10 species might be completely characterized by a sample of 100 individuals, but the same-size sample would greatly underestimate the richness of a tropical rain forest with 300 tree species. Standardizing to a sample of 100 would not give a clear indication as to the difference between the 2 communities.

A solution is to compare samples using the idea of equal “completeness,” not equal size, and completeness is better determined by using the concept of coverage rather than sample size. Often, the completeness of a sample is determined by looking at the final slope of the rarefaction curve calculated from the species frequencies (“extrapolation” process). The rarefaction sample is considered to be nearly complete if and only if this slope is small. This curve, referred to as the (observed) species accumulation curve (SAC), is a plot of S_m , the number of species seen versus sample size m .

Let p_i be the relative abundance of species i . This is also the probability of detecting species i , the product of the detection probability, given present, times the probability of presence (Chao & Jost, 2012, 25–35). Then, given S species in the population, the sample coverage is defined as

$$C = \sum_{i=1}^S p_i I[x_i > 0],$$

where x_i is the number of species present in the sample. The more the species in the sample, the greater the C . We have from Good (1953)

$$E[S_m] = \sum_{i=1}^S [1 - (1 - p_i)^m] = S - \sum_{i=1}^S (1 - p_i)^m.$$

Also, if C_m is the coverage,

$$\begin{aligned} E[C_m] &= \sum_{i=1}^S p_i [1 - (1 - p_i)^m] \\ &= 1 - \sum_{i=1}^S p_i (1 - p_i)^m, \quad m > 0. \end{aligned}$$

These equations give us

$$1 - E[C_m] = E[S_{m+1}] - E[S_m].$$

This implies that the expected coverage deficit on the left-hand side is equal to the slope of the expected species accumulation curve on the right-hand side (the expected change along the y -axis ($E[S_{m+1}] - E[S_m]$) divided by the corresponding change in the x -axis, which is a sample increment of one individual). This proves that the slope of the expected SAC for a sample of size m is equal to the expected coverage deficit of the subsample. In practice, the *estimated* coverage deficit after m individuals have been sampled will be approximately equal to the line connecting the two points (m, S_m) and $(m + 1, S_{m+1})$. This means that, based on the data, two samples rarefied down to sizes m_1 and m_2 , respectively, are equally complete if and only if the slopes of their rarefaction curves (at x -axis of m_1 and m_2 , respectively) are the same.

Chao et al. (2013) showed that there is a close relationship between the Shannon entropy and the SAC previously described and used it to obtain an accurate, low-bias analytic estimator of entropy, based on species frequency counts. It does not require prior knowledge of the number of species.

Estimators

Chao and Shen (2010) proposed the following coverage estimator from the original data, namely,

$$\widehat{C}_n = 1 - \frac{f_1}{n} \left[\frac{(n-1)f_1}{(n-1)f_1 + 2f_2} \right], \quad (5.9)$$

which generally has a smaller mean square error than the usual Turing's estimator $1 - f_1/n$, where we recall that f_k is the number of species represented by exactly k individuals in the survey. They provided an approximate variance estimator and

associated confidence interval. However, if $C_{m,n}$ is the coverage for a subsample of size m from the reference sample of size n , Chao and Jost (2012) proposed the estimator

$$\widehat{C}_{m,n} = 1 - \sum_{x_i \geq 1} \frac{x_i}{n} \frac{\binom{n-x_i}{m}}{\binom{n-1}{m}}, \quad (m < n).$$

Here, x_i is the number of individuals observed for species i in the reference sample so that $\sum_{i=1}^S x_i = n$. They proved that this estimator is the unique minimum variance unbiased estimator of $E[C_{m,n}]$. This equation is analogous to the traditional rarefaction equation giving the estimated species richness for a subsample, namely,

$$\widehat{S}_{m,n} = S_{obs} - \sum_{x_i \geq 1} \frac{\binom{n-x_i}{m}}{\binom{n}{m}}, \quad (m < n),$$

where S_{obs} is the number of species observed in the reference sample of size n . The estimator $\widehat{S}_{m,n}$ is the unique minimum variance unbiased estimator for $E[S_{m,n}]$, the expected number of species that would be observed in a subsample size of m . The authors gave an algorithm for constructing the coverage-based rarefaction curve.

Chao and Jost (2012) introduced a coverage-based extrapolation method as follows. In a traditional size-based sampling curve given the data for a reference sample of size n , the extrapolation problem is to predict the expected number of species $E[S_{n+m*}]$ in an augmented sample of $n + m*$ individuals from the community. They gave the following, more accurate predictor:

$$\widehat{S}_{n+m*} = S_{obs} + \widehat{f}_0 \left[1 - \left(1 - \frac{f_1}{n \widehat{f}_0 + f_1} \right)^{m*} \right],$$

where \widehat{f}_0 can be estimated as follows using a Chao1 estimator described above, namely,

$$\widehat{f}_0 = \begin{cases} \frac{(n-1)}{n} f_1^2 / (2f_2), & \text{if } f_2 > 0, \\ \frac{(n-1)}{n} f_1(f_1 - 1) / 2(f_2 + 1), & \text{if } f_2 = 0. \end{cases}$$

When $m* \rightarrow \infty$, $\widehat{S}_{n+m*} \rightarrow S_{obs} + \widehat{f}_0$ implying that the Chao1 species richness estimator is the asymptotic value of our extrapolation formula. In a similar way, we have the estimator

$$\widehat{C}_{n+m*} = 1 - \frac{f_1}{n} \left[\frac{(n-1)f_1}{(n-1)f_1 + 2f_2} \right]^{m*+1},$$

which tends to 1 as $m* \rightarrow \infty$, indicating a complete coverage. When $m* = 0$, it reduces to the sample coverage estimator for the reference sample. We also have the result

$$1 - \widehat{C}_{n+m*} = \widehat{S}_{n+m*+1} - \widehat{S}_{n+m*},$$

so that the original slope coverage discussed previously is valid not only for the rarefaction part but also for the extrapolation part. An integrated coverage-based curve includes the rarefaction part \widehat{S}_m versus \widehat{C}_m ($m < n$) and the extrapolation part \widehat{S}_{n+m*} versus \widehat{C}_{n+m*} ($m* > 0$), with both parts joining smoothly at the reference point (\widehat{C}_m, S_{obs}) . The reader is referred to Chao and Jost (2012) for further details and the computational methods used.

Coleman Model

Referring to the Poisson model with intensity λ_i for species i , we consider a “random-replacement” rarefaction method due to Coleman (1981) and Coleman et al. (1982) and as discussed by the extensive paper of Colwell et al. (2012). The expected number of species in a random area of size a within a reference area of size A ($a < A$) is given by

$$E[S_{area}(a)] = \sum_{i=1}^S (1 - \exp[-a\lambda_i]) = S - \sum_{i=1}^S (\exp[-a\lambda_i]).$$

If x_i is the abundance of species i in area a , we have

$$\begin{aligned} \widetilde{S}_{area}(a) &= S_{obs} - \sum_{x_i \geq 1} (1 - (a/A))^{x_i} \\ &= \sum_{k=1}^n [1 - (1 - (a/A))^k] f_k \end{aligned}$$

which is a minimum variance unbiased estimator of $E[S_{area}(a)]$. Here, the area a as a fraction of A gives the rarefaction subsample compared with m the size of the subsample of n in a multinomial type model. If we can assume that individuals are randomly and independently distributed in space, then $a/A \approx m/n$.

In a detailed paper, Colwell et al. (2012) considered two estimation problems: (i) estimating the expected number of species (and its variance) in a random sample of a smaller number of individuals or a smaller area sampled (interpolation) and (ii) estimating the number of species (and its variance) that might be expected in a larger number of individuals or a larger area sampled (rarefaction). Extrapolation was given for two models, the multinomial and Poisson models. The multinomial

model is described above. For the Poisson model, if we knew S_{obs} , extrapolation would require the estimation of

$$E[S_{area}(A + a^*)] = S_{obs} - \sum_{i=1}^S (1 - \exp[-a^* \lambda_i]) \exp[-A \lambda_i],$$

where $A + a^*$ is an augmented area. Chao and Shen (2010) proposed the estimator

$$\tilde{S}_{area}(A + a^*) = S_{obs} + \hat{f}_0 \left(1 - \exp \left[-\frac{a^*}{A} \frac{f_1}{\hat{f}_0} \right] \right)$$

and gave a variance estimator. For a single set of incidence frequencies for species, a Bernoulli product model was also considered for both extrapolation and rarefaction. The methods were applied to data from tropical beetles, trees, and ants. Using species abundances, x_i , data, the authors found that the multinomial and Poisson models produced extremely close results for both interpolation and extrapolation.

5.5.7 Species Ranking

It is of interest to know something about the ranking of relative species numbers, the species-rank abundance distribution (RAD). This has usually used the conventional “plug-in” estimator $\hat{p}_i = x_i/n$, where x_i is the number of species i counted in a sample of size n and p_i is the population proportion or probability of species i . The \hat{p}_i are then plotted as the y -coordinates with their ranking number along the x -axis. The problem with this is that some species will be missing from the sample and we want to allow for both the observed and unobserved species. Chao et al. (2015b) provided a method for doing this using the Chao1 estimator described above. We begin with

$$E[\hat{p}_i \mid x_i > 0] = E[x_i/n \mid x_i > 0] = \frac{p_i}{1 - (1 - p_i)^n}$$

and

$$\begin{aligned} E[\hat{p}_i] &= E[\hat{p}_i \mid x_i > 0] \Pr(x_i > 0) + E[\hat{p}_i \mid x_i = 0] \Pr(x_i = 0) \\ &= \frac{p_i}{1 - (1 - p_i)^n} \times [1 - (1 - p_i)^n] + 0 \times (1 - p_i)^n \\ &= p_i, \end{aligned}$$

for both detection and non-detection. Recalling Equation (5.9), the authors defined a general r th-order coverage

$$C^{(r)} = \frac{\sum_{i \in \text{detected}} p_i^r}{\sum_{i=1}^S p_i^r} = \frac{\sum_{i=1}^S p_i^r I(x_i > 0)}{\sum_{i=1}^S p_i^r}$$

where $I(\cdot, \cdot)$ is an indicator function which equals 1 when $x_i > 0$ and 0 otherwise. This equation reduces to Eq. (5.9) when $r = 1$. We now let $C_d^{(r)} = 1 - C^{(r)}$ be the “deficit.” Then, for f_2 and f_3 both positive, the authors showed in an appendix that for $r = 2$, we have the accurate estimate

$$\widehat{C}_d^{(2)} = \frac{2f_2}{\sum_{x_i \geq 2} x_i(x_i - 1)} \left[\frac{(n - 2)f_2}{(n - 2)f_2 + 3f_3} \right]^2.$$

Using the above equations, some approximations, and some reparameterization, the authors obtained the following corrected proportions for observed species:

$$\widehat{p}_{oi} = \frac{x_i}{n} (1 - \widehat{\lambda} \exp[-\widehat{\theta} x_i]),$$

with the adjustment factor generally close to 1. Assuming $p_i = \alpha\beta^i$ for the undetected species, we have the estimate

$$\widehat{p}_{ui} = \widehat{\alpha}\widehat{\beta}^i, \quad i = 1, \dots, \widehat{f}_0,$$

where \widehat{f}_0 is the Chao1 lower bound described above for the number of undetected species (cf. Chao & Chiu, 2012, for a review). Combining the two methods and adjusting the proportions, accordingly up to the whole population, give final estimates for the RAD graph. The authors extended the method to data from the incidence method (Sect. 5.5.3).

5.5.8 Second Sample Adding New Species

Shen et al. (2003) considered the problem of predicting the number of new species that would be observed in a second survey, based on data from an initial survey. Given S species, let p_i be the relative frequency of species i (or the overall probability of observing species i). In the first survey, n individuals and S_{1obs} species were observed, so that $w = S - S_{1obs}$ is the number of unobserved species to predict. Let f_k be the number of species with k individuals in the survey ($k = 1, \dots, n$); then $S_{1obs} = \sum_{k=1}^n f_k$ and $n = \sum_{k=1}^n kf_k$. The goal is to predict S_2 , the number of new species that will be discovered in a second survey of m individuals, given the information of the initial survey.

As previously defined, the coverage of a survey is the total relative abundances of the species observed in a survey. For the w unseen species, we assume that they are indexed from 1 to w with species relative abundances of p_1 to p_w . Observed species correspond to abundances of p_{w+1} to p_S with sample coverage defined to be

$$C = p_{w+1} + p_{w+2} + \dots + p_S.$$

In other words, the sample coverage denotes the conditional (on data) probability of finding a species that has already been discovered in the sample if an additional observation were to be taken. The well-known estimator of C proposed by A. M. Turing (Good, 1953) is $\hat{C} = 1 - f_1/n$. Shen et al. (2003) showed that this estimator has links with previous estimators, particularly when $p_1 = p_2 = \dots = p_w$. They provided some justification for the estimator

$$\hat{S}_2 = \hat{w} \left[1 - \left(1 - \frac{1 - \hat{C}}{\hat{w}} \right)^m \right],$$

where \hat{w} is some estimate of w such as $\hat{w} = f_1^2/(2f_2)$.

For each of the unobserved species with abundance probability p_i , the probability that it will be missed in the second survey of size m is $(1 - p_i)^m$. Let $\delta_i = 1$ if the i th species is undiscovered in the initial survey of size n and $\delta_i = 0$ otherwise. Then

$$E[S_2 | \text{initial survey data}] = \sum_{i=1}^S [1 - (1 - p_i)^m] \delta_i = w - \sum_{i=1}^w (1 - p_i)^m.$$

Since $E[\delta_i] = \Pr(\delta_i = 1) = (1 - p_i)^n$, we have the unconditional expectation

$$E[S_2] = \sum_{i=1}^S [1 - (1 - p_i)^m] (1 - p_i)^n.$$

A rare species is defined to be one that has r or less individuals in the survey, and an abundant species has more than r individuals. It is reasonable to estimate the number of missing species using the observed rare species, as abundant species would be discovered in a sample survey anyway and thus do not contain any information about the missing species. Based on empirical experience, the authors used $r = 10$.

Let S_{rare} be the total number of rare species observed in the initial survey. The estimated survey coverage for the rare species is

$$\tilde{C} = 1 - f_1 \sum_{i=1}^r (if_i),$$

and an estimate of w is (Chao et al., 2000)

$$\tilde{w} = \frac{S_{rare}}{\tilde{C}} + \frac{f_1}{\tilde{C}} \hat{\gamma}^2 - S_{rare}.$$

Here,

$$\hat{\gamma}^2 = \max \left\{ \frac{S_{rare}}{\tilde{C}} \frac{\sum_{i=1}^r i(i-1)f_i}{(\sum_{i=1}^r if_i)(\sum_{i=1}^r if_i - 1)} - 1, 0 \right\}$$

is the estimated squared coefficient of variation (cv) of the species abundances. The value of cv characterizes the degree of heterogeneity among the species abundances. Then an estimator of S_2 is

$$\tilde{S}_2 = \tilde{w} \left[1 - \left(1 - \frac{1 - \tilde{C}}{\tilde{w}} \right)^m \right],$$

a function of the frequencies $(\tilde{w}, f_1, \dots, f_n)$, and the data can be modeled by an approximate multinomial distribution with an estimated number of species $\tilde{S} = S_{1obs} + \tilde{w}$ and cell probabilities $(\tilde{w}/\tilde{S}, f_1/\tilde{S}, \dots, f_n/\tilde{S})$. Using the delta method for a multinomial distribution (see Appendix A.4), we have the following variance estimator:

$$\widehat{\text{var}}[\tilde{S}_2] = \sum_{i=1}^n \sum_{j=1}^n \frac{\partial \tilde{S}_2}{\partial f_i} \frac{\partial \tilde{S}_2}{\partial f_j} \widehat{\text{cov}}(f_i f_j),$$

where $\widehat{\text{cov}}[f_i, f_j] = f_i(1 - f_i/\tilde{S})$ for $i = j$ and $\widehat{\text{cov}}[f_i f_j] = -f_i f_j/\tilde{S}$ for $i \neq j$. A computer program *SPADE* (Species Prediction And Diversity Estimation), written in C language, was used for the computations. This can be accessed through *SpadeR*, which carries out species richness prediction and diversity estimation in R (Chao et al., 2015c).

5.5.9 Choosing the Sample Size

In the previous section, we looked at increasing the sample size to observe more species, which is critical for rare species. This raises the question of how to estimate the minimum number of additional individuals, samples, or sampling area required to detect any arbitrary proportion (including 100%) of the estimated asymptotic species richness. Chao et al. (2009) addressed this question using the Chao1 and Chao2 nonparametric estimators of asymptotic richness described above, which are based on the frequencies of rare or infrequent species in the original sampling data. The authors considered separate methods as follows for the two types of data

previously mentioned, namely, abundance and incidence data. Examples were given of four abundance-based data sets (butterflies, fish, lizards, and seedlings) and three incidence-based data sets (beetles, ants, and seedlings), as well as simulated data.

Abundance Data

Beginning with abundance data, a random sample of size n is selected with replacement from the community. For large n , we have from above an estimate of the lower bound of species richness

$$\widehat{S}_{Chao1} = S_{obs} + \widehat{f}_0 = S_{obs} + f_1^2/(2f_2).$$

If an original sample of size n is available, Turing and others (Good, 1953 and 2000) proved that, for the next individuals sampled, the probability of encountering each of the f_r species in frequency class r , $r = 0, 1, \dots$ is approximately

$$(r + 1)f_{r+1}/(nf_r).$$

As a special case, the probability of encountering each of the undetected species ($r = 0$) is thus $f_1/(nf_0)$ so that the probability q_0 that the next individual sampled represents a previously undetected species can be estimated by

$$\widehat{q}_0 = f_0 \times f_1/(nf_0) = f_1/n.$$

The following stopping rule was used for m additional observations: Continue until there are no singletons ($f_1 = 0$) in the enlarged sample of size $n + m$. Equivalently, the expected number of singletons in the enlarged sample of size $n + m$ is less than 0.5 (because the theoretical expected value may not be an integer). If $x = m/n$, the authors proved after some algebra that the approximate number of singletons in the enlarged sample with size $n + m$ is $2f_1(1 + x) \exp[-x(2f_2/f_1)]$, which is less than 0.5 if and only if

$$2f_1(1 + x) < \exp[x(2f_2/f_1)].$$

If x^* is the value of x for which we have equality in the above equation, our minimum value of m is $m^* = nx^*$. In many cases, the sampling effort required to reach the asymptote may be prohibitively large. A bootstrap percentile method described in the Supplement of Chao et al. (2009) can be used to construct a lower confidence limit. However, a large fraction, say g ($0 < g < 1$), of \widehat{S}_{Chao1} may be reached with considerably less sampling. The objective then becomes finding the additional sample size m_g such that the number of species reaches the target value $g\widehat{S}_{Chao1}$, i.e., the number of previously undetected species discovered in the additional sample is $g\widehat{S}_{Chao1} - S_{obs}$, where $g > S_{obs}/\widehat{S}_{Chao1}$. After some algebra,

it then follows from Shen et al. (2003, equation 10) that

$$m_g \approx \frac{nf_1}{2f_2} \log \left[\frac{\hat{f}_0}{(1-g)\hat{S}_{Chao1}} \right].$$

Incidence Data

In most biodiversity studies, individual organisms are not sampled randomly and independently, but instead, multiple individuals are collected or censused in traps, baits, quadrats, plots, or timed surveys. It is these sampling units, and not the individual organisms, that are actually sampled randomly and independently. For very abundant organisms (such as microbes), or taxa with clonal growth forms (such as many plants and invertebrates), it may not even be possible to count individuals within each sampling unit, and only their presence or incidence can be recorded. However, estimation is still possible for a set of replicated samples in which the incidence of each species is recorded in the sample. Using incidence data based on T replicated samples, let g_1 and g_2 represent the number of species in exactly one or two samples, respectively. We then have Chao's estimator

$$\hat{S}_{Chao2} = S_{obs} + [(T-1)/T][g_1^2/(2g_2)].$$

Using the same approach as for abundance data, we end up with the number of additional samples needed to achieve a fraction g of \hat{S}_{Chao2} as

$$m_g \approx \frac{\log \left[1 - \frac{T}{T-1} \frac{2g_2}{g_1^2} (g\hat{S}_{Chao2} - S_{obs}) \right]}{\log \left[1 - \frac{2g_2}{(T-1)g_1 + 2g_2} \right]}$$

Chao et al. (2017) extended the above method, using replicated incidence data with T replicates, described below. They noted that a counter-intuitive, but fundamental, concept in estimating the number of undetected species for incidence data is that frequent species (which are likely to be detected in most sampling units) contain almost no information about the richness of undetected species. By contrast, infrequent species (which are likely to be either undetected or detected in only a few sampling units) contain nearly all the information about undetected species richness. We now consider their method.

Let $w_{ij} = 1$ if species i is detected in the j th replicate (sampling unit) and 0 if not detected, and define $\mathbf{W} = (w_{ij})$ to be the $S \times T$ matrix. Let $y_i = \sum_{j=1}^T w_{ij}$ be the incidence-based frequency of species i , ($i = 1, \dots, S$). It is assumed that the detection probability of the i th species in the j th sampling unit is affected by two sources of heterogeneity: a species effect (π_i) and sampling unit effect (v_j) in a multiplicative manner. Then

$$w_{ij} | v_j \sim \text{Bernoulli}(\pi_i v_j)$$

and

$$\Pr(y_i) = \binom{T}{y_i} \lambda_i^{y_i} (1 - \lambda_i)^{T-y_i},$$

where $\lambda_i = \pi_i \int v h(v) d\nu$, so that y_i is a binomial random variable with detection probability λ_i . Recalling that g_k is the number of species that are detected in exactly k sampling units ($k = 0, 1, \dots, T$), let $g_{2+} = g_2 + \dots + g_T$, so that $S_{obs} = g_1 + g_{2+}$. The authors showed that for large T ,

$$g_{2+} \approx \frac{g_1^2}{2g_2} \left[-\log \left(1 - \frac{2g_2}{g_1} \right) - \frac{2g_2}{g_1} \right],$$

which has a unique solution for g_2 that can be substituted into S_{obs} and \widehat{S}_{Chao2} . The error for the above approximation is less than 10% if $T > 10$ and less than 5% if $T > 100$. In summary, the curve of the Chao2 estimator using the estimated g_2 deviates from that using the observed g_2 to only a limited extent. Bootstrapping using a log transformation was used to obtain confidence intervals.

5.5.10 Using eDNA

Olds et al. (2016) assessed fish communities in a stream using two methods, removal electrofishing and environmental DNA metabarcoding (eDNA) from water samples. The aim was to test the hypothesis that eDNA provides an alternative means of determining species richness and species identities for a natural ecosystem. In a Northern Indiana stream, electrofishing yielded a direct estimate of 12 species and a mean estimated richness (biased adjusted Chao2 estimator, Sect. 5.5.5) of 16.6 species (12.8–42.2), while eDNA sampling detected an additional 4 species. If S is the number of species present and S_{obs} is the number observed, then the Chao2 biased corrected estimator of minimum species richness is given by

$$\widehat{S}_{Chao2-bc} = S_{obs} + \frac{n}{n-1} \frac{g_1(g_1-1)}{2(g_2+1)},$$

where n is the sample size, g_1 the number of species with only one incidence of detection, g_2 the number of species with two incidences of detection, and $(n-1)/n$ a correction for small samples. In the authors' example with $n = 10$ electrofishing reaches, $N_{Sob} = 20$ unique species were observed, with four species ($g_1 = 4$) detected only in one reach and three species detected in only two reaches ($g_2 = 3$). Limitations of eDNA metabarcoding were discussed. They concluded that "eDNA metabarcoding can improve the accuracy of species detection for aquatic environments and will be a transformative tool for monitoring aquatic biodiversity on a changing planet."

5.5.11 *Biodiversity*

Here, we consider briefly one other topic with species already alluded to above, which is that of looking at patterns of diversity for species as it has been controversial, and there is a huge range of theories about it (Gelfand et al., 2005). However, following a multi-author Ecology forum (Ellison, 2010), a consensus seems to have emerged that Hill numbers (Hill, 1973, “numbers equivalent”) should be the species diversity measure of choice. Diversity estimates are often used to identify priority areas for protection in conservation planning decisions ranging from local (e.g., patch-specific) to landscape (e.g., municipality) scales. The loss of biodiversity can affect the transmission of infectious diseases (Keesing et al., 2010) by changing the abundance, the behavior, and the condition of the host or vector. Although biodiversity is not one of our main topics, with our main focus on estimating abundance, it is related to species models, so we will try and give an overview.

Buckland et al. (2005) discussed earlier methods of monitoring change in biodiversity through composite indices. Buckland et al. (2017) noted that “In 2002, nearly 200 nations signed up to the 2010 target of the Convention for Biological Diversity, ‘to significantly reduce the rate of biodiversity loss by 2010’. To assess whether the target was met, it became necessary to quantify temporal trends in measures of diversity.” The authors provided a survey of methods for doing this, which provides a useful summary. The 2002 target number was superseded by 20 targets for 2020.

Buckland and Johnston (2017) mentioned key principles and possible pitfalls in the large-scale monitoring of the biodiversity of regions. They stated that well-designed monitoring programs should meet the following criteria: (1) representative sampling locations, (2) sufficient sample size, (3) sufficient detections of target species, (4) representative sample of species (or all species), and (5) a temporal sampling scheme designed to aid valid inference. There needs to be a clear specification of the region, species, and timescale.

Advancing technology is available like camera traps, acoustic detectors, high-resolution photographic imagery used with aircraft and long-range drones, satellite imagery, environmental DNA, citizen science, and computer methods of analysis (e.g., bootstrapping). They note that there is inevitably a trade-off between ideal sampling designs and designs that are realistic and achievable. Rocchini et al. (2016) considered the pros and cons of remote sensing using satellites. Acoustic recordings have been used for bats, insects, amphibians, and birds.

Wimmer et al. (2013) discussed the use of acoustic recordings to determine bird species richness. Acoustic recordings can be analyzed automatically by call-recognition software or manually by humans to identify species-specific calls. They noted that automatic analysis can be subject to high levels of false-positive and false-negative result. Good software for analyzing the calls is essential.

Froidevaux et al. (2014), using ultrasonic passive acoustic sampling of bats, assessed the performance of 21 acoustic sampling schemes with 3 temporal

sampling patterns and 7 sampling designs. A total of 145,433 echolocation call sequences of 16 bat species were recorded. Three microhabitats were investigated in a given forest location simultaneously throughout the entire night. Some information about ultrasonic detection systems for bats is given by Britzke et al. (2013).

5.5.12 Biodiversity Measures

There are various types of diversity measures, relating to phylogenetic, taxonomic, and functional diversity of biological assemblages (Jarzyna & Jetz, 2016). For example, Henderson (2021) reviewed species richness and α -, β -, and γ -diversity measures. He said that a useful classification is α -diversity (within a community or habitat), β -diversity (a measure of the rate and extent of change in species along a gradient from one habitat to others), and γ -diversity (richness in a range of habitats in a geographical area, e.g., an island). Mendes et al. (2008) considered some of the older popular ones such as the Shannon entropy and the Gini-Simpson index. Such measures can be defined generally, but are typically modified for different kinds of diversity studies. There is also some controversy with regard to which ones should be used.

For some further references, see Magurran and McGill (2011). There is also the seven-volume encyclopedia series (Levin, 2013) that covers a wide variety of applications. Various measures are available in the program **Estimate S** (Colwell & Elsensohn, 2014; version 9.1.0, 2019, with some limitations mentioned at <http://viceroy.eeb.uconn.edu/estimates/>). Rocchini et al. (2016) listed 12 types of alpha and beta diversity indices. Ozdemir et al. (2018) computed some indices using satellite data, while Mason et al. (2018) considered the effect of measurement error on biodiversity metrics.

At <http://chao.stat.nthu.edu.tw/wordpress/software>, there are **R**-based tools for diversity, and there are the useful measure known as a Hill number and a computer package **iNEXT**(Hsieh et al., 2016). These Hill numbers, as well as various biodiversity and similarity indices, were discussed by Chao (2005, similarity indices), Chao et al. (2014), Chao et al. (2014), Cox et al. (2017, four assessment methods), and Chao et al. (2019). Gotelli and Chao (2013) said that a comprehensive measure of species diversity should include components of both species richness and the relative abundances of the species that are present. More recently, Chao et al. (2021) integrated the three dimensions of diversity mentioned in the previous paragraph under a unified framework of Hill numbers and their generalizations. Callaghan et al. (2022) considered determining the optimal distribution of sampling effort for inference of species diversity patterns (e.g., species richness, Shannon diversity, and Simpson's diversity) using the relationship between species diversity and land cover from citizen science data.

Norris and Pollock (1998) gave estimates for four measures: Simpson's index that measures the probability that two randomly chosen individuals (chosen with replacement) will be of the same species; Shannon's diversity index that uses

information theory; Hill's diversity numbers 1 and 2 that attempt to measure the “number of abundant species” and the “number of very abundant species,” respectively; and Shannon's evenness index. Environmental DNA (eDNA) can also be a tool for considering biodiversity, as was found by Yoccoz et al. (2012) with soil sampling.

Multispecies Occupancy Models

Iknayn et al. (2014) referred to a number of diversity measures (metrics) that have been classically divided into alpha (site level), beta (turnover across multiple sites), and gamma (composite of all sites in a region) components, already referred to above. The fundamental unit of all diversity metrics is a count of species, individuals, or both. They have a helpful overview of hierarchical multispecies occupancy models (MSOMs) and multispecies abundance models (MSAMs) for estimating occupancy and abundance. They used a hierarchical model that included three levels, one each for species (i), site (j), and replicate (k), and applied data augmentation. The first level represents the true occurrence states (w_i) within the community of all partially observed and never observed species. The second level is the ecological process governing presence (z_{ij}) of partially observed species. The third level explains the detection history (y_{ijk}) from the replicated surveys. The three levels are modeled as follows:

$$\begin{aligned} w_i &\sim \text{Bernoulli}(P) \text{ superpopulation process(data augmentation)} \\ z_{ij} \mid w_i &\sim \text{Bernoulli}(w_i \psi_{ij}) \text{ (ecological process)} \\ y_{ijk} \mid z_{ij} &\sim \text{Bernoulli}(z_{ij} p_{ijk}) \text{ (observation process)} \end{aligned}$$

The second and third levels are always present in a single-species model with ψ_{ij} the probability of presence where $z_{ij} = 1$ if present and 0 otherwise and p_{ijk} the probability of detection, given present, where $y_{ijk} = 1$ if detected. Here, $w_i = 1$ if the species is present in S sites, with probability P , and let x be the total number of unobserved species for which $w_i = 1$. Then gamma diversity is represented by $S_{obs} + x$, which estimates the total number of species in the sample of sites. For estimation, they used data augmentation (Appendix A.5). Broms et al. (2016) used a similar model and considered some methods of model selection and assessment for multispecies occupancy models. They included Bayesian p -values and various diagnostic plots.

Tingley et al. (2020) compared the MSOM and the nonparametric iChao2 estimators. They found that in many cases, MSOM may provide superior estimates that also account more accurately for uncertainty. However, they have higher data requirements, can be computationally demanding and non-intuitive to fit, and contain numerous assumptions that may not be met by the data. On the other hand, iChao2 estimators are simple to calculate, are flexible for many situations, and have relatively fewer assumptions. Both methods can show strong bias when average occupancy is very low. Using simulations, both methods performed well across a

broad range of community types. Results were similar regardless of the number of species (20 or 125) and number of sites surveyed (30 or 60).

Multispecies Abundance Model

The MSAM is similar to the MSOM except that the indicator variable y_{ijk} is replaced by n_{ijk} , an observed count for species i at site j and replicate k . If N_{ij} is the number of species i at site j , we have the following probability distributions:

$$\begin{aligned} w_i &\sim \text{Bernoulli}(P) \\ N_{ij} \mid w_i &\sim \text{Poisson}(w_i \lambda_{ij}) \\ n_{ijk} \mid N_{ij} &\sim \text{Binomial}(N_{ij}, p_{ijk}) \end{aligned}$$

In the above, the Poisson distribution could be replaced by some other distribution such as a zero-inflated Poisson or a negative binomial.

Dorazio et al. (2011) developed the following model for ant species using pitfall traps to capture the ants. Let y_{ik} denote the number of traps located at site k that contained the i th of S_{obs} distinct species captured in the entire sample of $S = 22$ sites. At each site, 25 pitfall traps were deployed during each of 2 sampling periods (early- and late-season replicates) so that the total number of replicate observations per site was constant ($J_k = 50$). Here, y_{ik} can take the values $\{0, 1, \dots, J_k\}$. It is assumed that $J_k > 1$ for some, hopefully all, sample sites because information from within-site replicates allows both occurrence and detection probabilities to be estimated for each species. In the absence of this replication, these two parameters are confounded. The data forms an $S_{obs} \times S$ matrix \mathbf{Y}_{obs} , which is then analyzed using parameter-expanded data augmentation. They then provided estimates of several diversity measures such as the gamma, alpha, and beta diversities. If ψ_{ik} , the probability of occurrence of species i at sample site k , $\text{logit}(\psi_{ik})$ is expressed as a linear regression in terms of covariates. Here, again $w_i = 1$ if species i is a member of the community so that $w_i \sim \text{Bernoulli}(\Omega)$ and $z_{ik} \mid w_i \sim \text{Bernoulli}(w_i \psi_{ik})$. The species captured are then modeled as

$$y_{ik} \mid z_{ik} \sim \text{Binomial}(J_k, z_{ik} p_{ik}),$$

where $z_{ik} = 1$ if present and 0 otherwise and p_{ik} is the conditional probability of capture of species i at site k , given present ($z_{ik} = 1$).

Chao et al. (2020), considered the question of how to compare diversities among assemblages of species. The main problem is that the estimation of species richness will depend on the sample completeness of a biological survey. The authors developed a novel class of measures to quantify sample completeness parameterized by an order $q \geq 0$ to control for sensitivity to species relative abundances. For example, when $q = 0$, species abundances are disregarded, and our measure reduces to the conventional measure of completeness, that is, the ratio of the observed species richness to the true richness (observed plus undetected). When $q = 1$, the measure reduces to the sample coverage (the proportion of the total number of

individuals in the entire assemblage that belongs to detected species). To compare diversity across multiple assemblages, they proposed an integrated approach based on the framework of Hill numbers to assess (a) the sample completeness profile, (b) asymptotic diversity estimates to infer true diversities of entire assemblages, (c) non-asymptotic standardization via rarefaction and extrapolation, and (d) an evenness profile.

5.5.13 Environmental Links

Species richness is closely linked to the environment, and selecting environmental species traits can provide an important role using, for example, generalized linear mixed models (Jamil et al., 2013). Newbold et al. (2009) compared some methods for generating climate-based estimates of spatial patterns of butterfly and mammal species richness, especially in a country like Egypt with a typical arid environment for flora and fauna. Records were taken from a variety of sources, and they began with two methods.

For the first, they used initial distribution models for the 40 butterfly species and 68 mammal species with at least 8 records of occurrence using **MAXENT** and explaining how it fitted a distribution. They then converted each model output into a binary prediction of presence or absence around a threshold value and then summed individual predictions across all species to give an estimate of species richness. A species was considered present in a cell if it had been recorded at least once. Species richness was modeled using two generalized linear models (GLMs) with Poisson and negative binomial errors. They applied **R** along with the log link using the same independent variables as in the species distribution models.

Their second method of modeling species richness was to model observed species richness values directly. That part of the study was concerned with the total number of species recorded in each cell (site) rather than individual records of species. Habitat and climate change variables were recorded, and, because of the large number of the latter along with elevation variables, four principal components were used (Seber, 1984: Section 5.2). In conclusion, the authors found that habitat and climate variables both had a significant effect on butterfly and mammal distributions, and on patterns of species richness, at least at the local scale at which the study was carried out. Summing the individual distribution models produced the best estimates of species richness. Also, “seemingly accurate estimates of species richness can be made using patchy, incomplete data, allowing us to predict the species richness of sites that have not been surveyed.”

Temporary Emigration

Species richness, when there is temporary emigration, was discussed by Kéry et al. (2009) using adult data from the Swiss butterfly monitoring program. This problem of nonclosure is common in communities of short-lived organisms, such as insects or plants, and even in relatively long-lived vertebrate communities when

there are migratory or hibernating species. The method consisted of a robust design model, using up to seven primary units and two secondary units (transects or repeat observations) within each primary unit. The authors followed the usual route of considering the two components, namely, presence spatially and in time (a process model) and detectability given presence. Using their notation, the model is as follows.

Let $y_{ijk} \in \{0, 1, 2\}$ be the number of times (in two secondary sampling periods or replicates) in which at least one adult of species k was observed in primary sampling period j ($j = 1, \dots, 7$) and transect i (they used 13), with p_{ijk} the probability of detection given present. When two secondary samples were taken at each period and transect, y_{ijk} is conditionally binomial, namely,

$$y_{ijk} \mid z_{ijk} \sim \text{Binomial}(2, p_{ijk}z_{ijk}),$$

where $z_{ijk} = 1$ if species k is present as an adult (able to be detected) during period j in transect i and $z_{ijk} = 0$ if it is not. A logistic normal regression model with covariates was then used for p_{ijk} . It was assumed that

$$z_{ijk} \sim \text{Bernoulli}(\psi_{ijk}),$$

where ψ_{ijk} is 1 minus the temporary emigration probability, namely, the probability that species k is available as an adult, or exposed to sampling, in period j and transect i . A quadratic regression model with a date covariate was fitted to $\text{logit}(\psi_{ijk})$. Species abundances were estimated. The authors carried out model fitting, data augmentation, and Bayesian analysis with prior distributions for various parameters. The reader is referred to their paper for further details.

Random temporary emigration was considered by Chandler et al. (2011) using a superpopulation hierarchical model and an integrated likelihood. Because their model only assumed that plots are closed during the sampling occasion and not among sampling occasions, an estimate of population density was provided.

5.5.14 Species-Area Relationships (SARS)

The long established power form of the species-area relationship (SAR) can be used to set baseline targets for species conservation using biodiversity survey data (Desmet & Cowling, 2004). SAR is modeled for each important habitat type, which is then used to estimate the proportion of habitat area required to represent a user-specified percentage of species. Many mathematical models can and have been used to describe the SAR (Lomolino, 2000), but the power relationship or Arrhenius equation is probably the most popular in the literature, namely,

$$S = kA^c,$$

where S is species richness (number of species), A is the area, k is a scaling factor that relates to the sample size used, and c typically ranges from 0.10 to 0.50. If S and A are replaced with proportion of species and proportion of area (S') and (A'), respectively, there is no need to estimate the constant k , and the equation can be transformed to a straight line without using k , namely,

$$\log(S') = c \log(A'), \text{ or } \log(A') = \log(S')/c.$$

We can then use a log-log plot and estimate the slope $1/c$ (cf., Kärenyi et al., 2016). The estimated c value can then be used to calculate the proportion of area required to represent a given species percentage and then multiply this up for the total population area. Here, various estimators can be used for species richness, and some were compared by Metcalfe et al. (2013). Rondinini and Chiozza (2010) gave a review of quantitative methods for defining percentage area targets for habitat types in conservation planning.

Lomolino (2000) noted that two critical shortcomings of such models are that they lack an asymptote (for the larger ecosystems) and that they ignore the possibility of what has been termed the small island effect. First, because isolated faunas are ultimately derived from a limited pool of species, the species-area relationship should asymptotically approach or level off at that maximum value of richness. Second, richness may be independent of island area for collections of relatively small islands. Examples are given by the authors.

Matias et al. (2014) commented that the use of the SAR and EAR (endemics-area relationship) for forecasting species loss resulting from habitat loss strongly depends on the ecological context. Further shortcomings of the SAR were mentioned by Kitzes et al. (2014) who proposed two new metrics, the extinction-area relationship and probabilistic species-area relationship, to address the shortcomings.

Yamaura et al. (2016) looked particularly at the SAR for plants and birds since it seems to be generally known that species richness increases with area and applied various models. For instance, we have the following Bayesian approach. Let Z_{ij} be the number of different species i in patch (site) j of area A_j , and assume that Z_{ij} is distributed as Poisson with mean λ_{ij} . To relate species abundance to area, the following model was used:

$$\log(\lambda_{ij}) = \beta_{0i} + \beta_{1i} \log(A_j) + \boldsymbol{\beta}'_i \mathbf{x}_j + b_{ij},$$

where \mathbf{x}_j are patch-specific covariate(s) and b_{ij} are random site effects, assumed to be normally distributed as $N(0, \sigma_b^2)$. The negative binomial distribution rather than random site effects could be used to account for extra-Poisson variation in abundance. It was further assumed that the two β parameters have normal distributions shared by all species (including rare species), namely, $\beta_{hi} \sim N(\mu_{\beta_h}, \sigma_{\beta_h}^2)$, $h = 0, 1$.

Let N_{ij} be the number of species i in patch (site) j exposed to sampling and ϕ_i be the sampled area divided by the patch area. Then we can model

$$N_{ij} \sim \text{Binomial}(Z_{ij}, \phi_j).$$

To consider imperfect detection, we can use different detection models depending on the survey methods. With simple counts, we can use the binomial model

$$y_{ijt} \sim \text{Binomial}(N_{ij}, p_i),$$

where y_{ijt} is the number of individuals of species i in patch j detected on visit t ($t = 1, \dots, T$) and p_i is the probability of detecting an individual of species i , conditional on presence.

Alternatively, the encounter frequencies, say, y_{ijH} ($H = 0, \dots, T$), can be modeled as a multinomial distribution with index N_{ij} and probabilities π_{iH} , the probability of species i being encountered H times (assumed independent of site number). Here, π_{i0} is the probability of no encounters and is one minus the sum of the other probabilities. Since the vector of all the counts y_{ijH} is a sufficient statistic for a capture-recapture model to estimate the number of individuals in a single species, the so-called heterogeneity model M_h , we can consider our model in this light where the different species lead to heterogeneity among individuals. Therefore, the previous multinomial distribution can be remodeled as a binomial distribution for the number of individuals n_{ij} encountered as

$$n_{ij} \sim \text{Binomial}(N_{ij}, \pi_{i\cdot}),$$

where $\pi_{i\cdot} = \sum_{H=1}^K \pi_{iH}$. Conditioning on the n_{ij} , we have that the y_{ijH} have a multinomial distribution with index n_{ij} and probabilities $\pi_{iH}/\pi_{i\cdot}$. Further models are given by the authors.

Karenyi et al. (2016) noted that the SAR method of setting habitat-specific conservation targets assumes that detectability of species is perfect, but this assumption is often violated. They compared SAR-based target setting with multispecies occupancy models (MSOMs, in Sect. 5.5.12) that incorporate incomplete detection to better represent communities with a large proportion of rare species. Also, they mentioned that although there is general agreement that bootstrap estimates are the preferred estimates for calculating conservation targets, those estimates increase with increasing sample size even at very large sample sizes. This means that estimators based on that method are not stable; they compared several methods.

5.6 Species Misidentification

A problem associated with species is errors in identification, in particular false-positive errors where a specimen is falsely declared to be a particular species. Shea et al. (2011) considered this problem using as an example freshwater mussel species (Bivalvia: Unionidae) identification. To evaluate the prevalence of and potential reasons for species misidentification, they conducted a laboratory-based identification exercise and quantified the relationships between mussel species characteristics,

observer experience, and misidentification rate. They focused their analysis on identifying how species characteristics and observer experience contributed to false-positive errors and fitted hierarchical logistic regression models to the species identification data to account for dependence among repeated species identifications by observers. Here, λ_{ijk} takes the value 1 if the i th observation for the j th species by observer k is misidentified and 0 otherwise. They used the model

$$\text{logit}(\lambda_{ijk}) = \eta_{ijk} = \beta_0 + u_{0k} + \beta_q x_{qijk} + u_{1j},$$

where β_0 is the intercept and β_q is the effect of observer k and species j with specific covariates x_{qijk} on misidentification. The random components u_{0k} and u_{1j} , assumed to be normally distributed with zero means, represented unique effects associated with observers and species, respectively, which were unexplained by observer experience and species-specific characteristics. To accommodate the complex model structure, they used Markov chain Monte Carlo as implemented in WinBUGS software, version 1.4 (Spiegelhalter et al., 2006) to fit candidate hierarchical logistic regression models. A total of 1217 species identifications were made by the 18 observers. This number was fewer than the total number of possible identifications (1332) because, in some cases, observers did not make a species identification. Of the 1217 species identifications, 323 were incorrect.

5.7 Opportunistic Surveys

In Sect. 4.16, we discussed the use of opportunistic surveys and citizen science with respect to occupancy modeling and for single species. They can also be used for multiple species. The use of volunteers in biodiversity monitoring schemes overcomes constraints caused by a lack of funding for professional researchers and increases public perception and involvement (Sewell et al., 2010: amphibian survey.) An opportunistic survey can use “citizen science” to collect a massive amount of auxiliary data at negligible cost and has been shown to be a useful supplement to traditional analyses. For example, Higa et al. (2015) found that “citizen data with spatially biased sampling effort can be appropriately utilized for large-scale biodiversity distribution modeling with the use of occupancy models, which encourages data collection by citizen scientists.”

Giraud et al. (2016) made use of opportunistic data along with planned schemes characterized by a known sampling effort for monitoring relative abundances of species using multispecies and multi-site data. Here, there are I species at J sites and two data sets indexed by k , where data set $k = 0$ has been collected using a standardized protocol and the data set $k = 1$ is of an opportunistic nature. Using a thinned Poisson model, it was assumed that x_{ijk} , the count gathering all the records of individuals of species i from all visits to site j for data set k , has a Poisson distribution. We assume that an individual is only counted once during a single visit, though it can be counted several times due to the possible multiple visits to a

site j for a data set k . In particular, we may have x_{ijk} larger than the number N_{ij} of individuals of species i at site j . The authors gave mathematical details to show that there is a theoretical gain in combining the two data sets. As an application, data from 34 species was used, and supplementary material is given in web appendices of the authors' paper.

5.8 Summary

This chapter is essentially a follow-on from the previous chapter, but with an additional level of complexity. One can simply add an additional subscript to allow for species type and allow for a community approach. Species distribution models for a single species are considered for which there are a large literature and extensive reviews.

Environment and species modeling is discussed in several places, as well as various Bayesian models. Community species play an important role, leading naturally into the consideration of various multivariate species models, including occupancy and spatial models, and time-to-detection models. This topic is closely related to determining species interactions, including the use of Markov networks and state-space modeling. Sometimes, absence information can be used.

An important topic is the determination of the number of species present in a region, called "species richness." This can be estimated when there are replicate visits. Parametric and nonparametric methods are given for estimating species richness using both incidence and abundance and are discussed in detail. Choosing the sample size for both types of data is considered.

A miscellaneous set of topics follows, namely, the use of environmental DNA, biodiversity with its many measures, the problem of species misidentification, and opportunistic and citizen science surveys and their design.

Chapter 6

Closest Distance and Nearest Neighbor Methods



6.1 Description

In Chap. 2 we considered plot methods for estimating population sizes for a wide variety of species, which can be time-consuming in establishing, for example, quadrats. In this short chapter, we look at relatively immobile or stationary populations (particularly trees). We note that the theory of this chapter is not usually referred to as distance sampling even though distances are involved. That term is reserved to describe numbers of animals detected from some fixed point and involves two topics, point and line transect sampling (Buckland et al., 2015), discussed in Chaps. 7 and 8. Computer packages for handling distance data are **Distance** (cf. Thomas et al., 2010) with goodness-of-fit tests and the package **R**. We now consider four distinct distance methods that are available instead of just counts on plots (usually circular).

The first method that we consider, which we call the closest-individual , involves choosing a point at random in an area and then measuring the distance to the nearest object (or r th), for example, a tree if we are interested in estimating tree density.

The second method, which we shall call the nearest-neighbor , involves choosing an object at random and measuring the distance to the nearest neighboring object, or more generally the r th nearest neighboring object. Choosing an object at random requires being able to identify all the objects. This would lead to having the same distance distribution as the closest-individual distance only if the objects were randomly distributed, for then choosing an object at random is equivalent to choosing a point at random. A third related method involves choosing a point at random, then choosing the nearest object from which to measure the nearest neighbor.

A fourth method involves choosing a point at random and measuring the distances to the r closest objects, where the circular “plot” involved and measured

out to the r th object will be variable. The distance methods proposed below can be less affected by terrain variation than plot methods.

For a discussion of the earlier developments of the theory of the closest-individual distance and nearest-neighbor distance that are still relevant see Seber (1982, sections 2.1.4 and 12.5). A more recent and extensive review of past methods is given by Cogbill et al. (2018) who found 2183 publications on the topic, and evaluated the accuracy of sixteen common PDEs (plotless density estimators) by applying them to simulated patterns. The two types of distance measurements have also been used either singly or together by various authors as tests of randomness, though some are concerned more with small-scaled patterns, and the two types have also been combined to estimate population density (cf. Seber, 1982: 472–475). Shen et al. (2020) mentioned that, over time, there have been at least 20 distance-based methods being proposed in the literature to improve the performance of density estimation or to reduce sampling effort. They also listed formulae for five of the most conventional ones in their supplementary material.

In the past, most of the estimates depended on the assumption of a random population distribution, so that tests of randomness need to be carried out. A common method of choosing a sample point is to choose a quadrat at random and then locate a point in the plot using a pair of random numbers to give coordinates. The question of edge effect discussed in Sect. 2.4 needs to be considered as selected points need to be away from the boundary. Below we will firstly demonstrate just the general theory of the methods by using the negative binomial distribution as an example, as the assumption of a random distribution of the population may not apply. For example, a study done by McGarvey et al. (2005) of the shellfish abalone on the sea floor found a high degree of clustering, which led to biased estimates.

6.2 Closest-Individual Distance

Let X be the distance from a randomly chosen point to the nearest individual. Suppose that the individuals are distributed according to a negative binomial distribution

$$h(u) = \frac{k(k+1)\cdots(k+u-1)}{u!} P^u Q^{-u-k}, \quad u = 0, 1, \dots, Q = 1 + P, k > 0,$$

where $h(u)$ is the probability of finding u individuals in a randomly chosen plot, with a random center, of area πx^2 , and $h(0) = Q^{-k}$. If D is the population density, the number of individuals expected to be in an area of πx^2 is $D\pi x^2 = E[u]$ ($= kP$) so that $P = D\pi x^2/k$. If we assume that k , the “heterogeneity” parameter is

independent of X (which may not be true), and then the distribution function of X is

$$\begin{aligned} F(x) &= \Pr(X \leq x) \\ &= 1 - \Pr(\text{no individuals in plot of area } \pi x^2) \\ &= 1 - h(0) \\ &= 1 - (1 + P)^{-k} \\ &= 1 - \left(1 + \frac{D\pi x^2}{k}\right)^{-k}, \end{aligned}$$

and

$$f(x) = F'(x) = 2D\pi x \left(1 + \frac{D\pi x^2}{k}\right)^{-k-1}.$$

It is helpful to make the transformation $Y = \pi X^2$, which has density function

$$g(y) = D \left(1 + \frac{Dy}{k}\right)^{-k-1}.$$

Given a sample y_1, \dots, y_s from s sample points, we can obtain maximum likelihood estimates of D and k from the likelihood function $\prod_{i=1}^s g(y_i)$.

6.2.1 Poisson Model

For the Poisson model, we let $k \rightarrow \infty$ in the previous section to give us (Moore, 1954)

$$f(x) = 2D\pi x \exp(-D\pi x^2) \quad \text{and} \quad g(y) = D \exp(-Dy).$$

The maximum likelihood estimate of D corrected for bias (s replaced by $s - 1$) is the unbiased estimator

$$\widehat{D} = (s - 1) / \sum_{i=1}^s \pi x_i^2,$$

with variance $D^2/(s-2)$. Unfortunately \widehat{D} is not robust and is sensitive to variations in D and will be a poor estimator for most practical (non-random) situations. It is noted that for both the negative binomial and Poisson distributions, we have $D = g(0)$, which is true for a number of different distributions (Patil et al., 1979). This suggests considering a nonparametric estimate of $g(0)$, and the previous authors

considered using the order statistics $y_{(1)} < y_{(2)} \dots < y_{(s)}$ to do this. They proposed using $\tilde{D} = \hat{g}_s(0) = k(s)/(sy_{(k)})$, where $k(s)$ is a sequence of real numbers such that $k(s) \rightarrow \infty$, $k(s)/s \rightarrow 0$ as $s \rightarrow 0$, and $k = [k(s)]$, the integral part of $k(s)$. It was found that the above method underestimated density in aggregated patterns and overestimated density in regular patterns. The above theory was also developed for the case where searching is restricted to a distance R . This can lead to edge effects, which are discussed in detail by Seber (1982, sections 2.1.4 and 12.5) and considered below.

A more recent method for handling the edge effect applied particularly to trees is called the “mirage point”. Here, the perpendicular distance d between the boundary and the sample point or plot center is first measured. The mirage point is then established at a distance d outside the tract boundary on a line perpendicular to the boundary running through the original sample point or plot center. Gregoire (1982) showed that the mirage method is design-unbiased for the fixed-radius plot and point sampling. Lynch (2012) described some of the studies using this and other methods such as the walkthrough method, the pull-back method (pull the sample back from the edge so that it will not overlap the boundary), the buffer method, the probability sampling method of Grosenbaugh (1958), and the measure-probability-of-inclusion method (cf. Gregoire and Scott, 1990, and Gregoire & Valentine, 2008: section 10.7). Lynch and Gove (2014) also proved the unbiasedness of a generalized mirage method for critical height sampling and related methods.

6.3 The r th Closest Distance

In some stationary populations such as trees, the above theory can be extended to deal with the distance $X_{(r)}$ to the r th nearest object from a randomly chosen point (rCD). For a *random* distribution of objects

$$\begin{aligned} F_r(x) &= \Pr(X_{(r)} \leq x) \\ &= \Pr(r \text{ or more objects in a circular plot of area } \pi x^2) \\ &= \sum_{i=r}^{\infty} \exp(-D\pi x^2) \frac{(D\pi x^2)^i}{i!} \end{aligned}$$

and (Thompson, 1956)

$$f_r(x) = F'_r(x) = \frac{2(D\pi)^r}{(r-1)!} x^{2r-1} \exp(-D\pi x^2).$$

If $Y = \pi X_{(r)}^2$, then Y had probability density function

$$g_r(y) = D^r y^{r-1} \exp(-Dy)/(r-1)!.$$

To estimate D , suppose s sample points are chosen giving distances y_i ($i = 1, \dots, s$). If $y_+ = \sum_{i=1}^s y_i$, then an unbiased estimator of D is Moore's (1954) estimate

$$\widehat{D} = (sr - 1)/y_+ \quad (6.1)$$

with variance

$$\text{var}[\widehat{D}] = D^2/(sr - 2).$$

Also $2Dy_i$ is χ^2_{2r} so that $2Dy_+$ is χ^2_{2rs} , which can be used for constructing confidence intervals. Eberhardt (1967) suggested the unbiased estimate

$$\widetilde{D} = (r - 1) \sum_{i=1}^s (y_i^{-1}/s) = \frac{(r - 1)}{s} \sum_{i=1}^s \frac{1}{\pi x_i^2} = \sum_{i=1}^s \widetilde{D}_i/s, \text{ say} \quad (6.2)$$

with, for $r \geq 3$,

$$\text{var}[\widetilde{D}] = \text{var}[\widetilde{D}_i]/s = D^2/[s(r - 2)].$$

Although greater than $\text{var}[\widehat{D}]$ above, it has an unbiased variance estimate

$$v(\widetilde{D}) = \frac{\sum_{i=1}^s (\widetilde{D}_i - \widetilde{D})^2}{s(s - 1)}.$$

Eberhardt (1967) (and given by Seber, 1982: 46–50) showed that \widetilde{D} (which we refer to as Eberhardt's estimate) is still unbiased for both the binomial and negative binomial distributions. However, as it is not design-unbiased, it can be biased for other spatial distributions.

With the negative binomial distribution, $h(u)$, given above is the probability of finding u individuals in a randomly chosen plot of area πx^2 . If $X_{(r)}$ is the distance to the r th closest individual, then

$$f_r(x) = \frac{2(D\pi)^r x^{2r-1}}{k^r} \frac{\Gamma(r+k)}{\Gamma(r)\Gamma(k)} \left(1 + \frac{D\pi x^2}{k}\right)^{-r-k}.$$

If $Y = \pi X_{(r)}^2$, then arguing as above, Y has probability density function

$$g_r(y) = \frac{\Gamma(r+k)(Dy)^{r-1} k^k D}{\Gamma(k)(r-1)!(k+Dy)^{r+k}},$$

which is related to the Beta distribution. We can then use this distribution to show that \tilde{D} is unbiased, and for $r \geq 3$,

$$\text{var}[\tilde{D}] = \frac{D^2(1 + (r - 1)/k)}{s(r - 2)}.$$

For the Poisson distribution

$$g_r(y) = D^r y^{r-1} \exp(-Dy)/(r - 1)!.$$

This was derived independently by Moore (1954), Morisita (1954), and Thompson (1956) and generalized to any number of dimensions by Dacey (1963) for a random distribution of objects, which can be used for three-dimensional distances.

6.3.1 The r th Nearest Neighbor

If the population distribution is random., i.e., Poisson, then choosing a point at random is the same as choosing an object at random, which leads to the same theory as before. However, if the population distribution is not random but negative binomial, then the r th nearest neighbor distribution (individual to r th nearest individual) is given by Gao (2013), namely,

$$g_r(x) = \frac{2(D\pi)^r x^{2r-1}}{k^r} \frac{\Gamma(r + k + 1)}{\Gamma(r)\Gamma(k + 1)} \left(1 + \frac{D\pi x^2}{k}\right)^{-r-k-1}.$$

We now look at nonrandom distributions in general.

6.3.2 Nonrandom Distributions

We previously considered used the negative-binomial distributions as an example of sampling from a clustered population. We now need to consider more general nonrandom distributions, and we refer first to Morisita's (1957) model, which endeavored to deal with such distributions. Using the same notation as above to estimate density D , s objects or (focal) points are sampled at random (i.e., sampled independently with equal probability). Centered on each focal point, q equal angle sectors are divided up (e.g., for a circle or a rectangle area we can have $q = 4$ quadrants). Let $x_{rij,pot}$ denote the distance from the i th ($i = 1, \dots, n$) focal point to its r th closest object in the j th sector ($j = 1, \dots, q$). Morisita assumed that the population distribution in the small area is random even though the overall

population is nonrandom and obtained the following two density estimators:

$$\begin{aligned}\widehat{D}_{1,ptot} &= \frac{r-1}{\pi s} \sum_{i=1}^s \sum_{j=1}^q \frac{1}{x_{rij,ptot}^2} \\ \widehat{D}_{2,ptot} &= \frac{rq-1}{\pi s} \sum_{i=1}^s \frac{q}{\sum_{j=1}^q x_{rij,ptot}^2},\end{aligned}$$

with respective variances

$$\begin{aligned}\text{var}[\widehat{D}_{1,ptot}] &= \frac{D^2}{sq(r-2)}, \\ \text{var}[\widehat{D}_{2,ptot}] &= \frac{D^2}{s(rq-2)}.\end{aligned}$$

Cogbill et al. (2018) tested and compared the performance of these two estimators against other existing plotless methods currently available and found the above estimates behaved best for nonrandom populations. These estimators were designed more for “locally random but globally heterogeneous” populations (Diggle, 1977).

Diggle (1975) introduced two composite estimators to try and correct for bias for nonrandom distributions. The first was modified by Shen et al. (2020) to correct for bias with random distributions, namely,

$$\widehat{D}_c = \frac{q \Gamma(s_1 qr) \Gamma(s_2 qr)}{\pi \Gamma(s_1 qr - 1/2) \Gamma(s_2 qr - 1/2) \sqrt{R_{1,ptot} R_{2,tott}}},$$

where

$$\begin{aligned}R_{1,ptot} &= \sum_{i=1}^{s_1} \sum_{j=1}^q x_{rij,ptot}^2 \\ R_{2,tott} &= \sum_{i=1}^{s_2} \sum_{j=1}^q x_{rij,tott}^2.\end{aligned}$$

Here, $x_{rij,ptot}$ and $x_{rij,tott}$ are distances from point to object and object to object, respectively, and s_1 and s_2 are the respective sample sizes. If $s_1 = s_2 = s/2$,

$$\text{var}[\widehat{D}_c] = D^2 \left(\frac{\Gamma^4(\frac{1}{2}rsq)}{(\frac{1}{2}rsq - 1)^2 \Gamma^4(\frac{1}{2}rsq - \frac{1}{2})} - 1 \right).$$

A key assumption underlying the above theory is that the population is locally random, that is random in each sector.

The authors introduced two other estimators by assuming that the population has a negative binomial distribution *in each sector* and sample distances are independent. They gave two moment estimators

$$\widehat{D}_{neg.ptot} = \frac{q(2r - 1) \sum_{i=1}^s \sum_{j=1}^q x_{rij,ptot}^{-1}}{\pi \sum_{i=1}^s \sum_{j=1}^q x_{rij,ptot}} - \frac{rsq^2}{\pi \sum_{i=1}^s \sum_{j=1}^q x_{rij,ptot}}$$

$$\widehat{D}_{neg.ttot} = \frac{rsq^2}{\pi \sum_{i=1}^s \sum_{j=1}^q x_{rij,ttot}}.$$

Confidence intervals are given in their supplementary material. Compared with other traditional estimators, the authors found that $\widehat{D}_{neg.ptot}$ performed best and $\widehat{D}_{neg.ttot}$ perfomed worst for simulated and real data. However, with the former, there remains a noticeable consistent underestimation for highly aggregated populations.

6.4 Methods for Trees

As trees are a good application of distance methods, we consider further methods for trees with a main focus on population size. Although the focus of this chapter is just on distance methods, it should be noted how technology has helped in the development of this topic in forestry such as GPS location methods, airborne laser scanning, digital aerial photogrammetry, and very high spatial resolution satellite optical imagery (White et al., 2016, especially issue 6). However, the density estimates below apply generally and not just for trees. The subject is somewhat a minefield as there are many estimators. An early helpful resource on tree sampling and related topics is given by Gregoire and Valentine (2008).

A variety of sampling methods are used for estimating a wide variety of characteristics for trees such as tree density, basal area per hectare, crown diameter, and so forth. Marquardt et al. (2010) provided a good review of such characteristics and compared a wide range of sample plot methods for riparian forests of western Oregon. Riparian areas rank among the most complex, variable, and dynamic terrestrial habitats in the world. They found that rectangular strip plots outperformed all other plot shapes and the circular-shaped plots performed better than the alternatives that used horizontal line sampling.

6.4.1 Plot Methods

We begin with plot methods as they are often compared to the distance methods to be considered. Circular fixed plot sampling (FPS) or fixed-radius circular plot

sampling, where the number of trees counted on a circular plot of fixed radius, is one of the oldest methods of forest sampling and is still commonly used throughout much of the world. When quadrat plot sampling is difficult or too costly (e.g., in low density populations or mountain areas), distance sampling is favored. In natural populations characterized by an irregular, possibly clustered, distribution of trees, the precision of a density estimate from distance sampling can be better than the precision obtained with fixed area plot sampling (Lurdes et al., 2017). One disadvantage of the method is that a fixed plot size may not work so well with low density populations, whereas a variable-radius method allows one to use a similar effort with varying densities. Adaptive sampling with circular plots lends itself particularly to sampling sparse but clustered clumps of trees (Yang et al., 2016).

Variable-radius plot sampling (also called angular-count or is the first plot method, considered briefly. It relies on the relationship between the diameter at breast height of observed trees and the distance between those trees and an observer who stands at the plot center (Grosenbaugh, 1952). Haxtema et al. (2012) noted that “HPS has historically been preferred by many forest inventory professionals in the Pacific Northwest because, when a prism or Relascope is employed, horizontal distance measurements are unnecessary except to check the in/out status of borderline trees—a great asset when working in steep and brushy terrain.” As well as tree density, accurate and efficient estimation of stem density and basal area of trees on an area are also crucial to the success of active restoration or management programs in forests adjacent to headwater streams.

Forest managers most often use a prism to estimate stand basal area for forest management purposes. Prism plots are designed to provide an estimate of basal area, but stem density can also be calculated if sample tree diameters are known. Conversely, any estimator for stem density can also provide an estimate of basal area, again if diameters of sample trees are measured (Thompson et al., 2006). Packard and Radtke (2007) gave a method for combining both fixed FPS and HPS methods to estimate total stand volume.

More recently, the above methods have been modified through the inclusion of remotely sensed auxiliary information, in particular light detection and ranging-derived (LiDAR) variables. High-resolution, discrete-return LiDAR data provide spatially explicit information well correlated to the three-dimensional structure of the forest canopy (Melville et al., 2015). Melville and Stone (2016) introduced a new centroid method using distance methods. Hsu et al. (2020) noted that such data may lead to forest inventory estimates that are generally unbiased at landscape levels but may be biased locally. They provided a method to correct local bias using ground-based double sampling with ratio estimation along with covariates to estimate a correction or calibration ratio.

6.4.2 Distance Sampling for Trees

For a short history of this topic relating to forestry, see Kleinn and Vilčko (2006a). One distance method is to measure the distances out to the n nearest trees so that the plot size varies, usually referred as an n -tree distance-sampling (NTDS) method. One advantage of this approach is that the number of trees counted is constant so that the inventory costs are nearly independent of tree density. Lessard et al. (1994) observed that the method proved to be cost-effective compared to point and plot sampling. Lessard et al. (2002) found that if the trees are randomly dispersed, their n -tree method was at least as precise as the sample plot method if the fixed-radius plot size is less than the ratio of $(n - 2)$ to the expected density. They also found a similar result holds for a clustered forest, where the ratio is multiplied by a factor involving a constant of heterogeneity. Although the number of trees remains fixed for each sample, the distribution of the size of the circular plot varies with the tree spatial pattern.

A variation of distance sampling where the plot size is variable is to measure the distances from a point out to the r th-closest tree (rCD or r th closest distance). The distance from the plot center to the center of the r th tree forms the radius of a circular plot. The r -tree, the T-square (Besag & Gleaves, 1973; Seber, 1982: 475–477), and point-centered (Pollard, 1971; Seber, 1982: 44) methods were compared with benchmark quadrat sampling in a review by Lurdes et al. (2017). Six 900 m² plots were established at *Pittosporum undulatum* stands in Sao Miguel Island. The authors tested a total of 15 estimators using a data resampling approach and recommended the use of the T-Square method.

Use of Sectors

We mentioned above that in an early paper, Morisita (1957) suggested dividing up a circular plot area a into q equiangular sectors such that the distribution of individuals in each sector is either random or fairly uniform. Let N_j be the number of trees in sector j with area a_j ($j = 1, \dots, q$). Here $N = \sum_{j=1}^q N_j$, $A = \sum_{j=1}^q a_j$, $D_j = N_j/a_j$, and $D = N/A$. If a sample point is chosen at random, the probability that it falls in a_j is a_j/A . Let x_{rij} be the distance from sample point i to the r th nearest tree in sector j . For a random distribution in the population, we have, by extending \tilde{D} of (6.2) to q sectors, the estimate

$$\tilde{D}_q = \frac{q(rsq - 1)}{\sum_{j=1}^q \sum_{i=1}^s \pi x_{rij}^2}$$

with variance

$$\text{var}[\tilde{D}_q] = \frac{D^2}{rsq - 2}.$$

For further details and discussion, see Morisita (1957) with $m = D/\pi$. Since, for large n (say $n > 50$), $(2\chi_n^2)^{1/2} - (2n - 1)^{1/2}$ is approximately $N(0, 1)$, we have

(Shen et al., 2020, supplementary materials)

$$\left(4\pi Dq^{-1} \sum_{i=1}^s \sum_{j=1}^q x_{rij}^2\right)^{1/2} - (4rsq - 1)^{1/2} \approx N(0, 1),$$

which can be used to construct confidence intervals for D .

Use of rth Closest Distance, rCD

Prodan (1968) introduced a density estimate that takes the general form for a single plot and r th neighbor (he used $r = 6$), namely,

$$\widehat{D}_r = \frac{10,000}{\pi x_r^2} (r - 0.5)$$

trees per hectare, where x_r is the distance to the r th nearest tree. This can be averaged over the sample plots as above for Eberhardt's method, Eq. (6.2). Unfortunately Prodán's density estimator is generally biased as it uses incorrect inclusion probabilities. However, Lynch and Rusydi (1999) compared Prodán's and Eberhardt's methods for Indonesian teak plantations and, according to their criterion, found that 5-tree distance sampling using Prodán's estimator was the most efficient of the techniques compared that did not show bias for these data. They believed this was probably due to the uniform spatial distribution characteristic of plantations.

Kleinn and Vilčko (2006a) noted that because the plot size at each sample point is computed as the smallest that could contain r trees, this uncorrected estimator will systematically overestimate the value of any attribute on a per-5acre basis. They introduced an rCD empirical method that used the average of the distances to the r th and $(r + 1)$ th trees (mean distance estimator), namely $x = (x_r + x_{r+1})/2$, as well as one that used a radius $x = \sqrt{[(x_r^2 + x_{r+1}^2)/2]}$ (mean area estimator). They compared their methods with those of Eberhardt and Prodán using real and artificially generated tree maps and focused on basal area and number of stems. It was found that for all but the uniform maps, fixed area plots have smallest errors, where the difference from the other plot designs was particularly large for small values of r .

For the NTDS methods, Eberhardt's method of was superior in terms of bias for random and clustered patterns for all $r = 1, \dots, 12$, but not for uniform patterns, where it had a negative bias. The methods of Kleinn and Vilčko (2006a) had the smallest bias for uniform maps. Their further conclusions were as follows: (1) Fixed area plots produced smaller errors than the r-tree estimators for estimation of both basal area and number of stems for all but the maps with uniform patterns, and (2) the performance of r-tree sampling does depend considerably on the spatial pattern of the population, where marked differences are observed over the spectrum ranging from uniform to random or clustered patterns.

Kleinn and Vilčko (2006b) introduced an rCD method using the framework of design-based probabilistic sampling. It uses the idea of defining an inclusion zone around each tree. The size of the inclusion zone is therefore a measure of the individual tree's inclusion probability when sampling is done with random sample points. Once the inclusion probabilities are known for all sampled trees, the Horvitz-Thompson estimator can be used as an unbiased estimator for any stand variable. In point-to-tree distance sampling, the inclusion zone of a particular tree depends exclusively on the spatial arrangement of the neighboring trees. Such inclusion zones are determined by r -order Voronoi polygons, where r is the number of trees being sampled per sample point.

Magnussen et al. (2008) commented that the method requires identification of, and the measuring of distances to, a large number of additional trees, possibly as many as $2r^2$ or even $2r^3$. Such a burden eliminates any practical advantage over sampling with fixed area plots. An easier method was given by Fehrman et al. (2012), which includes a fixed number of 3 units per sampling location and is based on partial triangulations over their natural neighbors to derive the individual inclusion probabilities.

ORBIT and GAMPOI Estimators

Magnussen et al. (2008) briefly summarised some of the estimators used at the time and introduced two more model-based estimators they called the ORBIT and GAMPOI estimates. The ORBIT estimator is based on the principal that, as the single site density is estimated by $(r - 1)/(\pi x_r)^2$ (cf. 6.2), it often fluctuates substantially with r . This suggests that if all r distances are combined to produce a single estimate, it may hopefully having a dampening effect. This means allocating an area a_j ($j = 1, \dots, r$) to each tree and then estimating the density as r divided by the sum of these r areas. Assigned areas are defined to give an approximately unbiased estimator under an assumed random (Poisson) point pattern. The area allocated to the j th nearest tree is that of an annulus with an inner radius of $\sqrt{x_{j-1} \times x_j}$ and an outer radius of $\sqrt{x_j \times x_{j+1}}$ giving an area of $a_j = \pi x_j(x_{j+1} - x_{j-1})$. The density estimator is then

$$\widehat{D}_{\text{orbit}} = r \times \left(\sum_{j=1}^r a_j \right)^{-1} = \frac{r}{\pi(x_j x_{j+1} - x_0 x_1)}.$$

There is no observation of x_0 and x_{r+1} ; they must be predicted. The inner radius x_0 is taken as the radius of a disk centered at the sample location and for which the probability is 0.5 that it does not contain a randomly located tree in a co-located disk with radius x_1 . This definition gives $x_0 = \sqrt{\log(2)}x_1$ so that $\widehat{a}_1 = \pi x_1(x_2 - \sqrt{\log(2)}x_1)$. A prediction of a_r and hence of x_{r+1} is obtained from a

linear regression:

$$\hat{a}_r = r\hat{\beta} \quad \text{where} \quad \hat{\beta} = \frac{\sum_{j=1}^{r-1} ja_j}{\sum_{j=1}^{r-1} j^2},$$

where $\hat{\beta}$ is the estimate of the trend in the linear regression of a_j on j through the origin. If s sites are sampled, Magnussen et al. (2008) proposed the density estimator

$$\tilde{D}_{\text{orbit}} = \frac{sr}{\pi \sum_{i=1}^s (x_{ir} \hat{x}_{i,r+1} - \sqrt{\log(2)} x_{i1}^2)}.$$

Their second estimator GAMPOI is based on assuming a negative binomial distribution for the point pattern leading to the following density function for x_r , namely,

$$f(x_r | \alpha, \beta) = 2\pi^r x_r^{2r-1} \frac{\Gamma(r+\alpha)}{\Gamma(\alpha)\Gamma(r)} \beta^{-\alpha} (\beta^{-1} + \pi x_r^2)^{-(r+\alpha)}.$$

The authors gave expressions for the mean and variance of x_r , with restrictions on α . Estimates of α and β , and hence $\tilde{D}_{\text{GAMPOI}}$, are ideally obtained by maximizing the log-likelihood of sample data x_{ir} , ($i = 1, \dots, s$) or, in the case of unstable estimates (which the authors discussed), by the method of moments. They also briefly considered what happens when the distances are constrained to a maximum distance x_{\max} for the r th nearest tree. Seventeen estimators of density were listed with references and then compared by simulation. In general, the best choices for r were 4 and 6, and the two new estimators performed well. For $s \geq 30$, GAMPOI emerged as the best choice, with ORBIT as a contender.

Other Estimators

Northrup et al. (2010) used a completely different estimator that they call the reconstructed-based density estimator. They reconstructed the spatial structure of the sampled forest stand in a rectangular simulation window using the r-tree sampling data and additional, simulated tree positions in the space between the sample plots. Finally, the density of the reconstructed pattern in the simulation is used as an estimator of the true tree density. The method does require knowing the positions of all r trees on each sampled plot and was found to work best with $r = 6$ for density estimation. It compared favorably with other estimators.

Magnussen et al. (2012) found that as different estimators are sensitive to any clustering and can have different degrees of bias, some positive and some negative, a composite weighted average of 2 to 5 different estimators selected from 19 estimators, with weights w_c adding to one, can perform better than an individual

estimator. This combined estimator takes the general form

$$\widehat{F}_{com}(n_c, \mathbf{X}) = \sum_{c=1}^{n_c} \widehat{w}_c(\mathbf{x}) \times \widehat{D}_c,$$

where n_c is the number of estimators used and \mathbf{x} are possible covariates. Using estimators previously found to work the best in terms of root mean square errors, a final five were chosen as the best combination along with methods of estimating the weights. Nine forest tree populations with widely different densities and patterns of stem locations were used exclusively for validation purposes. A more robust less biased estimator was given by Magnussen (2014).

Zhu et al. (2014) introduced a triangle method based on the average \bar{d} of the distances to the three nearest neighbors and used the density estimator $\widehat{D} = 1/\bar{d}$. Although it was found to be useful in practice, the writer (GAFS) is unaware of any theory for the method. They also referred to an older method called the “point-centered quarter method” (Cottam et al., 1953, cf. Seber, 1982: 44), and considered in detail by Khan et al. (2016). A point is chosen at random, together with two perpendicular directions fixed in advance. If Z_i ($i = 1, \dots, 4$) are the distances from the point to the nearest individual in each of the four quarters and $Y = \frac{1}{4}\pi(\sum_{i=1}^4 Z_i^2)$, then $2DY \sim \chi_8^2$. If s sample points are used and $Y_s = \sum_{i=1}^s Y_i$, then $2DY_s \sim \chi_{8s}^2$, which can be used to construct a confidence interval for D . The estimate of density is

$$\widehat{D}_4 = \frac{4(4s - 1)}{\pi \sum_{i=1}^s \sum_{j=1}^4 x_{ij}^2},$$

where x_{ij} is the distance from the i th point in the j th quadrant ($r = 1$). The method requires the assumption of a random population distribution.

Some Comparisons

Burch and Sánchez Meador (2018) estimated the ages of trees in a forest and compared the properties of mean tree age and age-class distribution estimators using n -tree distance sampling, fixed-radius plot sampling, and variable-radius plot sampling. They found that the estimator based on fixed-radius plot sampling outperformed the other two estimators for a forest having a clustered spatial pattern. Estimators from variable-radius plot sampling did not fare so well. The decision to use the n -tree sampling method for a particular application may depend on a combination of its statistical properties, ease of implementation, and overall cost.

Ramezani et al. (2016) compared the fixed-plot unbiased method using simulation with three of the popular rCD (rth closest distance) methods previously described above (Moore, Prodan, Kleinn and Vilckos). At each site the coordinate position of each tree was recorded using the global positioning system (GPS). The distance methods are highly dependent on the spatial distribution of the trees and give biased estimates (except under a homogeneous spatial Poisson point process

for tree locations). The bias decreases with increasing r . They compared cost efficiencies (time required) and evaluated the mirage boundary correction method of Lynch (2012) described in Sect. 2.4, which generally helped but not for all spatial patterns.

It was found that fixed-sized plot sampling was more accurate than r -tree distance estimators in the estimation of tree density. The distance sampling methods, however, appeared to be compatible to the traditional plot sampling in all forest types with Moore's estimate performing the best of the three rCD methods in terms of bias. Any spatial pattern was first quantified using the CE index of Clark and Evans (1954). Their results showed that, in most cases, r -tree distance estimators were superior to fixed-sized plot sampling in terms of cost-efficiency.

Basiri et al. (2018) compared the various methods available for trees with regard to accuracy, precision, bias, time spent, and efficiency. They found that the closest distance methods performed well for “natural pure and mixed stands of *P. euphratica*”. Also, continued technological advances in range measurements and field computing have added to the attraction of closest-distance sampling because measurements can be completed ever faster.

Haxtema et al. (2012) used simulation, to compare three n -tree distance sampling (NTDS) estimators due to Moore, Prodan, and Kleinn and Vilčko. They found that the Moore estimator tended to underestimate density and basal area for small values of r , whereas the Prodan and Kleinn-Vilčko estimators tended to give upwardly biased estimates. The Moore estimator emerged as the best candidate among the NTDS estimators examined, having the lowest bias for estimation of density and basal area on most stem maps. Similarly, the Moore estimator had the lowest relative root-mean-square error (RRMSE) values on most stem maps when larger sample sizes were considered, particularly for the estimation of density. The Prodan estimator performed poorly on the stem maps examined. In comparing the three estimators, Kleinn and Vilčko (2006a) also found that the Prodan estimator had the highest bias on all stem maps except those with a uniform spatial distribution. Lynch and Rusydi (1999) found that the Prodan estimator had negligible bias in uniformly spaced teak plantations, where the Moore estimator tended to underestimate volume and density.

As predicted by the theory, FPS (Circular fixed plot sampling) and HPS (Variable-radius plot sampling) described above had the lowest RRMSE for the estimation of density and basal area, respectively. Also, in estimating basal area, RRMSE values for the Moore estimator and FPS appeared to converge with those of HPS with increasing sample size. For the estimation of density, RRMSE values for the Moore estimator and HPS appeared to likewise converge with those of FPS on most stem maps. The increasing competitiveness of FPS and HPS for basal area and density estimation, respectively, with larger sample size is reflective of the design-unbiasedness of these sampling methods.

6.4.3 Three Problems with Closest Distance Methods

The above closest distance methods can be subject to three problems. The first is that many forests have spatial distributions that may not follow the Poisson or negative binomial distributions. Many natural stands are clumped, and plantations may have uniform spatial distributions, which have led to a variety of outcomes and biased estimates, as seen above. This means that Eberhardt's estimate can have low bias in some situations but high in others, as already noted. Alternative distributions are given by Dacey (1964, 1965, 1966), used for geographical research.

The second problem is the edge effect referred to above, where a random point location leads to the r th closest tree, and possibly closer ones, being outside the population area. A method for handling this situation is to use the so-called mirage point previously mentioned but which is different in this case as there are no fixed plot sizes. As described above, the perpendicular distance d between the boundary and the sample point or plot center is first measured. Then the r sample trees closest to either the original sample point or the mirage point are selected for use in one of the r -tree sampling estimation methods. Lynch (2012) described the method in detail for the r th closest distance methods and considered issues such as the mirage point is inaccessible (e.g., close to water), the mirage boundary correction is in a corner, and the need for piecewise linear boundaries.

The third problem is that unlike FPS and HPS mentioned above, selection probabilities of trees cannot be known unless distances and azimuths to many additional trees are acquired (Kleinn & Vilčko, 2006b). As noted above, this means that design-unbiased estimation for this method is currently not operationally feasible. Because the plot size at each sample point is computed as the smallest that could contain n trees, this uncorrected estimator will systematically overestimate the value of any attribute on a per-acre basis (Kleinn & Vilčko, 2006a).

6.5 Summary

This is a specialized and short chapter that deals with stationary populations of objects such as plants and trees and gives many methods currently used for estimating population density and related parameters. There are two basic methods. The first uses the distance of the closest or r th closest object from a chosen *point*, and the second uses the closest (nearest neighbor) or r th nearest neighbor to a chosen *object*.

The estimators, which are extensively reviewed, have been referred to as plotless density estimators. Sometimes the two methods have been combined to test for population randomness. Estimators and some variances are given for objects that have a spatial Poisson, negative-binomial, and general nonrandom distributions.

Trees are given a prominent place, where the above methods are also compared with other population methods such as plot sampling and distance sampling from points and transects. Several studies comparing various methods are described and concludes with three shortcomings of the closest distance method.

Chapter 7

Point Counts and Point-Distance Methods



7.1 Introduction

The term “point count,” a commonly used term, can have two meanings in the literature. The first is that it can refer to the number counted in a circular plot with incomplete detection and centered on a point chosen by some random method. This topic is discussed in Chap. 2. The second meaning, often referred to as the point-distance or point transect sampling, is the main emphasis of this chapter. Here, we look at methods of estimating the number or density of objects or animals in a population using their distances from some observation location at a point or, as in the next chapter, from a point on a line transect. We would expect such distances to be related to population density in some way as more shorter distances will arise from more dense populations. The term “distance sampling” was introduced by Buckland et al. (1993), and a foundation book on the subject is Buckland et al. (2001), with later editions.

The problem of incomplete detection is considered in this chapter in detail, as in Chap. 4 on occupancy. Using distances from a location was mentioned early on by Emlen (1971) and Reynolds et al. (1980) to provide density estimates. Distances measured from a line transect are discussed in the next chapter (Chap. 8). Also, in that chapter, point and transect distance counts are sometimes considered together in some references, as the theory is very similar, though the application is different. The two methods are also compared in that chapter (Sect. 8.10). A helpful practical discussion of both methods was given by Buckland et al. (2008). A readable, simple introduction to point-distance sampling, which includes multiple observers and the time to detection method, was given by Stanislav et al. (2010). Both Chaps. 7 and 8 of this book can be considered together.

Distance methods are widely used with many innovations. A major problem with such a diversity of methods is that many different protocols are used so that it can become difficult to compare studies. Sólymos et al. (2013) revived the importance

of having common protocols, referring specifically to those standards introduced by Ralph et al. (1993) and Ralph et al. (1995a). They noted that many research projects did not follow the recommendations. The standards made two recommendations regarding the count period of time: (1) The time spent at each point-count station should be 5 or 10 min, depending on the travel time between points; and (2) the data should be separated into those individuals first seen or heard in the intervals 0–3, 3–5, and 5–10 min, the latter only for a 10-min count. However, the time spent at a station will depend on the species, as there are various recommendations. It should be noted that counts of birds, for example, increase in a nonlinear fashion with a lengthening of either count period or count (Ralph et al. 1995b).

Two recommendations were made concerning the count radius: (1) All individual birds detected at a count station should be recorded, which implies that unlimited-distance surveys should be conducted to maximize the number of birds detected; and (2) birds first detected within 50 m of the point should be recorded separately from birds detected beyond 50 m. Counts within 50 m of the point-count station would allow comparisons of abundance among species because species differences in detection rates would be minimized within this small count radius. Design questions are discussed further in Sect. 7.2.2.

Sólymos et al. (2013) recommended that researchers undertaking more complex data collections should maintain the common standards whenever possible. Also, if multiple-observer methods are used, then one or all observers should conduct an independent survey that is not influenced by the presence of the other observers. The data from the multiple-observer surveys should be clearly separated by observer, and only one of the replicates should be combined and jointly analyzed with data from single-observer surveys.

7.2 Point-Distance Sampling

In this method, an observation point (site) is chosen at random away from a habitat edge, and the radial distances are recorded of all animals detected out to a fixed distance W from the site, i.e., within a circle of area πW^2 . Suppose that s such sites are chosen (suitably far apart to avoid duplicate counts with some species) so that the total area observed is $s\pi W^2$. Let n_i be the number of animals observed from site i so that $n = \sum_{i=1}^s n_i$ is the total number of animals observed. If the circular plot is small so that all animals on the plot are seen, then the density can be estimated by $n/(s\pi W^2)$.

If not all animals are seen, let r_i be the radial distance of the i th animal observed ($i = 1, \dots, n$), and let $g(r; \theta)$ or simply $g(r)$ for convenience, the so-called detection function, be the probability that an animal is detected, given it is at distance r from the observation post ($0 \leq r \leq W$). It is initially assumed that $g(0) = 1$ if animals very close to the observer will not be missed. Let P_W be the proportion of animals available for detection. We now provide a general theory.

7.2.1 Theory for Point Distances

Since there is some looseness in the literature about which probability distributions are conditional, we follow Seber (2001) to give a rigorous approach. In this theory, it is helpful to use capital letters for random variables, although it is not generally used elsewhere in this book because of notation complications. We use a bivariate distribution with one discrete and one continuous random variable.

Let $X = 1$ if a given animal in the observation area is observed and 0 otherwise. If $f_{X,R}(x, r)$ ($0 \leq r \leq W$) is the joint density function of discrete X and continuous distance R , then $g(r) = f_{X,R}(1 | r)$ is the detection function and the probability of detecting an individual at distance r . Let $f(r)$ ($= f_R(r)$) represent the density function for the distribution of animals (whether detected or not) with radial distance r . The probability that an animal is in an annulus with internal radius r and external radius $r + dr$ with respect to a circle of area πW^2 is $2\pi r dr / (\pi W^2)$, so that $f(r) = 2r/W^2$ ($0 \leq r \leq W$). The probability that an animal is observed is

$$P_W = f_X(1) = \int_0^W f_{X|R}(1 | r) f(r) dr = \int_0^W g(r) \frac{2r}{W^2} dr = \frac{2}{W^2} \mu_r, \quad (7.1)$$

and the probability density function of r , given the animal is detected, is

$$f_c(r) = f_{R|X}(r | 1) = \frac{f_{X|R}(1 | r) f(r)}{f_X(1)} = \frac{rg(r)}{\int_0^W rg(r) dr} = \frac{rg(r)}{\mu_r}.$$

The subscript “c” denotes that the density function is conditional on detection. Now, $f_c(r)/r = g(r)/\mu_r$, and, since $g(0) = 1$, we have

$$\lim_{r \rightarrow 0} \frac{f_c(r)}{r} = \frac{1}{\mu_r}.$$

The left-hand side of the above equation is the slope of $f_c(r)$ at $r = 0$, say $h(0)$. Then, if \hat{h} is an estimate of $h(0)$, we have the estimate

$$\hat{\mu}_r = 1/\hat{h}.$$

Also,

$$\hat{P}_W = \frac{2}{W^2} \hat{\mu}_r,$$

so that the population density from the s circles of total area $s\pi W^2$ can be estimated by

$$\hat{D}_W = \frac{n}{s\pi W^2 \hat{P}_W} = \frac{n}{2\pi s \hat{\mu}_r} = \frac{n\hat{h}}{2\pi s}. \quad (7.2)$$

If N_W is the total number of animals in the total observed area (called the “covered” area), and $E[\hat{h}] \approx h(0)$, then we see that \hat{D}_W is an unbiased estimate of D_W as follows:

$$\begin{aligned} E[\hat{D}_W] &= E_n[E(\hat{D}_W | n)] \\ &\approx E[n] \frac{h(0)}{2\pi s} \\ &= N_W P_W \frac{h(0)}{2\pi s} \\ &= N_W \frac{2}{W^2} \frac{\mu_r h(0)}{2\pi s} \\ &= \frac{N_W}{s\pi W^2} \\ &= D_W. \end{aligned}$$

Now,

$$\text{cov}[n, \hat{h}] = E_n\{E[(n - E[n])(\hat{h} - h(0)) | n]\} \approx 0,$$

so that by the delta method (cf. Appendix A.4),

$$\begin{aligned} \text{var}[\hat{D}_W] &\approx D_W^2 \left\{ \frac{\text{var}[n]}{(E[n])^2} + \frac{\text{var}[\hat{h}]}{[h(0)]^2} \right\} \\ &= D_W^2 [C_n^2 + C_{\hat{h}}^2] \end{aligned}$$

or

$$C^2 = C_n^2 + C_{\hat{h}}^2,$$

where C , C_n , and $C_{\hat{h}}$ are the coefficients of variation of \hat{D}_W , n , and \hat{h} , respectively. If $\bar{n} = n/s$, then $\text{var}[n]$ can be estimated by

$$\widehat{\text{var}}[n] = s \sum_{i=1}^s (n_i - \bar{n})^2 / (s - 1).$$

An important issue is how long to spend in each area searching. For example, with birds, if the search period is too short, then birds will be overlooked, while if the period too long, the birds can move around and perhaps get counted multiple times. Also, birds may move in and out of the study area.

Johnson (2008) has two tables listing some of the problems relating to availability and detectability (which he calls perceptibility), especially with regard to bird surveys. There are both variations with respect to the birds and their calls and with

respect to the observer(s). Also, there is a requirement to have a large number of detections (e.g., 60–100 Buckland et al. 2001) to satisfactorily estimate the detectability function $g(r)$ at $r = 0$. This means using distance sampling for just the most common species that are typically of lesser conservation interest or else pooling species in some way. The latter can be uncertain because of lack of information about possibly different distance functions for different species. Such pooling should be based on similarity of features that influence detectability, such as size, behavior, vocalization type and pattern, and microhabitat use, and not just because numbers are small. The methods are sensitive to having inaccurate measurements of distance, particularly those near the count station.

Heterogeneity

Efford and Dawson (2009) discussed the problems arising from distance-related heterogeneity and its effect on estimates using methods available at the time. For example, unmodeled individual heterogeneity using removal and capture-recapture methods, as discussed in later chapters, can lead to the underestimation of population size. They noted that simulation methods needed to incorporate such heterogeneity in, for example, the practical method of simulated field trials used by Alldredge et al. (2007a) and Alldredge et al. (2007c) for a population of known size. Efford and Dawson (2009) used simulation to assess nine methods for the conduct and analysis of point counts when the data included distance-related heterogeneity of individual detection probabilities. The detection probability $g_1(r)$ per occasion was modeled as a half-normal function of distance, namely,

$$g_1(r) = \exp[-r^2/(2\sigma^2)], \quad r < W.$$

In the case of repeated counts, the probability of recording a bird at least once on T occasions is

$$g_T(r) = 1 - \{1 - \exp[-r^2/(2\sigma^2)]\}^T,$$

assuming independence between occasions and the same detection function in each. For distance r_i to the i th individual, we have $p_{Ti} = g_T(r_i)$, and from Sect. 7.2.1, we have

$$\mu_{pt} = E[p_T] = 2W^2 \int_0^W r g_T(r) dr$$

and the “effective detection area” is

$$a = 2\pi \int_0^W r g_T(r) dr = \pi W^2 E[p_T] = \pi W^2 \mu_{pt}.$$

The authors measured the heterogeneity of detection probability within the population as the variance divided by the maximum variance for a distribution with the same mean (Dorazio & Royle 2003; Pledger 2005), namely,

$$\eta = \frac{\text{var}[p_T]}{\mu_{p_T}(1 - \mu_{p_T})},$$

where, from Efford and Dawson (2009),

$$\text{var}[p_T] = 2W^{-2} \int_0^W r[g_T(r) - \mu_{p_T}]dr,$$

and the conventional coefficient of variation

$$CV(p_T) = \frac{\sqrt{\text{var}[p_T]}}{\mu_{p_T}}.$$

The above measures were evaluated using numerical integration for varying σ/W and $T = 4$. The dependent and independent double-observer, and multiple-observer, methods were used as well as the “time-to-first-detection” or removal method and repeated counts (binomial N-mixture) method. Two-stage Monte Carlo trials were used, and for $0.2 \leq \sigma/W \leq 0.6$, μ_{p_T} and η were very sensitive to that ratio.

Bias

Bias was a problem that was not solved by using mixture models or including distance as a covariate. Efford and Dawson (2009) concluded that “no existing method allows effective estimation of population size from point counts.” In addition to birds (e.g., frogs), there are many field situations where animals are detected only by their vocalizations, and so it can be difficult to obtain precise distance information. The clear message from the simulated study by Alldredge et al. (2007a) is that the ability of observers to accurately gauge distances to singing birds is highly questionable. Consequently, distances are frequently recorded by grouping into discrete distance intervals, as described below.

7.2.2 Experimental Design

With regard to point location, Strindberg et al. (2004) noted that geographic information systems (GIS), together with advances such as the Global Positioning System (GPS) and remote sensing and satellite technology, mean that geo-referenced data of various sorts are readily available. Also, to take full advantage of GIS, there is a need to develop a range of automated design algorithms so that

efficient designs for various circumstances can be generated. The authors made a number of suggestions for designing a point transect design, which are described as follows. When a point is chosen, this creates a circular plot of radius W . The first recommendation, usually readily carried out, is that the number of sampled points K should be at least 20 and preferably many more. Although a simple random sample of points is always mentioned, there can be some problems associated with this design. For example, there is the possible edge effect, mentioned in Sect. 2.4. Here, points can end up too close to the boundary of the region under investigation so that counts out to a distance W may end up outside the region. If this problem is ignored, the estimated detection function is then biased, because it is now a composite function, reflecting both probability of detection and availability of objects (Buckland et al. 2001, 214). However, provided this composite function is well approximated by the model for the detection function, abundance estimation remains asymptotically unbiased.

Alternatively, we can go outside the boundary, provided the density is similar on both sides of the boundary, and include these distances in the buffer zone of width W around the region in the analysis to estimate density and then multiply by just the area of the investigated region excluding the buffer zone to estimate the population size for the region. If the density varies in the buffer zone, another approach is to survey the region plus the buffer zone, but only detections that fall within the survey region are used. The best option, which is implemented in the computer **Distance 4** using GIS functionality, is to record the effort for a point as the proportion of the circle of radius W and centered on the point that is in the survey region (assuming each point is surveyed just once). Thus, points within a distance W of the boundary, or in the buffer zone, have effort less than one. To estimate abundance, estimated density is multiplied by the size of the survey region (excluding the buffer zone), as with a normal survey.

Another problem is that the circular plots might overlap, and this compromises the assumed independence between points, so that standard variance estimation is biased. Fortunately, standard methods in **Distance** are robust to this assumption failure, whether it is due to overlapping circles in a simple random design or is due to use of a systematic design coupled with an analysis that assumes it was a simple random design. However, it is an inefficient use of resources to have fieldworkers surveying points within $2W$ of each other. An alternative method is to randomly place a grid of squares of side at least $2W$ over the region and choose a simple random sample of squares and locate a point at the center of each selected square. Squares with centers falling outside the survey region are not included in the sampling frame. If a buffer zone is used, those centers that fall outside the survey region but within W of the boundary would also be included.

Uneven spatial sampling can occur accidentally with any realization of a design. In this case, stratified sampling might be needed, or spatial modeling might be a better option, though the latter is model-based. Within strata, simple random, systematic, or cluster sampling of points might be adopted, depending on circumstances. Also, if the survey region is large, a simple random or systematic sample of points may be inefficient, especially if it is costly to travel from one point

to another. Cluster sampling is then a sensible option so that once a fieldworker reaches a sample location, he or she can sample several points around that location. We now consider the problem of choosing the number of sample points.

Sample Size

There are two components to sample size: the number of point transects in the design and the number of animals (or animal clusters for clustered populations) detected. Buckland et al. (2001, 232) recommended a minimum of 10–20 points to allow reliable estimation of variance of the encounter rate. If the population is very patchily distributed, then more points are needed to ensure that the variability through the study area is adequately represented and estimated. Buckland et al. (2001, 240–241) also suggested that at least 75–80 animals (or animal clusters) should be detected for reliable estimation of the detection function.

To determine the number of points to use, one can carry out a pilot survey using K_0 points that provides sufficient data to estimate density D and its coefficient of variation by \widehat{D} and $cv_0(\widehat{D})$, respectively, with the target coefficient of variation set at $cv_t(\widehat{D})$. An estimate of the K points needed for the main survey is then

$$K = \frac{K_0[cv_0(\widehat{D})]^2}{[cv_t(\widehat{D})]^2}.$$

If the pilot data are insufficient for this method, and n_0 animals are detected, we can use

$$K = \frac{K_0}{n_0} \frac{b}{[cv_t(\widehat{D})]^2},$$

with b set equal to 3 (Buckland et al. 2001, 242).

7.2.3 Measurement Errors

This topic of measurement errors, whether multiplicative or additive, is not mentioned much in the literature. We now follow Borchers et al. (2010) who considered the problem for both point and line transect methods and gave references to previous methods used. We use a similar notation to that of the previous section, but now with both true (unknown) and observed distances, with covariates added.

For an error-free model, let $f_t(r | \mathbf{z})$, ($0 \leq r \leq W$) represent the density function (pdf) for animals (whether detected or not) with *true* radial distance r and covariates \mathbf{z} . Let $g(r | \mathbf{z}; \boldsymbol{\theta})$ be the probability of detecting an animal at distance (true) r with an associated vector of unknown parameters $\boldsymbol{\theta}$. It is assumed that all animals at distance $r = 0$ are detected, i.e., $g(0 | \mathbf{z}; \boldsymbol{\theta}) = 1$, that animals are distributed independently,

and that detections are independent. Then the pdf of r , given the animal is detected, with covariates \mathbf{z} , is the conditional density

$$f_{ct}(r \mid \mathbf{z}; \boldsymbol{\theta}) = \frac{g(r \mid \mathbf{z}; \boldsymbol{\theta}) f_t(r \mid \mathbf{z})}{g(\mathbf{z}; \boldsymbol{\theta})},$$

where

$$g(\mathbf{z}; \boldsymbol{\theta}) = \int g(r \mid \mathbf{z}; \boldsymbol{\theta}) f_t(r \mid \mathbf{z}) dr$$

with integration over the range of r . Then, given n observations, we have the usual likelihood

$$L_{true}(\boldsymbol{\theta}) = \prod_{i=1}^n f_{ct}(r_i \mid \mathbf{z}_i; \boldsymbol{\theta}).$$

We note that if the point is chosen at random, then from the previous Sect. 7.2.1 $f_t(r \mid \mathbf{z}) = 2r/W^2$ and the equations following Eq. (7.1) hold.

Considering now measurement error, let y be the observed distance (measured with error) and r the true distance to an animal with associated parameters $\boldsymbol{\phi}$ and density function $f_{err}(y \mid r, \mathbf{z}; \boldsymbol{\phi})$. Let $\mathbf{y} = (y_1, \dots, y_n)'$ be the n measured distances. Then the density function of the measured distance to a detection (given present) is the conditional pdf

$$f_{cy}(y \mid \mathbf{z}; \boldsymbol{\theta}, \boldsymbol{\phi}) = \int_0^\infty f_{err}(y \mid r, \mathbf{z}; \boldsymbol{\phi}) f_{ct}(r \mid \mathbf{z}; \boldsymbol{\theta}) dr.$$

Suppose that in addition to the survey data (\mathbf{y}, \mathbf{Z}) , where $\mathbf{Z} = (\mathbf{z}_1, \dots, \mathbf{z}_n)$, there are observations from an experiment in which both true and measured distances are observed. For convenience, these extra observations are numbered from $n + 1$ to $n + K$ and denoted by $\mathbf{y}^+ = (y_{n+1}, \dots, y_{n+K})$, $\mathbf{r}^+ = (r_{n+1}, \dots, r_{n+K})'$, and $\mathbf{Z}^+ = (\mathbf{z}_{n+1}, \dots, \mathbf{z}_{n+K})$. Given the data $\mathbf{y}, \mathbf{Z}, \mathbf{y}^+, \mathbf{Z}^+, \mathbf{r}^+$ and an error model $f_{err}(y \mid r, \mathbf{z}; \boldsymbol{\phi})$, the likelihood for $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ is

$$\begin{aligned} L(\boldsymbol{\theta}, \boldsymbol{\phi}) &= \prod_{i=1}^n f_{cy}(y_i \mid \mathbf{z}_i; \boldsymbol{\theta}, \boldsymbol{\phi}) \prod_{k=1}^K f_{err}(y_{n+k} \mid r_{n+k}, \mathbf{z}_{n+k}; \boldsymbol{\phi}) \\ &= L_n(\boldsymbol{\theta}, \boldsymbol{\phi}) L_{err}(\boldsymbol{\phi}). \end{aligned}$$

From the above equation, we can obtain maximum likelihood estimates of $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$. Alternatively, we can use a conditional approach of maximizing $L(\boldsymbol{\phi})$ for $\boldsymbol{\phi}$ to get $\hat{\boldsymbol{\phi}}_c$ and use this estimate to maximize $L_n(\boldsymbol{\theta}, \hat{\boldsymbol{\phi}}_c)$ to get $\hat{\boldsymbol{\theta}}_c$. Although this method has lower efficiency, it is easier to compute.

The authors gave an extension of the model to the case when no true distances are known but there are K independent pairs of distance measurements from two observers. They also considered density estimation, variances, and interval estimation.

The above model also holds for line transects, where r is now the true perpendicular distance of the animal from the line and y is the observed perpendicular distance. Methods for estimating detection functions, population density, variances, and confidence intervals are given for both point and line transects.

Correction-Factor Estimators

Borchers et al. (2010) extended a method of Marques (2004) to allow for multiplicative errors so that $y = r\epsilon$, where ϵ is the error random variable with density function $f_\epsilon(\epsilon | \mathbf{z})$, which does not depend on r . Then

$$f_{err}(y | r, \mathbf{z}) = (\epsilon/y) f_\epsilon(\epsilon | \mathbf{z}),$$

and, ignoring incidental parameters, we finally end up with

$$f_{cy}(y | \mathbf{z}) = \int_0^\infty \frac{1}{\epsilon} f_\epsilon(\epsilon | \mathbf{z}) f_t(y/\epsilon | \mathbf{z}) d\epsilon.$$

7.2.4 Detection Functions

A variety of detection functions $g(y)$ have been proposed (see also Sect. 8.2.5). Buckland et al. (2001) considered the model

$$g(y) = \frac{k(y)[1 + s(y)]}{k(0)[1 + s(0)]},$$

where $k(y)$ is a parametric key function, such as the uniform, half-normal, or hazard-rate models, and $s(y)$ is a series expansion with three described, namely, cosine, polynomial, and Hermite polynomial. These are described by Marques et al. (2007) who combined the detection function with multiple covariates, provided they were not strongly correlated, as with any regression model.

As already noted at the beginning of the chapter, there are a number of key issues involved with point sampling such as how long to spend at each sample point. It can become a compromise between the need to maximize the probability of recording individual birds and perhaps $g(0)$ and avoiding overestimating mobile and conspicuous species. For example, mobile birds could be detected more than once, or there is a chance that a bird too distant to detect will move into detection range.

Cimprich (2009) looked at the issue of how much time to spend at each point for black-capped vireos, which are typically active, sing intermittently, and sometimes move tens of meters between songs. He used different count times to see which gave the best results for the species and recommended that “investigators collect preliminary data to establish an appropriate count duration when accurate estimates of absolute, rather than relative, abundance are important.”

Peak (2011) considered the same problem using golden-cheeked warblers and recommended that “counts for other species in similar habitats with similar behavior and movement patterns also should be limited to 2 min when unbiased estimates are important and conducting field tests of the point transect distance sampling method.” He examined the performance of standard distance sampling estimators under model selection. Line and point transect estimators with distances simulated from two detection functions, hazard-rate and exponential power series, were compared for a range of sample sizes. A similar type of comparison was carried out by Prieto Gonzalez et al. (2017) for point and line estimators using two detection functions, hazard-rate and exponential power series.

7.2.5 *Point Sampling Along Lines*

Since it can be difficult to choose a point at random, a common method is to choose systematic points on a line and measure the distances out to individuals seen in a circle of fixed radius W . The theory is essentially the same as before, but it will depend on how the transect and points are chosen. For example, points can be located randomly on the line or else placed systematically with a random start. Lee and Marsden (2008) investigated what is the best search interval for birds. They noted that the length of the count period becomes a compromise between the need to maximize the probability of recording individual birds and with certainty at zero distance and avoiding overestimating mobile and conspicuous species. Incorporating a longer count period is advantageous in tropical forests where many birds are cryptic (good at hiding), especially with difficult vegetation. Various length time intervals were discussed, and it was found that incorporating a settling down period seriously underestimated abundance of nearly all species due to evasive movement and changes in detectability.

Strindberg et al. (2004) briefly mentioned the question of designing such a method. If the spacing between successive lines is the same as that between successive points on a single line, then the points fall on a regular grid and are therefore a systematic sample of points. If, however, the spacing between lines is greater than that between points within lines, then the lines should be treated as clusters of points, because points within the same line cannot be considered independent. In other words, the lines must be treated as the sampling units, rather than the individual points within lines.

Marques et al. (2010) considered using linear features such as roads or paths rather than straight-line transects for a population of Irish hares (*Lepus timidus*

hibernicus). A key requirement of distance sampling is that samplers (lines or points) are placed according to a randomized design, which ensures that samplers are positioned independent of animals. However, samplers are often placed along linear features such as roads or tracks, so that bias is expected if animals are not uniformly distributed with respect to distance from the linear feature. Several linear features can be used. Theory for the method follows.

It is assumed that all animals are available and points $i = 1, \dots, I$, are sampled. Let $g(r | \phi_1)$ represent the probability of detecting an animal at radial distance r from a point, given present (with distribution parameters ϕ_1). Let $\pi_r(r)$ be the probability density function (pdf) of distance r for an individual, irrespective of whether detected or not. With the so-called uniformity assumption, we have $\pi_r(r) = 2r/W^2$ for points (Sect. 7.2.1, $f(r)$). The probability of detecting an animal, given it is in the covered area, is

$$P = \int_0^W g(r | \phi_1) \pi_r(r) dr.$$

If the covered area is a , then the density D is estimated by $n/(a\widehat{P})$, where n is the number seen. Let θ be the angle the radial distance bears to the line with distribution $\pi_\theta(\theta)$, where r and θ are assumed to be independent. We then have a bivariate likelihood

$$L_\theta L_r = \prod_{i=1}^n \pi_\theta(\theta_i) \frac{g(r_i) \pi_r(r_i)}{\int_0^W g(r) \pi_r(r) dr}.$$

For convenience, the authors sometimes drop ϕ_1 and ϕ_2 below from the notation. Where the method is different from previous ones is that the sampling design now consists of first randomly choosing a square of side $2W$ so that the linear feature runs along its centerline, with origin at the square's center.

For an object in the square, let x be the x -coordinate of an object measured along the line, and let $h(x | \phi_2)$, defined on the interval $(0, W)$, be its probability density function (pdf) irrespective of whether detected or not. Here, ϕ_2 is a distributional parameter. Because of the random square placement, the pdf of object locations in the y -dimension, given x , is $\pi^*(y | x) = 1/W$ leading to

$$\pi^*(x, y) = \pi^*(y | x) h(x) = h(x)/W.$$

After obtaining the joint pdf of x and y conditional on $r \leq W$, and changing to polar coordinates, the authors obtained the joint pdf of observed (r, θ) as

$$f(r, \theta | \phi_1, \phi_2) = \frac{rh(r \sin \theta | \phi_2)g(r | \phi_1)}{\int_0^W \int_0^{\pi/2} rh(r \sin \theta | \phi_2)g(r | \phi_1)d\theta dr}.$$

This formulation explicitly allows the inclusion of a non-uniform model for $h(x \mid \phi_2)$. Given the n observed (r, θ) pairs, and assuming a parametric form for $h(x \mid \phi_2)$, the authors showed that the joint distribution of r and θ is

$$\pi(r, \theta) = rh(r \sin \theta \mid \phi_2)/(WP_c),$$

where

$$P_c = \int_0^W \sqrt{W^2 - x^2} h(x \mid \phi_2)/W dx,$$

which can be estimated by \widehat{P}_c from an estimated $h(x \mid \widehat{\phi}_2)$. We now have the likelihood

$$L(\phi_1, \phi_2 \mid \{r_i, \theta_i\}) = \prod_{i=1}^n \frac{\pi(r_i, \theta_i)g(r_i \mid \phi_1)}{\int_0^W \int_0^{\pi/2} \pi(r, \theta)g(r \mid \phi_1)d\theta dr}.$$

Analogous to the conventional case,

$$P = \int_0^W g(r \mid \phi_1)\pi_r(r \mid \phi_2)dr,$$

where

$$\pi_r(r) = r \int_0^{\pi/2} h(r \sin \theta \mid \phi_2)d\theta/(WP_c),$$

from which we can obtain maximum likelihood estimates $\widehat{\phi}_1$, $\widehat{\phi}_2$, and \widehat{P}_c . This leads to the estimate

$$\widehat{P} = \int_0^W g(r \mid \widehat{\phi}_1)\pi_r(r \mid \widehat{\phi}_2) dr.$$

The estimated density in the strip of half-width W about the linear feature is then

$$\widehat{D} = \frac{n}{a\widehat{P}\widehat{P}_c},$$

where $a = k4W^2$, the total area of the k squares containing the covered circles. When $h(x)$ is constant, $P_c = \pi/4$. A density estimator over the whole line of width $2W$ is given, and variance estimators are obtained using bootstrapping.

7.2.6 Cluster Sampling

Animals are often seen in clusters, where a cluster is defined to be a group of animals having a well-defined location for the group center point and the number s in the group is recorded. Buckland et al. (2009) considered the situation where there are plots and points are placed at random within each plot. The density estimate for plot k is then, from an adaption of (7.2) in Sect. 7.2.1 with $h(0) = f'_c(0)$,

$$\widehat{D}_k = \frac{1}{2\pi e_k} \sum_{i=1}^{n_k} s_i \widehat{h}(0 \mid \mathbf{z}_i) = n_k \frac{H_k}{2\pi e_k},$$

where e_k is the “effort” or total number of points on the plot multiplied by the (mean) number of visits per point; n_k is the number of animal clusters detected on plot k ; $\widehat{h}(r \mid \mathbf{z}_i) = \widehat{f}'_c(r \mid \mathbf{z}_i)$ is the estimated derivative of the probability density function for the i th detected cluster on plot k being detected at distance r , given present; and \mathbf{z}_i are the covariates for that cluster. We have the sample mean

$$H_k = \frac{1}{n_k} \sum_{i=1}^{n_k} s_i \widehat{f}'_c(0 \mid \mathbf{z}_i).$$

The authors discussed bootstrapping to get variance estimates, much the same as with line transect distances in Chap. 8.

7.3 Combined with Removal Method

We now consider using the point-count method combined with a distance removal method. Sólymos et al. (2013) considered 3 approaches, namely, the general survey data, the removal model, and the distance model, and estimated singing rates for 75 boreal forest songbird species. They found that for the survey protocol of length of the time interval, radius length, and covariates, the point-count radius explained most of the variation in detectability. However, environmental and temporal covariates (date, time, vegetation) affected singing rates.

Detection was described as two components: p_a (availability), the probability that an individual bird present at the time of the survey gave a visual or auditory cue and is therefore available for detection; and the conditional probability (p_d) that an available bird is detected, referred to as “perceptibility.” Then $p = p_a p_d$ is the overall probability of detection.

For a given point, let y_{jk} be the number of counts during an increasing time period t_j ($j = 1, \dots, J, t_0 = 0$) with time intervals of length $\Delta t_j = t_j - t_{j-1}$, with t_J being the total time interval. We also have increasing radii r_k ($k = 1, \dots, K, r_0 = 0$), with r_K being the maximum radius. This gives K distance bands of width

$\Delta r_k = r_k - r_{k-1}$. Let $p_a(t_j)$ be the value of p_a for period t_j and $p_d(r_k)$ the value of p_d for radius r_k . Let D be the point-level population density per unit area, let $y_{j\cdot} = \sum_{k=1}^K y_{jk}$ and $y_{\cdot\cdot} = \sum_{j=1}^J y_{j\cdot}$, and let A be the area sampled from a point. If known, the latter would be πr_K^2 . For the removal approach, an individual is recorded when first detected and then effectively removed. The expected count first detected during time interval Δt_j can be written as

$$E[y_{j\cdot}] = DA[p_a(t_j) - p_a(t_{j-1})]p_d(r_K),$$

with $p_a(t_0) = 0$, and

$$E[y_{\cdot\cdot}] = DA p_a(t_J) p_d(r_K). \quad (7.3)$$

It is then assumed that singing events by individuals follow a Poisson process and that the population (i.e., singing individuals, usually males) at a given location and time is homogeneous in its singing rate of ϕ . Then, if the singing rate is constant within a sampling event, the time to the first singing event will follow an exponential distribution $f(t) \exp(-t\phi)$ so that

$$p_a(t_J) = \int_0^{t_J} \phi \exp[-t\phi] dt = 1 - \exp[-t_J\phi].$$

As is the case with some multinomial models, by obtaining the multinomial distribution for the counts conditional on the sum of the counts, nuisance parameters can be eliminated. Here, we consider the $y_{j\cdot}$ conditional on $y_{\cdot\cdot}$ giving us (from the appendix of Sólymos et al. 2013)

$$f(y_{1\cdot}, \dots, y_{J\cdot} \mid y_{\cdot\cdot}, \{\pi_{j\cdot}\}) = \frac{y_{\cdot\cdot}!}{\prod_{j=1}^J y_{j\cdot}!} \prod_{j=1}^J \pi_{j\cdot}^{y_{j\cdot}}, \quad (7.4)$$

where, for element $y_{j\cdot}$,

$$\begin{aligned} \pi_{j\cdot} &= \Pr(y \in y_{j\cdot} \mid y \in y_{\cdot\cdot}) \\ &= \frac{DA p_d(r_K)[p_a(t_j) - p_a(t_{j-1})]}{DA p_a(t_J) p_d(r_K)} \\ &= \frac{p_a(t_j) - p_a(t_{j-1})}{p_a(t_J)} \\ &= \frac{\exp[-t_{j-1}\phi] - \exp[-t_j\phi]}{1 - \exp[-t_j\phi]}, \end{aligned} \quad (7.5)$$

which eliminates $DA p_a(r_K)$ and depends only on ϕ . The overall likelihood is the product of the above probability function (7.4) over all the sample points. If ϕ_i

corresponds to sample point i , then $\log(\phi_i)$ can be expressed as a linear model with covariates that affect singing rates and regression parameters, which can be estimated from the overall likelihood.

Introducing the distance part of the model, all available birds at $r = 0$ are assumed to be detected, and they are detected at their initial locations. A detection function was used, namely, $g(r) = \exp[-r^2/\tau^2]$, so that using P_W in Sect. 7.2.1 Eq. (7.1) with $W = r_K$, we have

$$\begin{aligned} p_d(r_K) &= \int_0^{r_K} 2rg(r)dr/r_K^2 \\ &= \frac{\pi\tau^2(1 - \exp[-r_K^2/\tau^2])}{\pi r_K^2}. \end{aligned}$$

To obtain an estimate of τ , we find from the authors that

$$\begin{aligned} \pi_{\cdot k} &= \Pr(y \in y_{\cdot k} \mid y \in y_{\cdot \cdot}) \\ &= \frac{u(r_k) - u(r_{k-1})}{u(r_k)}, \end{aligned}$$

where

$$u(r_k) = (1 - \exp[-r_k^2/\tau^2]).$$

We can then set up a conditional multinomial distribution like Eq. (7.4), but with the elements $y_{\cdot k}$ and $\pi_{\cdot k}$, which eliminates $DAp_d(r_K)$ and gives us a distribution depending just on τ . For sample point i , $\log(\tau_i)$ can be expressed as a linear model in covariates (e.g., observer, habitat type) leading to an overall likelihood function like the one involving the ϕ_i . We can now combine the components of removal (relating to p_a) and distance (relating to p_d). For sample point i ($i = 1, \dots, n$), and using estimates $\hat{\phi}_i$ and $\hat{\tau}_i$ in p_{ai} and p_{di} , respectively, we have from Eq. (7.3)

$$\hat{D}_i = \frac{y_{\cdot \cdot}}{A_i \hat{p}_{ai}(t_J) \hat{p}_{di}(r_K)}.$$

The authors considered a joint estimation of ϕ_i and τ_i using

$$\pi_{jk} = Pr(y_{jk} \mid y_{jk} \in y_{\cdot \cdot}) = \pi_{j \cdot} \pi_{\cdot k}.$$

However, they showed that as the marginal removal and distance sampling models are orthogonal to each other, the joint estimator does not improve statistical efficiency and consistency. The reader is referred to their paper for further details including some special cases and computational methods.

7.4 Trapping Web

Point transect methods work well for many species, but some species are insufficiently visible or noisy to allow adequate numbers to be detected by observers standing at random points. This has led to combining trapping studies with point transect sampling using a standard trapping web introduced by Anderson et al. (1983). Lukacs et al. (2004) commented that trapping webs are very resource demanding requiring at least 90 traps with a minimum of 15 points (webs). The traps are arranged rather like a circular spider's web, with a higher density of traps at the center than at the edges, and dropping off as distance from the center increases, leading to an upward bias in density estimates.

Typically, traps would be prebaited and placed by skilled technicians to minimize disturbance to the area being trapped. The purpose of the web design is to allow an assumption that all trappable animals at the center of the web are captured during the course of t days of trapping. A number of standard assumptions are made including assuming all animals at the center of the web are captured with probability equal to one by the end of the t occasions. This is achieved by ensuring that trap density is high in the vicinity of the point and decreases with increasing distance from the point.

A weakness of the trapping web method is that it depends on the movement of the animals in order to trap them. Because trap density is higher near the center of the web, animals that move over a large area are more likely to be trapped close to the center, causing upward bias in the density estimates. Also, some animals initially off the web will move onto it and encounter a trap, leading to a recorded distance that is biased low. On the other hand, an animal on the web is more likely to encounter a trap if it moves toward the center, where trap density is high, than if it moves away from the center, where trap density is low. This bias toward shorter distances results in positive bias in density estimates. Bias can also arise if a trap becomes unavailable once it captures an animal, as animals near the center point or line might not be caught because the traps in their home ranges are unavailable. Clearly, the method has problems. However, Lukacs (2002) described simulation software **WebSim** to help the user design a survey to give low bias.

Following Anderson et al. (1983), the method consists of L lines of equal length a_T radiating from a single randomly chosen center point, each with T live traps, where the i th trap is at distance $a_i = (i - 1)\theta + (\theta/2)$ from the web center ($i = 1, \dots, T$). Let $b_i = i\theta$ be a point midway between two traps, with $b_0 = 0$ the center of the web and b_T the boundary of the web that lies beyond the last trap. Because traps are at fixed, with equally spaced intervals, distance data should be treated as "grouped" data. The captures in the i th ring of traps arise from the ring defined by distances b_{i-1} to b_i . If $c_i = \pi b_i^2$ is the total area of the web out to the i th ring, then the area trapped by this ring is $\Delta_i = c_i - c_{i-1} = \pi\theta^2(2i - 1)$.

Using first captures, let n_i be the number of different animals captured in the i th ring, and let $n = \sum_i n_i$. The population density for a particular line and its catchable area is then estimated by $\widehat{D} = n \widehat{f}(0)$, where $f(c)$ is the probability density function of capture at distance c , estimated from the data. The population size N for the individual web is estimated by $\widehat{N} = \widehat{D} c_T$. Data for the L transects, each with a trapping web, can be combined and an average density for the whole population obtained. Details are given in the appendix of the authors' paper.

7.5 Single Trap or Lure Method

Buckland et al. (2006) applied the point-distance method for species that are problematic to survey, in particular Scottish crossbills (*Loxia scotica* Hartert) in Britain, for which traditional methods do not work so well. They ran a separate study alongside the point transect trapping survey in which the location of animals relative to the trap location is known at the time the trap is set. Animals are lured to a point or trapped at a point, and the number observed at each point is recorded. Point-line trapping then consisted of a main survey in which numbers of animals captured are recorded and a separate study for which a subset of animals have known initial locations with respect to the traps. For the main survey, the trapping locations should be determined by an appropriate randomized design as for a normal point survey described above.

The separate study is conducted on a subset of animals to record whether they respond to the lure or enter the trap for a range of known distances from the point. These data are used to estimate the probability that an animal will respond to the lure or enter the trap, as a function of its initial distance from the point. This allows the counts to be converted to an estimate of abundance in the survey region. Lure point transects extend the applicability of distance sampling to species that can be lured to a point, while trapping point transects potentially allow abundance estimation of species that can be trapped, with fewer resources needed than trapping webs and conventional mark-recapture methods.

For the main study (trapping or lure survey), let:

K = number of points

n_k = number of clusters detected from point k , $k = 1, \dots, K$

$n = \sum_{k=1}^K n_k$ = total number of clusters detected

s_i = size of cluster or flock size i , $i = 1, \dots, n$

If animals do not occur in clusters, set the $s_i = 1$. For the trial data (separate study) from which the detection function is to be modeled, let:

m = number of trials (i.e., number of clusters tested)

$$y_i = \begin{cases} 0, & \text{cluster } i \text{ is not detected from the point} \\ 1, & \text{cluster } i \text{ reacts to lure, allowing detection from the point} \end{cases}$$

r_i = initial distance of cluster i from lure, $i = 1, \dots, m$

z_{ij} = value of covariate j for cluster i , $i = 1, \dots, m$, $j = 1, \dots, J$

The r_i cannot be observed for the n clusters detected during the main survey. We assume that covariates z can all be recorded for these clusters. One of these J covariates is likely to be cluster size. If $p_i = E[y_i]$, the probability the cluster is detected, the authors used a logistic regression model

$$\text{logit}(\hat{p}_i) = \alpha + \beta_0 r_i + \sum_{j=1}^J \beta_j z_{ij},$$

with the corresponding fitted values \hat{p}_i .

We now use this fitted model to estimate the probability of detection of those clusters detected in the main survey. For each of these detections, we can readily substitute values z_{ij} into the fitted model, but we do not know the r_i . This means that \hat{p}_i is also a function of the unknown r_i , namely, $\hat{p}_i \equiv \hat{p}(r, z_{i1}, \dots, z_{iJ})$ for $0 < r < W$, where W is some large distance at which a reaction to the lure is believed to be very unlikely. We therefore estimate the probability of detection of cluster i , unconditional on its distance from the point, by integrating over the unknown r , namely,

$$\hat{P}(z_{i1}, \dots, z_{iJ}) = \int_0^W \pi(r) \hat{p}_i(r, z_{i1}, \dots, z_{iJ}) dr,$$

where $\pi(r)$, $0 \leq r \leq W$, is the probability density function of distances of clusters (whether detected or not) from the point. In conventional point-distance sampling, clusters within one of the circles of radius W are assumed to be randomly positioned in the circle so that $\pi(r) = 2r/W^2$. This assumption is valid through random point placement (or, more usually, a systematic grid of points, randomly located, as described previously).

With regard to the edge effect, given random point placement and assuming that animals do not occur beyond the survey region boundary, the availability function is

$$\pi(r) = \frac{rq(r)}{\int_0^W rq(r) dr},$$

where $q(r) = \sum_{k=1}^K q_r(k)$ and $q_r(k)$ is the proportion of the circumference of a circle of radius r centered at point k that lies within the survey region for $0 \leq r \leq W$. If this proportion is always 1, then $\pi(r) = 2r/(W^2)$. Using a type of Horvitz-Thompson estimator, the estimated number of animals in the region covered by the survey is

$$\widehat{N}_c = \sum_{i=1}^n \frac{s_i}{\widehat{P}(z_{i1}, \dots, z_{iJ})}.$$

The estimated abundance for the entire survey region of area A is

$$\widehat{N} = \frac{A}{A_c} \widehat{N}_c,$$

where

$$A_c = 2\pi \sum_{k=1}^K \int_0^W r q_k(r) dr = 2\pi K \int_0^W r q(r) dr$$

is the area of the covered region. If $q(r)$ is always one, then $A_c = K\pi W^2$.

The authors, Buckland et al. (2006), discussed the choice of W along with methods for obtaining variance estimates and asymptotic confidence intervals. They also suggested a bootstrap approach using a resample corresponding to the main survey by sampling with replacement the K points along with their data and resampling the clusters from the experiment for fitting the detection function.

7.6 Some Model Variations

We now consider a number of methods that provide variations of the above method, including allowing the circular plot to vary in size, extending the above conditional likelihood to an unconditional likelihood, allowing repeated counts, and introducing spatial replication.

7.6.1 Variable Circular Plot

This method tended to be used prior to the focus on the fixed distance circular plot described above. A site is chosen at random, and, for a certain time period (allowing for the population to settle down after possible disturbance), all individuals seen out to a distance R when numbers start to fall off (so-called point of inflection) have their distances recorded (Ramsey & Scott 1979; Reynolds et al. 1980). This can be

done, for example, by counting the number of individuals seen in concentric bands. The circular area out to the point of inflection is then used as the sampled area.

Buckland (1987) suggested choosing R (now a random variable) so that 50% of those observed are used. Thus, for site i and given R_i , the above theory can be used by setting $W = R_i$ for that site. However, the randomness of R needs to be modeled. If we assume that at each site all animals observed are included, we effectively have $W = \infty$, which was used in early developments. Alternatively, it can be assumed that W is the maximum distance at which an animal can be observed, as with Quang (1993) who developed a kernel estimator of the population density along with a confidence interval.

7.6.2 Unconditional Likelihood

Up till now, we have used a conditional likelihood approach where the observations are conditional on detection. However, in Sect. 3.2.6, we introduced a model by Royle (2004a) that used an unconditional likelihood. This model was extended by Royle et al. (2004) as follows to incorporate distance measurements in distance classes and to provide a better link to abundance. Let $k = 1, \dots, K$ index the distance classes, with end points $(c_1, c_2), (c_2, c_3) \dots (c_K, c_{K+1})$, where $c_{K+1} = W$. Let y_{ik} be the observed count of individuals in distance class k for site i ($i = 1, \dots, s$), and let $\pi_k(\boldsymbol{\theta})$ be the probability that an individual is present and is detected in that distance class. This probability will depend on the detection function $g(y; \boldsymbol{\theta})$, the probability of detection at distance y , where $\boldsymbol{\theta}$ is to be estimated. It is found by integrating $g(y; \boldsymbol{\theta})$ over the area of the circle between c_k and c_{k+1} , namely,

$$\pi_k(\boldsymbol{\theta}) = \int_{c_k}^{c_{k+1}} \frac{2\pi y g(y; \boldsymbol{\theta})}{\pi W^2} dy.$$

Let N_i be the animal abundance at site i , and let $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{ik})'$ be the observations there. Then

$$f(\mathbf{y}_i | \boldsymbol{\theta}) = \frac{N_i!}{(\prod_k y_{ik}!) (N_i - Y_i)!} \left[\prod_k \pi_k(\boldsymbol{\theta})^{y_{ik}} \right] \times \left[1 - \sum_k \pi_k(\boldsymbol{\theta}) \right]^{N_i - Y_i},$$

where $Y_i = \sum_k y_{ik}$. We now assume that the N_i are random effects with some distribution and choose the Poisson for ease of exposition, namely,

$$N_i \sim \text{Poisson}(\lambda_i),$$

where λ_i , the mean of N_i , is the density per “sample site” (assuming all circular sites are the same size). Using a log-linear model, we have

$$\log(\lambda_i) = \alpha_0 + \alpha_1 z_i,$$

where z_i is the value of some covariate at site i such as foliage cover. As the N_i are not observable random effects, we now have to proceed differently by integrating (in this case summing) them from the likelihood, namely,

$$L(\boldsymbol{\alpha}, \boldsymbol{\theta} | \mathbf{y}_i) = \sum_{N_i=y_i}^{\infty} f(\mathbf{y}_i; \boldsymbol{\theta}) \frac{\exp[-\lambda_i(\boldsymbol{\alpha})] \lambda_i(\boldsymbol{\alpha})^{N_i}}{N_i!},$$

where $\boldsymbol{\alpha} = (\alpha_0, \alpha_1)'$. If the covariates are centered to have mean zero, then $\exp[\alpha_0]/(\pi^2 W)$ is the “mean” density.

One benefit of the Poisson abundance model is that the summation in the above equation can be done analytically, yielding the product Poisson likelihood

$$L(\boldsymbol{\alpha}, \boldsymbol{\theta} | \mathbf{y}_i) \prod_{k=1}^K \text{Poisson}(y_{ik}; \lambda_i(\boldsymbol{\alpha}) \pi_k(\boldsymbol{\theta})).$$

With other abundance models such as the negative binomial, we do not get such simplification, and numerical integration is involved. If we can assume that the data are independent across sites, for example, the sites are sufficiently far apart to avoid double counting, we have

$$L(\boldsymbol{\alpha}, \boldsymbol{\theta} | \{\mathbf{y}_i\}) = \prod_{i=1}^s L(\boldsymbol{\alpha}, \boldsymbol{\theta} | \mathbf{y}_i).$$

In studying avian populations, the authors used the half-normal detectability function. The choice of the detectability function $g(y; \boldsymbol{\theta})$ is discussed in Sect. 8.2.5. If we do not use covariates, we can also assume $N_i \sim \text{Poisson}(\lambda)$. If $\hat{\lambda}$ is obtained from an integrated likelihood, and N_{tot} is the sum of the N_i over all the s locations, we use the estimate $\hat{N}_{\text{tot}} = s\hat{\lambda}$ (Royle 2004b). Sillett et al. (2012) extended the model slightly to estimate habitat-specific abundance and rangewide population size of a bird species of management concern, the island scrub jay (*Aphelocoma insularis*).

Royle et al. (2004) noted that the above methods can also be applied to line transect methods described below using points on the transect as sites. They also have an **R** package for carrying out the computations. Goodness of fit was evaluated for the best fitting Poisson model using the parametric bootstrap procedure based on model deviance. This indicated no significant lack of fit ($P = 0.726$). A negative binomial distribution was also considered. They noted that there can be problems when a particular covariate affects both detection and abundance.

The unconditional approach was also used by Alldredge et al. (2007b) who focused on accounting for variation in detection probabilities associated with the singing rates of birds. As above, their model also accounts for both availability and detection by modeling the *combined* probability that a bird sings during the count and the probability that it is detected when it sings. If the presence of a bird is not

recorded, it means that it did not sing during a given interval (i.e., was not available) or it sang but was not detected by the observer. In open habitats such as grasslands, detection may be primarily visual, such that availability depends on a bird being present and not hidden from view. However, in heavily vegetated habitats, such as forests, detections are primarily auditory.

The authors surveyed several species at the same time, so that if there are some common parameters, data can be pooled and the number of population parameters reduced. Covariates can be used and can, for example, include species and distance. They used the program **DISTANCE** for computations. Alldredge et al. (2006) also considered incorporating multiple observers. Mathematical models are not given in the above papers, but overviews of the value of the methods are given, and reference is made to other papers.

7.6.3 *Repeated Counts Model*

Sólymos et al. (2018) formulated and evaluated a continuous time removal for avian point counts by estimating the availability probabilities of 152 terrestrial bird species using a large database of point-count surveys conducted across boreal and hemiboreal regions of North America.

7.6.4 *Spatially Replicated Counts Model*

Royle (2004b) developed a general model using birds as an example. He said that the key idea was “the introduction of a metapopulation model that characterizes the (spatial) variation in abundance of the spatially referenced populations being sampled.” By introducing a very general model, he included a number of different methods under the same umbrella, which we now briefly consider.

Let N_i be the number of birds available to be counted at site i ($i = 1, \dots, s$). Sampling provides a vector of counts $\mathbf{y}_i = \{y_{ik}\}$ for each sample location. The precise nature of the data vector \mathbf{y}_i depends on the sampling protocol used such as the following.

- (1) For independent double- or multiple-observer protocols (see Sect. 7.7 on double observers below), k indexes an “observer detection history.” With $k = 3$, we have two independent observers and three sets of data—two sets each seen by just one observer and the third by both, giving rise to a three-dimensional \mathbf{y}_i .
- (2) For a removal protocol, k indexes the time interval of (first) detection, i.e., y_{i1} is the number of birds first seen in interval 1, y_{i2} in interval 2, y_{i3} in interval 3, and so on.

- (3) For distance sampling, the count statistics are indexed by distance, so that y_{ik} is the number of birds seen in distance class k at site i .
- (4) For conventional capture-recapture experiments, the data structure is analogous to that obtained under multiple-observer sampling except that the capture history is organized in time giving \mathbf{y}_i , a vector of capture histories.

The above general approach leads to standard multinomial distributions. With a metapopulation approach, we assume that the N_i have a discrete distribution $g(N \mid \boldsymbol{\theta})$, for example, a popular one is Poisson(λ) for a uniform distribution of individuals. For a non-uniform distribution, we can use a negative binomial distribution or introduce covariates or include a random effect such as $\log \lambda_i = \beta_0 + \epsilon_i$, where $\epsilon_i \sim \text{Normal}(0, \sigma^2)$. Further details are given by Royle (2004b), and it is clear that such a general inclusive approach can be useful in developing computer packages.

7.7 Double Observers

Double observers are used extensively in this chapter and the next on line transects. Having two observers means that detection probabilities can be estimated more directly. One method is to designate one person as the primary observer who collects data on those seen and heard at the observation site and designate the other as the secondary observer who records all detections by the primary observer (passed on to the secondary observer) and those not detected by the primary observer as well. Observers can alternate between the two roles. This method is usually referred to as a “dependent” method. For example, Nichols et al. (2000) alternated roles of the two observers and followed the method and notation of Cook and Jacobson (1979). The counts for the primary observer included all birds detected, whereas the counts for the secondary observer included only birds detected by this observer that were missed by the primary observer.

The above method is often less efficient (i.e., larger standard deviation) than with independent observers, but it does exclude errors and time constraints associated with the matching process. The process of communicating observations during the count can be cumbersome and distracting. Although the method perhaps belongs to Chap. 12 on capture-recapture where the two-sample method is discussed in detail, we see below that distances can be introduced via covariates. A major advantage of this approach is that it uses both observers’ data to provide information about each observer’s probability of detection. The theory follows Nichols et al. (2000) and Cook and Jacobson (1979). It is developed further in Chap. 8 where line transects are used.

7.7.1 Theory

Let p_i be the detection probability for observer i , which is assumed to be the same whether observer i is serving as the primary or the secondary observer. Roles are swapped over about half-way through, and the authors referred to the part of the survey in which observer i is the primary observer as the i th half, $i = 1, 2$. Let N_i denote the true number of birds exposed to the sampling procedure in the i th half. The observers are assumed to be independent. Let x_{ij} be the number of individuals counted by observer i ($i = 1, 2$) when observer j ($j = 1, 2$) was the primary observer. The counts for the primary observer included all birds detected, whereas the counts for the secondary observer included only birds detected by this observer that were missed by the primary observer.

The joint distribution of (x_{11}, x_{21}) is the product of $\text{Binomial}(N_1, p_1)$ and $\text{Binomial}(N_1 - x_{11}, p_2)$, while the joint distribution of (x_{22}, x_{12}) is the product of $\text{Binomial}(N_2, p_2)$ and $\text{Binomial}(N_2 - x_{22}, p_1)$. Assuming that the pairs (x_{11}, x_{21}) and (x_{22}, x_{12}) are independent, the joint distribution for all four random variables is the product of the four binomial distributions. We need to estimate $N = N_1 + N_2$, the area covered by the survey, which is $s\pi W^2$ if s circular plots are surveyed.

Let p_i be the probability that a bird is observed by observer i ; then the probability that a bird is detected by at least one observer is $p = 1 - (1 - p_1)(1 - p_2)$. Then conditioning the data on being detected by at least one observer, the product of the four binomial variables can be rewritten as

$$\text{Bin}(N_1, p) \text{Bin}(x_{11}, +x_{21}, p_1/p) \text{Bin}(N_2, p) \text{Bin}(x_{22} + x_{12}, p_2/p).$$

Using the two terms not involving the N_i , we obtain the following maximum likelihood estimators (Cook & Jacobson 1979), namely,

$$\hat{p}_1 = \frac{x_{11}x_{22} - x_{12}x_{21}}{x_{11}x_{22} + x_{22}x_{21}}, \quad \hat{p}_2 = \frac{x_{11}x_{22} - x_{12}x_{21}}{x_{11}x_{22} + x_{11}x_{12}}, \text{ and } \hat{p} = 1 - \frac{x_{12}x_{21}}{x_{22}x_{11}}.$$

We note that \hat{p}_2 follows from \hat{p}_1 by interchanging subscripts 1 and 2. The above estimates can also be obtained as moment estimates by solving the equations from the expected values of the four variables. If N_s , where s the number of points for the two halves is even, is the total number of animals on the sample area $s\pi W^2$, then a natural estimate of N_s is

$$\hat{N}_s = \frac{x_{..}}{\hat{p}},$$

where $x_{..} = x_{11} + x_{12} + x_{22} + x_{21}$. An associated variance estimator is

$$\widehat{\text{var}}[\hat{N}_s] = \frac{(x_{..})^2 \widehat{\text{var}}[\hat{p}]}{\hat{p}^4} + \frac{x_{..}(1 - \hat{p})}{\hat{p}^2}.$$

A confidence interval can be based on the estimated number of birds not detected, $\widehat{f}_0 = \widehat{N}_s - x_{..}$ (Chao 1989; Rexstad & Burnham 1991). The $\log(\widehat{f}_0)$ is treated as an approximately normal random variable, yielding the following 95% confidence interval $(x_{..} + \widehat{f}_0/C, x_{..} + \widehat{f}_0C)$, where

$$C = \exp \left(1.96 \left[\log \left(1 + \frac{\widehat{\text{var}}[\widehat{N}_s]}{\widehat{f}_0^2} \right) \right]^{1/2} \right).$$

Nichols et al. (2000) made a number of useful comments about possible extensions. Most North American bird counts contain many species for which only small numbers of individuals are detected, making estimation of species-specific detection probability problematic or impossible. In an attempt to reduce the number of parameters to be estimated and thus reduce variances, they suggested considering models in which detection probabilities are similar for the two observers and for species, leading to grouping data. They also mentioned some practical problems involved in keeping the observers physically independent. For example, one suggestion is to have a third person as a recorder for the first observer. Also, observer training is important, especially with different species involved and hard-to-detect birds. The above model is essentially a two-sample removal capture-recapture model with the first observer doing the “removing” (Sect. 12.1).

A second approach is to assume that the two observers are on an equal footing in that they each record their own observations. In the past, it has usually been assumed that the two observers are independent (FI—full independence) so that we have a two-sample capture-recapture model. However, sighting probabilities are going to be heterogeneous for at least one factor such as distance as well as other possible covariates. Laake et al. (2011) used the first method and recorded the distances into two groups (“bins”), (0–50] meters and (50–100] meters. A removal model was fitted with distance as a covariate. They also noted that observer independence does not guarantee statistical independence as any unmodeled heterogeneity can lead to dependence.

Point Independence

The detection probabilities of two observers may be correlated even when the observers act independently and objects are always available. For example, if an object is a long way from the site, then it is more likely to be detected if, say, it is a large object. Thus, if observer 1 detects an object with a long perpendicular distance, it is likely to be a large object, and, because it is a large object, observer 2 is also more likely to detect it. An important idea is known as “point independence” (PI) that assumes that there is no unmodeled heterogeneity (i.e., that there is independence) at a given point, typically at the observation point (cf. Burt et al. 2014, for a helpful discussion of this problem).

Dependence

Forcey et al. (2006) compared the two double-observer methods for bird species and preferred the dependent method. It resulted in higher observer-specific and joint detection probabilities, as well as lower standard errors of detection probability across most cover types. The observers also used double sampling (Sect. 3.2.3). A major advantage of this approach is that it uses both observers' data to provide information about each observer's probability of detection. Kissling and Garton (2006) introduced a method they called the paired-observer, variable circular plot (POVCP) to estimate detection probabilities and effective area surveyed. They examined differences in detection probabilities at the plot (site) center, the effective area surveyed, and the densities for five bird species in closed-canopy forests of Southeast Alaska.

Four survey methods were compared: single-observer fixed-radius (50-m) plot; single-observer, variable circular plot (SOVCP); double-observer fixed-radius (50-m) plot; and POVCP. Average detection probabilities for paired observers increased by about 8% (standard error = 2.9%) for all species once estimates were corrected for birds missed at the site center. Density estimators of fixed-radius survey methods were likely negatively biased, because the key assumption of perfect detection was not met, which turned out to be the case. They developed logistic regression models to estimate correction factors to account for incomplete detection at the site center for 32 of 35 combinations of observers and species and concluded that the POVCP method had least bias and the highest precision. The double-observer approach assumes that the same bird is potentially detectable by both observers which is probably not always true and potentially introduces a serious bias in estimating detection probabilities. This availability bias is present in all methods involving two or more observers.

Strickfaden et al. (2020) found that the dependent method reduced the number of false-positive errors, that is, detections of individuals that are absent but counted as present because of misidentifications or double counts.

7.7.2 Unreconciled Double-Observer Method

Riddle et al. (2010) introduced the “unreconciled double-observer” method, an independent double-observer method that does not require observers to match or otherwise reconcile individual observations, but they need to be independent. They focused particularly on bird counts. The modeling of the counts uses the repeated counts model (Royle 2004a) discussed in Sect. 3.2.6. Here, each observer functions as a “visit.” Because both “visits” occur simultaneously, the detection estimate provided by this method is only for p_d (the probability of detection given present and available), even though the mathematics involved is the same as with the repeated counts models. Royle and Dorazio (2009, 177) were the first to suggest

that unreconciled independent multiple-observer counts could be modeled in this way.

Riddle et al. (2010) gave the following simple model for y_{ij} , the observed counts at site i by the j th observer, with associated probability of detection of p_{ij} , and N_i being the abundance at site i , namely,

$$y_{ij} \sim \text{Binomial}(N_i, p_{ij}), \text{ and } N_i \sim \text{Poisson}(\lambda).$$

They compared estimates of the detection probability and abundance from the unreconciled double-observer method with estimates from the traditional independent double-observer method (which requires matching of individual animals) using field-simulated bird-point-count data. The unreconciled double-observer method provided point estimates of detection probability and abundance that were essentially identical to the results of the independent double-observer method, even though much less effort was required to collect data with the unreconciled method. However, estimates of detection probability from the unreconciled double-observer method were usually less precise, and estimates of abundance were always less precise than those from the independent double-observer method. This was because there is less information available in the unreconciled double-observer approach. They found that the method did not work with some species.

7.7.3 Aerial Counts

Much of the early work on double observers was with aerial censusing (Cook & Jacobson 1979). Various configurations of observers that can be used (e.g., those occupying front and back seats) are mentioned, for example, in Sect. 8.12. This was also done by Koneff et al. (2008) with aerial waterfowl surveys, who also introduced visibility factors to adjust counts for incomplete detection. Because of rugged, remote terrain, ground counts were out of the question, and, instead, helicopters were traditionally used to conduct slow, hovering counts on a subsample of survey segments. These were designed to flush birds or otherwise cause bird movement and aid in detection. They used the independent double-observer method where a rear-seat observer collects an independent count, which is reconciled to the count of each front-seat observer to build an encounter history for each observation (i.e., a bird or group of birds). To avoid difficulties encountered when reconciling counts post-survey, they used a technique referred to as on-the-fly reconciliation in which front- and rear-observer recordings were matched immediately rather than later.

Let p_1 and p_2 be the respective detection probabilities of observer 1 in the front seat and observer 2 in the rear seat, given the presence of the bird. Let p_{ik} be the probability that observer i detects a group of size k , which is modeled as

$$\text{logit}(p_{ik}) = a_i + \beta(k - 1).$$

Because of the nose-cone configuration of the aircraft, the visible area of each sampled transect was greater for the front observers than for the rear-seat observer, though front-seat observers were the authors' main concern. This led to dividing the total sampled area of a transect into two subareas A and B, where A was that area exposed to sampling by the front-seat observer only (i.e., the region of increased forward visibility) and B was that area exposed to sampling by both the front- and rear-seat observers. The probability that a bird is in B is ψ and for those in A ($1 - \psi$). Detections by the two observers can be described as 11, and 10, 01, being observed by just one. Using

$$\Pr(X) = \Pr(X|Y)\Pr(XY) + \Pr(X|\text{not } Y)\Pr(\text{not } Y)$$

we have

$$\begin{aligned}\Pr(11) &= p_1 p_2 \psi + 0(1 - \psi) = p_1 p_2 \psi = p_{11} \\ \Pr(10) &= p_1(1 - p_2)\psi + p_1(1 - \psi) = p_1(1 - p_2\psi) = p_{10} \\ \Pr(01) &= (1 - p_1)p_2\psi + 0(1 - \psi) = (1 - p_1)p_2\psi = p_{01}.\end{aligned}$$

If we set $P_2 = p_2\psi$, we have the standard two-sample capture-recapture model where P_2 is the apparent detection probability of the rear-seat observer and represents the confounding of the rear-seat observer's usual conditional probability of detection with the parameter ψ . The front-seat observer's detection probability, which is of primary interest ($p_{11} + p_{10} = p_1$), is not confounded by ψ . If n , the total number detected, consists of x_{11} , x_{10} , and x_{01} , then conditional on n , we have a multinomial distribution.

The main interest of the authors was to detect groups and extended the model by replacing p_i by p_{ik} for observer i and group size k and using the logistic model

$$\text{logit}(p_{ik}) = \alpha_i + (k - 1)\beta.$$

We note that when $k = 1$, we have $\alpha_i = p_i$. A variety of models relating to position, observer, and aircraft effects were considered. Distance sampling was not found to be helpful because of logistic problems.

7.8 Components of Detectability

The probability of detection P can be split up into three possible components: the probability that an individual associated with the sample area is present during the count (P_p), the probability that an individual is available (e.g., with a bird, vocalizing or not visually obscured) given that it is present ($P_{a|p}$), and the probability that an individual is detected given that it is present and available ($P_{d|a,p}$). The full detection process can therefore be represented as $P = P_p P_{a|p} P_{d|a,p}$. From

conditional probability arguments, we have $P_{ap} = P_{a|p}P_p$, $P_d = P_{d|ap}P_{ap}$. For ease of exposition, we shall, at times, write more briefly $P = P_pP_aP_d$. As described by Riddle et al. (2010), various methods for estimating one or more components of P are now available, including distance measurements, multiple observers, time-of-removal data, time-of-detection data, and repeated counts. The terms “availability” and “presence” are often confused and merged as one term in the literature. We consider these below.

Distance sampling and multiple-observer methods provide only estimates of P_d and assume that P_p and $P_{a|p}$ are both equal to one or are constant among sites or study areas. Time-of-removal and time-to-detection data provide an estimate of the product $P_aP_{d|a}$ and assume that P_p is equal to one or is constant among sites or study areas. Separate estimates of P_a and $P_{d|a}$ are not available unless the methods are combined with distance or multiple observers (e.g., Diefenbach et al. 2007).

Repeated counts methods provide the full estimate of $P_pP_aP_d$, but none of the components of the detection process are separable unless this method is combined with time of removal, or time of detection and distance, or multiple observers. The population size provided by repeated counts is actually a superpopulation estimate that may be difficult to translate or relate to habitat area or bird density in many instances. This means that the population sampled with this method includes all birds that have territories that overlap the survey area even if some birds were not present in the survey area on each visit. It is clear that to obtain estimates of the separate components, some of the methods need to be combined.

Nichols et al. (2009) added a fourth component, the probability that the bird’s home range includes at least a portion of the sampling unit. This is generally dealt with through study design.

Riddle et al. (2010) presented the following modified point-count method with a focus on birds that allowed one to estimate two components (availability and detection given availability) of the detection process using a combined dependent double-observer time-to-detection approach. We recall that in the dependent double-observer approach, the primary observer records all birds seen or heard. The secondary observer records all birds detected by the primary observer, but also records birds he or she detects that the primary does not. Observers reverse roles with each new point count. For example, with four time intervals, the observers’ detection history might be (11,01,00,11) so that in the first time interval, both detected a bird, while in the second time interval, only the second observer detected a bird, and so on. In combining the models, suppose there are t time intervals and two observers (usually $t = 3–5$ so that heterogeneity models could possibly be fitted). This combined method is equivalent to Pollock’s (1982) robust design capture-recapture model with t primary periods (the time intervals) and two secondary periods (the observers) within each primary period.

Using the idea of a classic random temporary emigration model (cf. Seber & Schofield 2019, Section 11.4), γ_i is the probability that an animal is a temporary emigrant in period i , and γ_i does not depend on its value in previous periods. Riddle et al. (2010) suggested that γ_i may be thought of as the probability that a bird is unavailable for detection in interval i so that we can write $p_{ai} = (1 - \gamma_i)$

($i = 1, \dots, t$) as the probability a bird is available in interval i . They also include conditional detection probabilities for each observer in each period, namely, (p_{d1i}, p_{d2i}) for all i . The survival probabilities in the emigration model are all equal to one. For ease of exposition, we consider just $t = 2$ time intervals so that the detection histories are 11, 01, 00 for each primary observer, which gives us a total of $k = 3^t = 9$ for each primary observer or a total of 18 cell probabilities. For example, the detection history (11,01) has cell probability

$$p_{a1}p_{d11}p_{a2}(1 - p_{d12})p_{d22},$$

so that the bird has to sing in each interval to be detected by at least one observer. Another history (11, 00) has probability

$$p_{a1}p_{d11}[p_{a2}(1 - p_{d12})(1 - p_{d22}) + (1 - p_{a2})],$$

where 00 in the second interval means there were two components for the probability, the first where the bird sings but is missed by both observers and the second where the bird does not sing. A third example history (00,00) has for the first observer being the primary observer the cell probability

$$\{[p_{a1}(1 - p_{d11})(1 - p_{d21})] + (1 - p_{a1})\} \{[p_{a2}(1 - p_{d12})(1 - p_{d22})] + (1 - p_{a2})\}.$$

Let n_{ij} ($i = 1, \dots, k - 1$) be the number that have the i th detected history when observer j is the primary observer. Then the likelihood function for the population size N is the product of two multinomial distributions

$$\begin{aligned} L(N; \boldsymbol{\theta}) &= \frac{N! \tau}{n_{11}! \cdots n_{k-1,1}!(N-n)!} [p_{11}(\boldsymbol{\theta})^{n_{11}} \cdots p_{k-1,1}(\boldsymbol{\theta})^{n_{k-1,1}}] \\ &\times \frac{(1-\tau)}{n_{12}! \cdots n_{k-1,2}!} [p_{12}(\boldsymbol{\theta})^{n_{12}} \cdots p_{k-1,2}(\boldsymbol{\theta})^{n_{k-1,2}}] p_k(\boldsymbol{\theta})^{N-n}, \end{aligned}$$

where $n = \sum_{i=1}^{k-1} \sum_{j=1,2} n_{ij}$ is the total number of detected birds; p_{ij} represents the multinomial cell probabilities that are known functions of $\boldsymbol{\theta}$, availability and detection probability parameters; and $p_k = 1 - \sum_{i=1}^{k-1} \sum_{j=1,2} p_{ij}$. Also, τ is the fraction of the detection histories where observer 1 is the primary observer (often 0.5 by design), and, thus, $1 - \tau$ is for detection histories where the roles are reversed. The reason why $k - 1$ is used is that it is assumed that the k th detection history is the undetected history when either observer 1 or 2 is the primary observer. This can be done because in any time period when you have no detection, the probability is the same regardless of which observer is the primary observer.

We can now put it all together under the framework of the capture-recapture model M_t of Sect. 12.5.1, namely,

$$L(N; \boldsymbol{\theta}) = L_1(N; p_k(\boldsymbol{\theta})) \times L_2(\boldsymbol{\theta}),$$

where

$$L_1(N; p_k(\boldsymbol{\theta})) = \frac{N!}{n!(N-n)!} [1 - p_k(\boldsymbol{\theta})]^n p_k(\boldsymbol{\theta})^{N-n}$$

$$L_2(\boldsymbol{\theta}) = \frac{n}{n_1! \cdots n_{k-1}!} q_1(\boldsymbol{\theta})^{n_1} \cdots q_{k-1}(\boldsymbol{\theta})^{n_{k-1}}$$

with

$$q_i(\boldsymbol{\theta}) = \frac{p_i(\boldsymbol{\theta})}{1 - p_k(\boldsymbol{\theta})}, \quad i = 1, \dots, k-1.$$

Maximum likelihood estimates of the availability and detection probabilities (i.e., of $\boldsymbol{\theta}$) can be obtained by maximizing L_2 . Then, from Sanathanan (1972a), the maximum likelihood estimate \widehat{N} of N the population size is obtained by maximizing L_1 . For any given \widehat{p}_k , this is the greatest integer less than or equal to $n/(1 - \widehat{p}_k)$. Optimization can be carried out using, for example, *constrOptim* in **R**, which constrains probability estimates to be between 0 and 1. The authors applied the method to northern bobwhite birds as well as to some simulated data having low availability and detection probabilities, using a wide range of models, with some variations such as random availability.

Pollock et al. (2004, 2006) considered the separate estimation of the components P_a and $P_{d|a}$ and applied their method to the aerial census (using double observers) of the marine mammal dugong, which may be in the sampling area but mostly underwater and therefore not available. Even if available, it may not be detected (lack of “perception”). The authors used the term probability of perception $P_{d|a}$, the conditional probability of a dugong being detected by at least one observer, given it is available, and referred to the (unconditional) detection probability $P_d = P_{d|a}P_a$. Dugongs have a patchy distribution and often occur at very low densities. Here, \widehat{P}_a , an estimate of P_a , is difficult to obtain and must be done externally to the aerial survey using additional data collected in a dedicated study, e.g., using artificial dugong models (Marsh & Sinclair 1989). The two-sample Petersen capture-recapture model (Sect. 12.2) applied here giving \widehat{p}_i , the estimate of being seen (“captured”) by observer i ($i = 1, 2$) and the estimator

$$\widehat{P}_{d|a} = (1 - \widehat{p}_1)(1 - \widehat{p}_2).$$

Let p_{area} be the proportion of the population area sampled, and let $\widehat{P}_d = \widehat{P}_a \widehat{P}_{d|a}$. If n is the number of animals seen, the population size N can be estimated by

$$\widehat{N} = \frac{n}{p_{\text{area}} \widehat{P}_d}.$$

The method is readily extended to groups of animals where n groups are detected with g_i animals in group i ($i = 1, \dots, n$). Then Pollock et al. (2004) gave the

generalized Horvitz-Thompson estimator

$$\widehat{N}_g = \sum_{i=1}^n \frac{g_i}{p_{\text{area}} \widehat{P}_{id}}.$$

An estimate of its standard error can be obtained using simulation or a parametric bootstrap method. The authors considered further applications to birds and salamanders with regard to availability, e.g., birds not singing and salamanders below the ground for a cool and moist environment.

Mordecai et al. (2011) also considered a model using the idea of availability in studying episodic or mobile species by applying a multi-scale design, with waterthrushes used as an example. Here, availability a_{it} (their u_{ij}) is the probability that at site i the species is available for detection at time t .

As noted by Simons et al. (2009), we have seen that different methods estimate different components of the detection process. For example, distance sampling and multiple-observer approaches assume that all birds on a given sample plot are available (sing during the count interval) and they estimate the probability of detection given availability. Time-to-detection methods provide estimates of the product of availability and detection given availability, but they cannot separate the two components. Repeated count methods estimate the product of availability, detection given availability, and a third component of the detection process, the probability that an individual is present in the sample area.

Schmidt et al. (2013) developed a model, with a focus on passerines to estimate $P = P_p P_a P_d$. Let n_{ijt} represent the observed counts at each point (site) $i = 1, \dots, I$ points, during each repeated count $j = 1, \dots, J$ counts, in each year $t = 1, 2, \dots, T$ years. Let N_{it} be the site abundance, assumed to be Poisson (λ_{it}). Let $z_{it} = 1$ if present with probability ψ_{it} and 0 otherwise, and let p_{ijt} be the detection probability, given present. We thus have the following sequence of probability distributions, including a binomial mixture:

$$\begin{aligned} n_{ijt} &\sim \text{Binomial}(N_{it}, p_{ijt} z_{it}) \\ z_{it} &\sim \text{Bernoulli}(\psi_{it}) \\ N_{it} &\sim \text{Poisson}(\lambda_{it}). \end{aligned}$$

Here, $\text{logit}(p_{ijt})$, $\text{logit}(\psi_{it})$, and $\log(\lambda_{it})$ are expressed as linear models using covariates, regression parameters, and random effects. Two covariates addressed variation in P_p of males within territories and P_a due to singing behavior. The regression parameters were given prior normal distributions. A similar type of model was considered by O'Donnell et al. (2015) for salamanders.

7.8.1 Validation and Comparison of Point Methods

As there are few studies where the population is completely known, we mention Nelson and Fancy (1999) who used radio telemetry to locate a known population of birds that were investigated using the point transect method. Numbers were small, and they concluded that for some species, movement during the count period, large errors in distance estimates, undercounting, or misidentification makes it impossible to obtain accurate density estimates. For species with a large effective detection radius, interstation distances may have to be calibrated to reduce double detections.

Alldredge et al. (2008) developed a clever field simulation method to consider the effectiveness of two methods, namely, the use of distance sampling with three different detection functions and the use of independent observers. They simulated a field survey that mimicked bird calling in a field environment. The system allowed them to vary singing rate, species, distance, the complexity of points, and other factors. They found that common assumptions about the accuracy of distance estimation, and the ability of observers to accurately map and match observations, are often not met on auditory point counts. This is serious as aural detections often comprise 95% of all bird detections in forested habitats (Simons et al. 2007). For the double-observer method, two pairs of observers tended to underestimate true population size, while the third pair tended to double-count birds, which overestimated the population. Detection probabilities were always higher and population estimates lower when observers subjectively matched birds compared to an objective rule and showed a slight negative bias and good precision.

Thompson et al. (2008) compared indices of abundance estimated as annual means of counts and from a mixed Poisson model to abundance estimates from a count removal model with three time intervals and a distance model with three distance bands. This led to five methods, and trend estimates were calculated from an autoregressive, exponential model fitted to annual abundance estimates from the previous methods and also by estimating trend directly by treating year as a continuous covariate in the mixed Poisson model. This led to 12 comparisons and was done for 6 forest songbird species and 2 physiographic areas and from 1997 to 2004. Although there was strong evidence that detection probabilities varied among species and years, there was “good overall agreement across trend estimates from the 5 methods for 9 of 12 comparisons.” It was found that “estimates of detection probabilities differed greatly between removal and distance models, no doubt because the methods estimated different components of detection probability and the data collection was not optimally designed for either.”

Murray et al. (2011) compared three methods using roadside point counts, namely, distance sampling, independent double observers, and removal methods, for whistling northern bobwhites (*Colinus virginianus*). Two observers independently recorded each whistling bird heard, distance from the observer, and time of first detection at 362 call-count stops. Once detected, later detection is ignored so it is effectively “removed.” Covariates for year and observer effects were used for each method, and distances from observer were used as covariates for the double-

observer and removal methods. Although double-observer surveys provided more precise estimates, the authors recommended using the removal method to estimate detectability and abundance of bobwhites. The removal method provided precise estimates of density and detection probability and required half the personnel time as double-observer surveys. Also, the likelihood of meeting model assumptions was higher for the removal survey than with independent double observers. There were some problems such as the probability of detection near the observer was less than 1 or the roadside survey points were not randomly distributed with respect to the observer. Also, distances were not consistently recorded by individual members of observer pairs.

Riddle et al. (2008) noted that distance sampling and multiple-observer methods do not allow estimation of the detection probability $P_a P_{d|a}$, whereas the time-to-detection and repeated count methods do. We recall that P_a is the probability of a bird being available and $P_{d|a}$ is the probability of a bird being detected, given it is available. They compared the time-to-detection method with a logistic regression method (Wellendorf et al. 2004), which also includes estimation of the total detection probability. It used the following covariates: number of calling coveys, wind speed, cloud cover, and changes in barometric pressure prior to the count. They found that both methods gave comparable results for bobwhite covey detection and recommended that both methods be used if possible.

7.9 Combined Methods

As noted by Alldredge et al. (2007a), there are three distinct methods for estimating abundance and modeling the detection process from unrepeated point-count surveys. The first models the probability of detection as a function of distance from the observer, which we have considered above. The second uses multiple-observer methods to estimate the probability of detection by each observer using a removal (primary-secondary observer) or capture-recapture (independent observer) method. The third, referred to as the time-to-/time-of-detection method, estimates the probability of detection over multiple time intervals and is described in Sects. 4.3.5 and 5.4.7. In this case, the method estimates the combined probability that an individual is available for detection and that it is detected given that it is available. These methods can be combined in various ways leading to the various components of detectability discussed in Sect. 7.8 being estimated separately.

Two Methods Combined

Farnsworth et al. (2005) proposed a method, particularly for birds, that combined distance and temporal removal sampling. Here, the birds are “removed” (ignored) after being seen the first time to estimate P , the probability of detection. Here, $P = P_a P_b$, where P_a is the probability a bird vocalizes within a time interval and P_b is the conditional probability a bird is detected, given that it vocalizes.

The authors considered the simplest case, for ease of exposition, where a count is divided into two time intervals of equal length and every bird detected is identified as within a fixed radius or beyond a fixed radius from the observer (double observers could also be used), giving two time intervals and two distance categories. In practice, the authors recommended at least three intervals. The assumptions made are as follows: no moving of birds during the count (i.e., a short count time); P_a is constant over space and time and over the count period; and birds are correctly counted (i.e., properly identified, assigned to the proper distance category, and no double counting). They used the half-normal distance function (with σ^2 instead of $2\sigma^2$), namely,

$$P_b = \exp(-r^2/\sigma^2)$$

for r , the distance from the observer to the bird, and $P = P_a P_b$. Let x_{ij} be the number of birds first recorded in the i th distance category during the j th time interval, giving us four sufficient statistics, and let $x..$ be the sum of the four x_{ij} 's. For example, x_{11} is the number of birds detected within the radius r_1 of the observer during the first time interval. Having a count divided into at least two time intervals allows for the estimation of P_a , and classifying birds detected into at least two distance categories allows for the estimation of the decline in P_b with increasing distance. Let π_{ij} be the probability that a bird is a member of x_{ij} given it is a member of $x..$ Farnsworth et al. (2005) eliminated the unknown density D by conditioning on $x..$, a “trick” we have seen several times in this book. Then the conditional multinomial density function, using the authors’ Appendix 1, is

$$f(\{x_{ij}\} | x..) = \frac{x..}{x_{11}!x_{12}!x_{21}!x_{22}!} \prod_{ij} \pi_{ij}^{x_{ij}},$$

where

$$\begin{aligned} \pi_{11} &= \frac{1 - \exp(-r_1^2/\sigma^2)}{2 - (P_a/2)}, \\ \pi_{12} &= \frac{1 - \exp(-r_1^2/\sigma^2) + (P_a/2) \exp(-2r_1^2/\sigma^2) - (P_a/2)}{2 - (P_a/2)}, \\ \pi_{21} &= \frac{\exp(-r_1^2/\sigma^2)}{2 - (P_a/2)}, \\ \pi_{22} &= \frac{\exp(-r_1^2/\sigma^2) - (P_a/2) \exp(-2r_1^2/\sigma^2)}{2 - (P_a/2)}. \end{aligned}$$

Maximum likelihood estimates $\hat{\sigma}^2$ and \hat{P}_a are found by maximizing the likelihood

$$L(P_a, \sigma | \{x_{ij}\}) \propto (\pi_{11})^{x_{11}} (\pi_{12})^{x_{12}} (\pi_{21})^{x_{21}} (\pi_{22})^{x_{22}}.$$

The authors used the program **SURVIV** to find the estimates and associated variance estimates. They then obtained the density estimate

$$\widehat{D} = \frac{X_{..}}{n\pi\widehat{\sigma}^2(2\widehat{P}_a - (1/2)\widehat{P}_a^2)},$$

where n is the total number of counts. The method will naturally extend to more than two equal time intervals and more than two distance categories. Such parameters will depend on the species under investigation.

Amundson et al. (2014) extended the above method using a Bayesian framework that facilitates evaluating covariates on abundance and detection probability, incorporating excess zero counts (i.e., zero inflation), accounting for spatial autocorrelation, and estimating population density. They broke down the probability of detection into several components. Firstly, there was the coverage probability, the probability that the location of the bird is within a sampling unit at the time of the survey and is the product of the probability that (1) a bird has a home range that overlaps the sampling unit (P_s) and (2) is present in the study area at the time of the survey (P_p). Secondly, the detection probability also consists of two components: availability (P_a), which is the probability that an animal is present in the survey area and signaling its presence to the observer (e.g., a bird is vocalizing or is in view), and perceptibility (P_d), which is the probability that an animal available for detection (e.g., a bird sings) is detected by the observer.

Three Methods Combined

Up till now, we have mainly considered two components of detection, the probability of detecting an individual, given it is available for detection as above, and the probability of being available as discussed above and in Sect. 4.1.3 on occupancy. Hostetter et al. (2019) considered breaking up the probability of detection into three components: probability of presence during a survey (P_p), probability of availability given presence ($P_a | P_p$), and probability of detection given availability and presence ($P_d | P_a, P_p$) as described above. They sampled S plots (sites), assumed to be independent, on $t = 1, \dots, T$ sampling occasions within a period of demographic closure (i.e., a single breeding season). Plot-specific counts were conducted within a short, standardized period (e.g., 10 min) divided into $k = 1, \dots, K$ sub-intervals (e.g., four 2.5-min intervals).

The method used was the point transect (point-count) survey, where the sample units are circular, fixed-radius plots as described above. The authors assumed a superpopulation of individuals during the entire study period N_{sup} , of which N_{sup}^s is the number at site s ($s = 1, \dots, S$). In a particular survey on occasion t , only some of the animals from the superpopulation, N_{pres}^{st} say, will be present within site s . Then, only N_{avail}^{st} of the individuals present at site s during survey t will likely be available for detection.

We require the following data. For animal i ($i = 1, \dots, n$), let r_i be the radial distance from site center to animal i , and let n_{st} be the number of distinct individuals seen on site s at time t . Let $x_{istk} = 1$ if the individual was detected at site

s on occasion t and observed during the k th sub-interval; $x_{istk} = 0$ otherwise ($i = 1, \dots, n_{st}$; $s = 1, \dots, S$; $t = 1, \dots, T$; and $k = 1, \dots, K$). We now have to put all this information together.

For distance data, Hostetter et al. (2019) assumed a half-normal detectability function

$$g(r) = \exp\left(-\frac{r^2}{2\sigma^2}\right).$$

From Sect. 7.2.1, we get

$$P_d = P_W = \frac{2}{W^2} \int_0^W rg(r)dr.$$

This integral can be approximated using a discrete number J of distance bins Kéry and Royle (2016). Let y_{jst} denote the number of individuals observed in distance bin j , at site s , on occasion t and \mathbf{y}_{st} be the vector $(y_{1st}, y_{2st}, \dots, y_{Jst})'$. Let $\mathbf{y} = (y_1, \dots, y_J)'$ denote the vector for the numbers of individuals observed in each of the J bins. The vector \mathbf{y} has a multinomial distribution with conditional cell probabilities $\pi_j^c = \pi_j \phi_j / P_d$, where ϕ_j is the proportion of the sample area within the j th bin, π_j is the probability of detecting an individual that is in distance bin j , and $P_d = \sum_j \pi_j \phi_j$ is the probability of detection across all distance bins, given availability and presence. Let $\boldsymbol{\pi}^c = (\pi_1^c, \dots, \pi_J^c)'$.

Time-of-/time-to-detection data are analyzed with closed population capture-recapture methods using x_{ik} , the detection encounter history for individual i during sub-interval k , which is assumed to be a Bernoulli random variable with conditional probability $P_A^c = P_a' / P_a$. Here, P_a' is the probability of detection in sub-interval k and $P_a = 1 - (1 - P_a')^K$ is the probability of being available at least once during the survey. This can be modeled in various ways depending on how the data are grouped.

We now assume that $N_{\text{sup}}^s \sim \text{Poisson}(\lambda)$ and have the following probability functions:

$$\begin{aligned} N_{\text{pres}}^{st} &\sim \text{Binomial}(N_{\text{sup}}^s, P_p) \\ N_{\text{avail}}^{st} &\sim \text{Binomial}(N_{\text{pres}}^{st}, P_a) \\ n_{st} &\sim \text{Binomial}(N_{\text{avail}}^{st}, P_d) \\ x_{ik} &\sim \text{Bernoulli}(P_a^c) \\ \mathbf{y}_{st} &\sim \text{Multinomial}(n_{st}, \boldsymbol{\pi}^c) \end{aligned}$$

Vague prior distributions were used for the parameters, in particular $\lambda \sim \text{Gamma}(0.1, 0.1)$, $\log \sigma \sim \text{Uniform}[0, 10]$, and P_a' and P_p both $\text{Uniform}[0, 1]$. Models were fitted using Markov chain Monte Carlo (MCMC) methods and **JAGS**

(Plummer 2003) software accessed through R version 3.3.1 (R Development Core Team 2016) with the jagsUI package (Kellner 2015).

In summary, we see that distance data provide information on P_d , time-of-detection data provide information via the capture-recapture method on $P_d P_a$, and repeated spatial and temporal surveys provide information on $P_d P_a P_p$.

Péron and Garel (2018) also combined the three methods, namely, point-distance measurements, multiobserver data, and time-to-detection information, in a complex model where the detection data was stratified across sites, years, within-year visits, robust design-style secondary occasions within visits, observers, and groups of individuals. To compare methods, the authors applied their method to Pyrenean chamois, Mediterranean mouflon, and feral cats.

Some Practical Considerations

A number of basic issues need to be thought through in designing a point-count survey. Some of these were discussed earlier in Sect. 7.2.2. Based on Huff et al. (2000), we give a further list with some additions, which relate to more general planning. Although focused on bird monitoring, they can apply more generally.

- (1) **Planning:** What is the best method for monitoring? Can it detect the sort of changes looked for? What are the long-term resources for monitoring? What habitats need to be studied and will stratification be needed? For birds, it is suggested three breeding seasons of observation for investigating habitat and five to ten more seasons for investigating trends in abundance.
- (2) **Site selection:** What sites should be considered for monitoring, bearing in mind that some sites may be too small? Photographs can be helpful in choosing competing sites that correspond to monitoring goals and have suitable access.
- (3) **Establishing the point-count sites:** The number of sites needs to be determined and needs to be randomly chosen. This is related to choosing the distance from the edge of the boundary location because of edge effect and the distance between points. Locating, accessing, exiting, and choosing the order of the stations should be carefully considered. A permanent marker can be used for each station. Also, how often should sites be revisited?
- (4) **Choose W, the distance out to which individuals are monitored:** One suggested method of choosing W is the distance following 10% truncation of the observations at the largest distances.
- (5) **Conducting the point counts:** Trained qualified observers who have been tested (for sight and sound if necessary) should be used as proper training for observers is essential, especially with volunteer observers in citizen science studies. The time of the day and weather conditions for the counts can be critical. Time spent counting needs to be carefully chosen. With regard to birds, there is the question of dealing with flyover birds. In some cases, juveniles, males, and females may need to be recorded separately.
- (6) **Good methods of recording data should be available.** Some thought is needed as to what methods and software will be used for data analysis. Automated sound recordings are being used more (cf. Brandes 2008) and are useful for bird species at risk (Holmes et al. 2014).

- (7) If individuals occur in clusters, the theory applies to the number of clusters. An estimate is then needed of the average cluster size to estimate the total number of individuals.

7.10 Summary

This chapter considers two closely related topics. The first consists of choosing a point by some random process such as totally at random or as point on a line transect or some linear feature such as a road or track. The number of animals (or clusters) or objects (e.g., animal signs) observed or heard is counted in a circular plot centered on the point with radius W . The second method also includes the distances from the center to each individual. The aim is to estimate population numbers or density using full or conditional likelihood functions.

A number of variations are given such as (1) increasing W until a certain fixed number of individuals are counted, (2) simply counting all individuals observed so that we effectively have $W = \infty$, and (3) grouping distances into distance categories (referred to as concentric bands or “bins”) or into just two categories, up to distance W and greater than W . Repeat visits can be made to the original points to provide additional data, and the time between visits is critical.

Mathematical theory for the methods is given, and heterogeneity is considered. Designing such an experiment is discussed such as choosing the number of points, dealing with points close to the boundary of the population area, and possible measurement errors. Using a model-based approach, a variety of probability of detection functions are proposed. The distance sampling can also be combined with a removal method.

As some species are insufficiently visible or noisy to allow adequate numbers to be detected by observers standing at random points, the point counts or distances can be combined with capture-recapture using a trapping web. A single trap or lure method can also be used. Standard or “unreconciled” double-observer techniques are described, along with aerial methods. With more data, detectability can be split into components involving availability, presence, and observance. Various methods are compared, as well as how to combine some of them.

Chapter 8

Line Transect and Distance Methods



8.1 Line Distances

In line transect sampling, an observer travels a total distance L across an area A containing a population of N animals in non-intersecting and non-overlapping lines counting n , the number of animals, or the number of animal clusters sighted out to a distance W . A cluster is defined to be a group of animals with a well-defined location for the group center. For each animal or cluster detected, the observer measures or estimates the perpendicular distance y of the animal or cluster center from the nearest part of the line. In many surveys, it is easier to measure or estimate the observer-to-animal (radial) distance r at the time the detection is made. If the sighting angle θ is also measured, then the perpendicular distance is $y = \sin \theta$.

This method can be regarded as plot sampling as in Chap. 2 with the plot consisting of a strip of area $2WL$ containing N_L animals, but with the difference of having incomplete detectability. The strip or detection area is often referred to as the “covered area.” A transect can be traveled continuously, or regular stops can be made as it takes time to make observations. Line transect surveys can be useful for the estimation of animal densities over large areas or if budgets are limited. They can also be used for sampling groups of all kinds of animals (e.g., Liu et al., 2008, blue sheep).

If there is incomplete detection, there are two approaches. We can model the detection probability using logistic regression with covariates, or we can measure detection distances from the line transect. It is then assumed that the probability of detection falls off smoothly with distance from the line and to use the recorded distances y of detected animals from the line to model the detection function, $g(y)$. This is defined to be the probability of detecting an animal that is at a perpendicular distance y ($0 \leq y \leq W$) from the transect. The theory for this parallels the point transect theory of Chap. 7 so that the two are often considered together in the literature. We now look at the line transect distance theory.

Suppose that the transect(s) is randomly located. Consider the random variable I_i taking the value 1 if the i th animal is sighted and 0 otherwise, and let P_L be the probability it is sighted. Then $E[I_i] = P_L$ and

$$E[n] = E\left[\sum_{i=1}^{N_L} I_i\right] = N_L P_L \text{ or } N_L = E[n]/P_L.$$

This result does not depend on the assumption of a random distribution of the animals, but a variance estimate does require further assumptions. From the above equation, we see that to obtain an estimate of the actual number N_L on the covered area, n has to be divided by an estimate of the probability of being detected there, P_L . To obtain this estimate, the observer records one or more of the following statistics at the time of first sighting:

- (a) The radial distance r_i ($i = 1, \dots, n$) from observer to animal
- (b) The right-angle distance y_i from the animal sighted to the path of the observer
- (c) The angle of sighting θ_i from the observer's path to the point at which the animal was first sighted

An obvious use for the method is for populations in which animals, and particularly birds, can only be seen when they are disturbed and flushed into the open. In this case, r_i is called the “flushing” distance.

8.1.1 Basic Assumptions

The basic assumptions underlying the above method are:

- (1) The lines are placed independent of animal locations. This can be difficult if the observer walks along well-traveled paths or drives along a road. Independent lines are best used for walking along.
- (2) The sighting of one animal is independent of sighting another.
- (3) No animal is counted more than once.
- (4) When animals are seen through being flushed into the open, each animal is seen in the exact position it occupied when startled.
- (5) The response behavior of the population as a whole does not substantially change in the course of running the transect.
- (6) The probability of an animal being seen, given it is a right-angle distance y from the line transect path (irrespective of which side of the path it is on), is a simple function $g(y)$, say, of y , such that $g(0) = 1$ (i.e., probability 1 of seeing an animal on the path). This assumption can be removed as we see later. Some earlier methods of estimating $g(0)$ were given by Laake and Borchers (2004).
- (7) The animals are homogeneous with regard to their response behavior, regardless of sex, age, etc. This implies that $g(y)$ is the same for all animals.

- (8) Distance measurements are accurate. One should use a laser rangefinder and have observers adequately trained. There is evidence that the detection probability may increase when observers gain more experience. However, data can be put in “bins” or distance classes, which occurs in some models in this chapter. Buckland et al. (2015, 67) comment that little precision is lost by binning data. However, information is lost for assessing whether assumptions hold and whether models for the detection function provide an adequate fit. If it is feasible to estimate exact distances, then such data should be recorded using aids (e.g., laser rangefinders, GPS, clinometers, reticles) to reduce estimation error.
- (9) It is assumed that $g(y)$ reduces smoothly with increasing distance from the line. Required properties of this function are discussed further in Sect. 8.2.5.

Some of these assumptions will be relaxed later. In the meantime, two comments need to be made. First, if the transect is located at random, each animal has the same probability of being in the coverage area, and its actual perpendicular distance from the line has a uniform distribution with density function W^{-1} on $[0, W]$. This is independent of the distribution of the animals and is a design feature. Second is the anomalous role, sometimes confused in the literature, played by $g(y)$. It is strictly a conditional probability, which is discussed in detail below, and is a model feature.

The theory tends to be described in terms of a single long transect, but in practice, a good number of lines are usually laid down over the population area. With a single transect, it is not possible to make valid estimates of the standard error and confidence limits of estimates for population density and population size because of lack of replication. Although replicate transects should be randomly located within subareas, a systematic (evenly spaced) sample is often used in practice. Systematic survey designs can reduce the true variability among different realizations of the design and can achieve better overall coverage. However, estimating the variance then becomes difficult, and estimators typically approximate the variance by treating the design as a simple random sample of lines.

Stratification can be helpful (Buckland et al., 2015, Section 2.3), such as stratifying by environmental conditions, by cluster size, by habitat, and by observer or platform, though perhaps less relevant with the use of covariates. Buckland, Rexstad, et al. preferred multiple covariate distance sampling.

In some applications, it may turn out that a number of individual lines will yield no observations. Eberhardt (1978) noted that assigning “zero” values to such lines will bias the overall estimate (underestimating the true population size). One practical way to deal with this problem is to design the survey so that a predetermined number of individual lines are combined as a single sampling unit.

Citizen science (cf. Sect. 4.16) has also been used in conjunction with line transects. For example, Dunham et al. (2013) used it for large mammals.

8.2 General Distance Theory

We follow Seber (2001) and let $X = 1$ if a given animal in the observation area is observed and 0 otherwise. If $f_{X,Y}(x, y)$ ($0 \leq y \leq W$) is the joint density function of discrete X and continuous perpendicular distance Y , we can define $g(y | \theta) = f_{X|Y}(1 | y, \theta)$, the probability of detecting an individual given it is at distance y (the so-called detectability function), where θ is an unknown vector parameter, or we condition on it for a Bayesian analysis discussed later. Also, $\pi(y) = f_Y(y)$ is the density function of the distance y irrespective of whether an animal is detected or not. Since the transect is randomly placed, Y has a uniform distribution $\pi(y) = 1/W$ on $[0, W]$. Also, the probability P_L that an animal is detected, given that it is in the area $a = L \times 2W$ of possible detection, is

$$P_L = f_X(1) = \int_0^W f_{X|Y}(1 | y, \theta) \pi(y) dy = \frac{1}{W} \int_0^W g(y | \theta) dy = \frac{\mu_L}{W}, \quad \text{say,} \quad (8.1)$$

and N_L is the number present in the area a . Because of the upper limit of W , the integral $\mu_L = \int_0^W g(y) dy$ will generally need to be found numerically and is sometimes referred to as the “effective strip width” because of (8.4) below, where the estimate of density is then the number seen n divided by the effective area of the transect (effective strip width times the length of the transect).

The density function of y , conditional on the animal being detected, is then

$$f_c(y) = f_{Y|X}(y | 1) = \frac{f_{X|Y}(1 | y) \pi(y)}{f_X(1)} = \frac{g(y)}{W} \frac{W}{\mu_L} = \frac{g(y)}{W P_L}, \quad (8.2)$$

$$f_c(0) = g(0)/\mu_L,$$

where the subscript “c” denotes that the density function is conditional. We note that $f_c(0) = 1/\mu_L$ if $g(0) = 1$, which means that an estimate $\hat{f}(0)$ of $f_c(0)$ gives us an estimate of μ_L and hence of P_L by (8.1). Also, if $P_a = a/A$, the probability of being in the detection area given that it is in the population, then the unconditional probability of detection is

$$\begin{aligned} P &= \Pr(\text{detection} | \text{available for detection}) \Pr(\text{available for detection}) \\ &= P_L P_a. \end{aligned} \quad (8.3)$$

If we are interested in just density, we can estimate the density in the covered area using

$$\widehat{D}_L = \frac{n}{\widehat{P}_L(2WL)} = \frac{n}{2\widehat{\mu}_L L} = \frac{n\widehat{f}}{2L\widehat{g}}. \quad (8.4)$$

Here, \hat{f} is an estimate of $f_c(0)$, \hat{g} is an estimate of $g(0)$ if not equal to 1, and we can then obtain an estimate of N from $N = D_L A$, provided D_L is representative of the overall density D . This will of course depend on the terrain and how the transects are selected. We again iterate that we do not need to assume that the animals are randomly distributed, much as we saw in plot sampling. However, some assumptions are needed for variance estimation. Generally, we have $g(0) = 1$ so that

$$\hat{D}_L = \frac{n\hat{f}(0)}{2L}.$$

If our observation unit is a school with s individuals, as in marine studies, our density estimate is now, from (8.4),

$$\hat{D}_L = \frac{n\hat{E}[s]}{\hat{P}_L(2WL)}.$$

A good example of the use of the above formula for sea birds using vessel transects was given by Ronconi and Burger (2009). This topic is discussed further in Sect. 8.13.1.

Williams and Thomas (2007) described an extensive study using the above results for a number of species of marine mammals including cetaceans (porpoises, dolphins, whales) and pinnipeds (seals and sea lions). A stratified survey design was used where the study area was divided into four strata, within which a sample of equal-spaced zigzag or parallel transect lines was placed. There was with a random start point to ensure equal coverage probability within strata. Zigzag transects are further discussed in Sect. 8.3.1.

8.2.1 Variance Estimation

We now consider an expression for the variance of $\hat{D}_L = n\hat{f}/(2L)$ using the fact that $E[n] = N_L P_L = 2LD_L\mu_L$ (from (8.1)). Assuming that for large n , $E[\hat{f}|n] \approx f_c(0)$, then

$$\begin{aligned} E[\hat{D}_L] &= E_n E[\hat{D}_L | n] \\ &\approx E[n] \frac{f_c(0)}{2L} \\ &= N_L P_L \frac{f_c(0)}{2L} \\ &= \frac{N_L \mu_L f_c(0)}{2LW} \end{aligned}$$

$$= N_L/a \\ = D_L,$$

since $a = 2LW$ and $g(0) = 1$ implies $f_c(0) = 1/\mu_L$. Also,

$$\text{cov}[n, \hat{f}] = E_n(E[(n - E[n])(\hat{f} - f_c(0)) \mid n]) \approx 0.$$

Hence, by the delta method (cf. Appendix A.4),

$$\begin{aligned} \text{var}[\hat{D}_L] &\approx D_L^2 \left\{ \frac{\text{var}[n]}{(E[n])^2} + \frac{\text{var}[\hat{f}]}{[f_c(0)]^2} \right\} \\ &= D_L^2 [C_n^2 + C_{\hat{f}}^2], \end{aligned}$$

or

$$C^2 = C_n^2 + C_{\hat{f}}^2,$$

where C , C_n , and $C_{\hat{f}}$ are the coefficients of variation of \hat{D}_L , n , and \hat{f} , respectively.

Assuming animals are independent with regard to sighting, we note that the probability function of n is Binomial(N_L, P_L). The joint probability density function of the $\{y_i\}$ conditional on n is

$$\prod_{i=1}^n f_c(y_i) = \mu_L^{-n} \prod_{i=1}^n g(y_i),$$

so that the joint distribution of the y_i and n is

$$\binom{N_L}{n} P_L^n (1 - P_L)^{N_L - n} \mu_L^{-n} \prod_{i=1}^n g(y_i).$$

If the animals are randomly distributed and P_L is small so that n is approximately Poisson, and $\text{var}[\hat{f}|n] \approx \sigma^2/n$ for some constant σ^2 , then Burnham and Anderson (1976, 329) showed that $\text{var}[\hat{D}_L]$ is approximately inversely proportional to L . Thus, if we have a sample of r transects of each of length L_i ($i = 1, \dots, r$), then the weighted mean is

$$\bar{D}_L = \frac{\sum_{i=1}^r L_i \hat{D}_i}{\sum_{i=1}^r L_i}$$

with an approximately unbiased variance estimate

$$v(\bar{D}) = \frac{\sum_i L_i (\hat{D}_i - \bar{D})^2}{(r-1) \sum_i L_i}.$$

8.2.2 Encounter Rate

From (8.4), we have

$$\hat{D}_L = \frac{1}{2W} \frac{n}{L} \cdot \frac{1}{\hat{P}_L},$$

where n/L is the encounter rate. Also, using the delta method as referred to above,

$$[CV(\hat{D}_L)]^2 = [CV(n/L)]^2 + [CV(1/\hat{P}_L)]^2,$$

where CV is the usual coefficient of variation. In discussing the above equation, Fewster and Jupp (2009) noted that the encounter rate component of the variance, $\text{var}(n/L)$, typically accounts for 70–80% of the total variance. A variance term can be given using the previous theory, but there are problems if there are population inhomogeneities. To get around this heterogeneity, we use r transects of length L_i ($i = 1, \dots, r$) as above, each with n_i observations. If $L = \sum_{i=1}^r L_i$, then n/L is a ratio of the means, \bar{n}/\bar{L} . In this case, the authors provided several methods of variance estimation and gave two basic approaches, a design-based and a model-based approach, which we now consider.

The first approach considers design-derived estimators for random designs with random line placement. The design consists of r lines with fixed orientation, and, for convenience, the region is oriented so that the lines are vertical. Their horizontal positions are selected uniformly and independently across the full width of the survey region. This ensures that the observations n_i and L_i each form a random sample with respect to the design, thus providing the replication necessary to estimate variance.

If individual line lengths, L_i , are equal for all possible lines, then the total L is fixed so that $\text{var}[n/L] = \text{var}[n]/L^2$ with estimator

$$\widehat{\text{var}}_1 \left[\left(\frac{n}{L} \right) \right] = \frac{1}{L^2} \widehat{\text{var}}[r\bar{n}] = \frac{r}{L^2(r-1)} \sum_{i=1}^r (n_i - \bar{n})^2,$$

where $\bar{n} = n/r$, the mean per line. If the region is rectangular, $\widehat{\text{var}}_1$ is design unbiased. If the line lengths L_i are random, the authors obtained a design-derived variance estimator

$$\widehat{\text{var}}_2 \left[\left(\frac{n}{L} \right) \right] = \frac{r}{L^2(r-1)} \sum_{i=1}^r L_i^2 \left(\frac{n_i}{L_i} - \frac{n}{L} \right)^2,$$

which strictly estimates the mean squared error of n/L . To reflect total effort if line L_i is surveyed t_i times, we replace L_i by $L_i t_i$. The authors did not recommend $r^{-1} \sum_{i=1}^r (n_i/L_i)$ as an encounter rate estimator because of its sensitivity to observations on short lines.

The second approach is model based as a sufficiently robust model might yield an estimator with better precision than a design-based model. It uses two spatial point processes for the objects in the survey area. The first process is a true contagion model with fixed L_i , and it is assumed that

$$E[n_i] = \beta L_i \quad \text{and} \quad \text{var}[n_i] = \sigma^2 L_i, \quad n_1, \dots, n_r \text{ uncorrelated},$$

having the variance of n_i proportional to its mean. This can be viewed as a Poisson-like model with overdispersion and is a homogeneous Poisson process when $\sigma^2 = \beta$. Under this model, we have the model-unbiased variance estimator

$$\widehat{\text{var}}_3 \left[\left(\frac{n}{L} \right) \right] = \frac{r}{L(r-1)} \sum_{i=1}^r L_i \left(\frac{n_i}{L_i} - \frac{n}{L} \right)^2.$$

The second process is an apparent contagion model using a compound Poisson process in which the rate of the Poisson process is itself a random variable. The process is called an apparent contagion because the varying rate of the Poisson process gives the impression of clustering across the region, although the process does not genuinely incorporate a clustering mechanism. A popular apparent contagion model is the Poisson-gamma model, which generates a negative binomial distribution for the n_i . Conditional on the L_i , this leads to

$$E[n_i] = \beta L_i, \quad \text{var}[n_i] = \beta L_i + \gamma L_i^2, \quad \text{the } n_i \text{ uncorrelated},$$

where $\gamma = \beta^2/\tau$ and τ is the shape parameter of a gamma distribution with scale parameter 1. As there are an unlimited number of model-unbiased estimators for $\text{var}[n/L]$, the authors proposed the following estimator:

$$\widehat{\text{var}}_4 \left[\left(\frac{n}{L} \right) \right] = \alpha \sum_{i=1}^r L_i^\phi \left(\frac{n_i}{L_i} - \frac{n}{L} \right)^2.$$

Here,

$$\alpha = \left\{ \sum_{i=1}^r L_i^\phi \left\{ \frac{L}{L_i} - 1 \right\} \right\}^{-1},$$

$$\phi \approx 2 + \left\{ \frac{2(S_2^2 - LS_3)}{LS_2 D_1 - 2S_2 D_2 + 2LD_3 - L^2 D_2} \right\},$$

where $S_j = \sum_{i=1}^r L_i^j$ ($j = 2, 3$) and $D_j = \sum_{i=1}^r L^j \log(L_i)$ ($j = 1, 2, 3$). When all line lengths are the same, we have $\phi = (2 - 2/r)(1 - 2/r)^{-1}$, and $\widehat{\text{var}}_2$ and $\widehat{\text{var}}_3$ reduce to $\widehat{\text{var}}_1$ in this case, which happens only when the region is a perfect rectangle.

When there is a trend in object density through the survey region, systematic designs usually provide better precision for estimating the encounter rate than random designs. However, a single sample from a systematic design strictly gives no replication, and variance estimation is well known to be difficult, as mentioned in the next section. The authors developed extensive theory for a post-stratification scheme with non-overlapping or overlapping strata with differing line lengths. The methods were compared using simulation and applied to a marine mammal survey in British Columbia, Canada.

8.2.3 Systematic Sampling

In sampling plots or line transects to estimate density, systematic designs will generally yield lower variance than random designs. However, although the variance is reduced, estimating it is a well-known problem, and existing methods such as treating samples as independent replicates or using a stratified approximation with overlapping or non-overlapping strata (Fewster et al., 2009) tend to overestimate the variance so that the gain in precision is not available.

Fewster (2011) developed a new “striplet” estimator based on modeling the encounter process over space. This has negligible bias and excellent precision in a wide range of simulation scenarios, including strip sampling, distance sampling, and quadrat sampling surveys, as well as including populations that are highly trended or have a strong aggregation of objects. Its improvement in precision over other methods was demonstrated using survey data for the spotted hyena (*Crocuta crocuta*) in the Serengeti National Park, Tanzania.

In general, for strip (transect) sampling on a rectangular region \mathcal{R} with complete detection, density D is estimated by $\widehat{D} = n/(2WL)$, where n is the number seen, $2W$ is the width of the strip, and L is the length of the transect. To estimate $\text{var}[\widehat{D}]$, we need to specify a conceptual population of replicate surveys that create the variance we want to estimate. We imagine repeating surveys numerous times according to some specified rules and obtaining an estimate \widehat{D} for each replicate

survey. The variance of these replicate estimates is the “true” variance. The set of rules for generating conceptual replicates is called the inferential framework, because it creates the distributions from which one evaluates bias, precision, and confidence interval coverage. This framework defines the distribution of \widehat{D} and its variance, as well as evaluates the bias and precision of the variance estimator.

Fewster (2011) used the inferential framework as follows. For, say, a rectangular area with K transects, we have a baseline with systemic transects perpendicular to the line (y -axis) and the x -coordinates of the baseline extend from 0 to a (a is scaled to 1 by Fewster). Each transect has width W on either side of its centerline, and the spacing between the centerlines is r . Every survey has a total of J search-strips (striplets) parallel to the y -axis with fixed systematic spacing (and fixed orientation), but with a random start point b for this systematic grid chosen from the uniform distribution on $[W, W + r]$. Such a sampling frame ensures that the first (leftmost) possible transect starts at $x = W$ so that it is inside \mathcal{R} , and additionally, J has to be chosen so that the last (rightmost) striplet is at least distance W from the right boundary with x -coordinate at most $a - W$, so it is also inside \mathcal{R} .

Since the first centerline from the left has x -coordinate b , the K centerlines of the transects have x -coordinates, $b, b + r, \dots, b + (K - 1)r$. The continuous sampling frame for b is now approximated by a discrete sampling frame $b \in S_b = \{b_1, b_2, \dots, b_B\}$ for some large B , where the b_i have equally spaced x -coordinates such that $b_1 = W$ and $b_B \leq W + r$. This means that the transect spacing r must satisfy $rK = (a - 2W)$, so that we can choose the start with either r or K . Such a sampling frame undersamples the regions $x \in (0, W)$ and $x \in (a - W, a)$. Consequences of this are likely to be minor, but, if necessary, a wrap-around sampling frame can be used instead (Fewster & Buckland, 2004, 293).

For a single replicate survey arising from a single random b , let $\mathbf{x} = (x_1, \dots, x_J)'$ be the number of objects in the J striplets giving a total of $n = \sum_{j \in S_b} x_j$ counts or n_k for the k th transect. Then, $E[n | b]$ and $\text{var}[n | b]$ can be estimated using the modeled distribution of x . One sums over b_1, \dots, b_B for the discrete uniform distribution for b using

$$\text{var}[n] = \text{var}_b[E(n | b)] + E_b[\text{var}(n | b)].$$

It is assumed that \mathbf{x} is modeled by the multinomial distribution with index N and probabilities $\{p_j\}$. Let striplet j have area α_j , and let the x -coordinate of the midpoint be t_j . Let $\lambda(\cdot)$ be a one-dimensional intensity function of objects per unit area at x -coordinate t ; then

$$p_j = \frac{\alpha_j \lambda(t_j)}{\sum_{m=1}^J \alpha_m \lambda(t_m)}.$$

If the transects have centerlines at t_1^*, \dots, t_K^* and areas a_1, \dots, a_K , then $E[n_i] = a_i \lambda(t_i^*)$, and a generalized additive model can be fitted using

$$\log(E[n_i]) = \log(a_i) + \log[\lambda(t_i^*)]$$

to obtain a smoothed estimate of $\lambda(t_j)$. If $Q(b) = \sum_{j \in S_b} p_j$, then $n \mid b \sim \text{Binomial}(N, Q(b))$ so that $E[n \mid b] = N Q(b)$ and $\text{var}[n \mid b] = N Q(b)[1 - Q(b)]$, which can be substituted into $\text{var}[n]$. This can then be estimated using sample means from the discrete distribution for b with probabilities b_j/B by

$$\widehat{\text{var}}[n] = \frac{1}{B} \sum_{b=b_1}^{b_B} [\widehat{N} \widehat{Q}(b)\{1 - \widehat{Q}(b)\} + \widehat{N}^2 \widehat{Q}(b)^2] - \left\{ \frac{1}{B} \sum_{b=b_1}^{b_B} \widehat{N} \widehat{Q}(b) \right\}^2,$$

where $\widehat{Q}(b) = \sum_{j \in S(b)} \widehat{p}_j$, $\widehat{p}_j = \alpha_j \widehat{\lambda}(t_j / \sum_{m=1}^J \alpha_m \widehat{\lambda}(t_m))$, and $\widehat{N} = A \widehat{D}$.

For details and further extensions, see Fewster (2011). She considered non-rectangular regions with complete detections and estimating the encounter rate n/L , allowing for incomplete detection with distance sampling. Fewster used two-dimensional surveys and alternative distributions for X including overdispersed distance and plot sampling.

8.2.4 Pooling Robustness

Burnham et al. (1980, 45) introduced the criterion of pooling robustness, where we want the estimation of the detection density function $f(y)$ to be robust to variations in detection probability. The probability of detecting an object at any given perpendicular distance will vary according to numerous factors, both external (e.g., weather, habitat, observer, etc.) and internal (e.g., age, sex, size, etc.) of the population being sampled. They considered partitioning line transect data into r strata, chosen so that there is little heterogeneity within each stratum. This also applies to heterodetection probabilities. The pooling robustness criterion is then defined as

$$n \widehat{f}(0) = \sum_{j=1}^r n_j \widehat{f}_j(0),$$

where n_j is the number of detections in stratum j , $\widehat{f}_j(0)$ is the corresponding fitted pdf of distance of detected objects from the line evaluated at zero distance, n is the total number of detections, and $\widehat{f}(0)$ is the pdf fitted to the pooled distance data, evaluated at zero distance.

Burnham et al. (2004) considered this topic further and concluded that “the pooling robustness criterion can be assumed to hold provided we have a sufficiently

flexible model for $g(y)$ to fit the observed distance data.” They noted that this result is surprising and has certain repercussions. For example, the training of shipboard observers often includes a warning that they should maintain a constant search effort. If observers search less efficiently in areas of low density through boredom and more efficiently when they are in areas of high density, then the effective strip width is systematically greater in high-density areas and less elsewhere. This was thought to generate bias in abundance estimates, whereas the above results show that it does not—so long as they maintain $g(0) = 1$ in both high- and low-density areas. It would be best for observers to search with high efficiency throughout, as a larger number of detections give higher precision, but abundance estimation is not compromised if efficiency is variable.

Buckland et al. (2015, 54–55) mentioned that with pooling robustness, “there is no need to record covariates on individuals (such as gender, whether calling, whether in a cluster), environment (such as habitat, thickness of vegetation, visibility) or observer (such as identity, experience, number of observers).” Also, “provided we can reasonably assume that $g(0) = 1$, and provided heterogeneity is not extreme, pooling robustness applies to an overall abundance estimate.” Further comments are made about stratification.

8.2.5 Detection Function

We now turn our attention to modeling the detectability (distance) function $g(y)$, previously introduced in Sect. 8.2. (Buckland et al., 2015, 54) allowed y to be either the perpendicular distance from the line or the distance r from the observation point and specified the following properties that are needed for $g(y)$:

Shoulder. The model should possess a shoulder. This means that the probability of detection should remain at or close to one as distance from the line or point increases away from zero, before falling away at larger distances. Generally, the wider and flatter the shoulder, the better the precision of estimation. With a narrow shoulder, different models can give rise to very different estimates of density, even if they fit the data equally well. The probability of detection should stay close to one for some distance from the line.

Non-increasing. We can expect sightability to go down with the distance.

Model robust. Since the true shape of the detection function is unknown, we need flexible models that can fit a variety of shapes.

Pooling robust. Detectability may be affected by a lot of factors causing heterogeneity and, if necessary, these which could perhaps be included as covariates such as those relating to environment, sex differences of the animals, or observer effects. Pooling robustness, as noted in the previous section, means that $g(y)$ is not affected by such heterogeneity and we do not need to take such factors into account. Sometimes, stratification can be used effectively.

Estimator efficiency. We always like to have an estimate with low bias, but we have to balance this with small standard errors to give us good precision.

Theoretical conditions indicate that a shoulder normally exists so that the slope of the detection function at zero distance, denoted by $g'(y)$, is zero. If the true detection function has a wide shoulder, then different models for the function will tend to give similar estimates of density. However, if the detection function has no, or only a narrow, shoulder, different models can give rise to very different estimates of density, even if they fit the data equally well. Buckland et al. (2015, 63) emphasized that field methods should be used that ensure that probability of detection stays close to one for some distance from the line or point.

For a survey of some of the earlier methods of fitting $g(y)$, see Seber (1982, Section 12.4.3). The first model considered rigorously, using the negative exponential distribution, was considered by Gates et al. (1968). However, its use is not recommended as it does not have a “shoulder.” Buckland et al. (2015) commented that assumption failure leads to spiked data, which results in many detections close to the line or point, with a sharp fall-off with distance. For example, if there is a poor estimation of distance, many detected animals may be recorded as on the line (zero distance). If all animals on the line are certain to be detected, it is unlikely that many animals just off the line will be missed.

Some Detection Functions

Burnham et al. (2004) considered a number of detection functions in detail. We begin with the half-normal distribution, sometimes referred to as a type of key function, $g(y) = g(y; \theta) = \exp[-y^2/(2\sigma^2)]$, where $0 \leq y \leq W$, with $g(0) = 1$, and $\theta = \sigma$ a scale parameter. Varying σ does not change the shape of the detection function, though it does affect how quickly probability of detection falls with distance from the line. Then, the density function of y , conditional on the animal being detected, is

$$f_c(y) = 2 \frac{1}{\sqrt{2\pi\sigma^2}} \exp[-y^2/(2\sigma^2)] dy, \quad 0 \leq y \leq \infty.$$

Given a sample of n distances y_i , the maximum likelihood estimate of σ^2 is $\widehat{\sigma}^2 = \sum_{i=1}^n y_i^2/n$, so that $\widehat{f}_c(0) = 1/\widehat{\mu}$, where $\widehat{\mu} = \int_0^W \exp[-y^2/(2\widehat{\sigma}^2)] dy$. If we can ignore the tail of the distribution, then $\mu_L \approx \mu_\infty$ and

$$\widehat{\mu}_\infty = \sqrt{\frac{\pi\widehat{\sigma}^2}{2}} = \frac{1}{\widehat{f}_c(0)}.$$

In fact, for a number of detectability functions defined on $(0, \infty)$, the tail can usually be ignored for reasonable W . Also, some of the largest distances can be truncated to avoid spurious bumps in the tail of the detection function (e.g., Aars, Marques, et al.

used a 5% trim for polar bears). We can also use covariates to model $g(y, \mathbf{z}; \boldsymbol{\theta})$ (see Marques & Buckland, 2003, 2004). For instance, for the half-normal model and the i th animal, we can use

$$\log(\sigma(\mathbf{z}_i)) = \alpha + \sum_{q=1}^Q \beta_q z_{iq}.$$

We then need to add $f(y_i | \mathbf{z}_i)$ for each i to the overall likelihood. It should be noted that covariates may lead to detectability functions that are not monotone (cf. Miller & Thomas, 2015, for some examples).

Another key function, due to Pollock (1978), is the exponential power series distribution

$$f_c(y) = \frac{\exp[-(y/\lambda)^p]}{[\lambda \Gamma(1 + 1/p)]}, \quad y > 0, \lambda > 0, p > 0,$$

so that maximum likelihood estimates of λ and p are required. The distribution includes the exponential with $p = 1$ and the half-normal with $p = 2$. Maximum likelihood estimators of λ and p are obtained by iteratively solving the maximum likelihood equations. Pollock gave some details and an approximate estimate of the variance of \hat{D} . He also considered dealing with grouped data.

At any one point in time, there is a “hazard” that an object will be detected by the observer; Buckland et al. (2015) used a hazard-rate model

$$g(y) = 1 - \exp[-(y/\sigma)^{-b}], \quad 0 \leq y \leq W,$$

which also has a scale parameter σ , but has shape parameter b as well, providing more flexibility. A similar type of model was considered by Eberhardt (1968), namely,

$$g(y) = 1 - \left(\frac{y}{W}\right)^a, \quad 0 \leq y \leq W.$$

Then $\mu = Wa/(a+1)$, and

$$E[\bar{y}] = E[y] = \frac{W(a+1)}{2(a+2)}.$$

One can either obtain a maximum likelihood estimate of a or use the above equation to obtain a moment estimate.

Nonparametric models have been used, and some details are given by Seber (1982, 463). For example, Burnham and Anderson (1976), who developed much of the above theory, used order statistics for the y_i , while Anderson et al. (1979) used a log-linear model $\log g(y) = \beta_0 + \beta_1 y + \beta_2 y^2$.

Burnham et al. (2004, Chapter 11) provided an excellent and extensive overview of the topic with theoretical examples and details such as some profile likelihood intervals. They also considered a model incorporating a moving observer as follows. Suppose that the perpendicular distance of the object from the line is y (the authors used x) and the observer is at distance x (the authors used z) from the point of closest approach to the object (i.e., the distance on the line of the observer to the foot of the perpendicular from the object to the line), and at radial distance r at angle θ to the line, so that $r^2 = x^2 + y^2$. If the observer approaches from a remote point on the transect so that x may be considered to decrease from ∞ to 0, assume for simplicity that the object cannot be detected once the observer has passed his or her point of closest approach. Let $h(x, y)$ denote the probability of detection function for x and y , given not sighted between ∞ and x . Then changing the variable of integration from x to r gives the detection function

$$g(y) = 1 - \exp\left[-\int_y^\infty \frac{r}{\sqrt{r^2 - y^2}} k(r, y) dr\right],$$

where $k(r, y) \equiv h(\sqrt{r^2 - y^2}, y)$ is referred to as a hazard-rate function. It is necessary to restrict the form of the hazard, and a plausible hazard should satisfy the following conditions:

- (1) $k(0, 0) = \infty$.
- (2) $k(\infty, y) = 0$.
- (3) $k(r, y)$ is non-increasing in r for any fixed y .

For example, suppose that the hazard belongs to the family defined by

$$\int_y^\infty \frac{r}{\sqrt{r^2 - y^2}} k(r, y) dr = (y/\sigma)^{-b},$$

for some σ and b , which was mentioned previously, and leads to the detection function

$$g(y) = 1 - \exp[-(y/\sigma)^{-b}].$$

Hayes and Buckland (1983) gave two hazard functions from this family

$$k(r, y) = cr^{-d}, \quad r \geq y,$$

a function of r only, and

$$k(r, y) = cr^{-d} \cos \theta, \quad \text{where } \sin \theta = y/r.$$

The above hazard-rate development assumes that objects are continuously available for detection, but this is often not the case. For example, whales that travel singly

or in small groups may surface at regular intervals, with periods varying from a few seconds to over an hour, depending on species and behavior, when the animals cannot be detected. Discrete hazard-rate models have been developed for this case (cf. Schweder, 1990; Skaug & Schweder, 1999; and Laake & Borchers, 2004, Sections 6.8.3 and 6.8.4, for further references).

If the fit of the key function is poor, three types of adjustment available from the program **DISTANCE** are available (Buckland et al., 2001, 58–68):

- (1) A cosine series, which, if used in conjunction with the uniform key function, gives the Fourier series model.
- (2) A Hermite polynomial series. These have orthogonal properties with respect to the half-normal key function, and together, these give the Hermite polynomial model.
- (3) A simple polynomial series.

Any adjustment type can be used with any key function (K plus A approach) providing a flexible range of shapes. Maximum likelihood methods are used to fit the models. One recommended strategy for most situations is to choose a small set of key function and adjustment combinations and for each combination to choose the number of adjustment terms using forward selection, i.e., start with no adjustment terms and fit an increasing number of terms, stopping when the Akaike information criterion (AIC) fails to decrease (Thomas et al., 2010).

Goodness-of-fit tests are available including the chi-square test, Kolmogorov-Smirnov test, and both the weighted and the unweighted versions of the Cramér-von Mises test. These are available in the **DISTANCE** software. With a goodness-of-fit test, counts are recorded in distance intervals with endpoints $c_0, c_1, \dots, c_r = W$ (usually $c_0 = 0$), with interval counts n_j . Then the test statistic is

$$\chi^2 = \sum_{j=1}^r \frac{(n_j - n\hat{\pi}_j)^2}{n\hat{\pi}_j},$$

where

$$\hat{\pi}_j = \int_{c_{j-1}}^{c_j} \hat{g}(y) dy$$

and $\hat{g}(y) = g(y; \hat{\theta})$. Here, χ^2 is approximately distributed as the χ^2 distribution with $r - q - 1$ degrees of freedom under the null hypothesis that our model is the true detection function, where q is the number of parameters in the model that we have estimated.

Miller and Thomas (2015) introduced detection functions that are based on mixtures of simple parametric key functions, such as

$$g(y, \mathbf{z}; \boldsymbol{\theta}, \boldsymbol{\phi}) = \sum_{j=1}^J \phi_j g_j(y, \mathbf{z}; \boldsymbol{\theta}_j).$$

For example, $g_j(\cdot, \cdot)$ could be the half-normal function $\exp[-y^2/(\sigma_j(\mathbf{z})^2)]$, where \mathbf{z} are covariates. This leads to expressions for the likelihood, the effective area of detection, and a Horvitz-Thompson-like estimator of N , the population size.

8.2.6 Group (Cluster) Sampling

Animals are frequently seen in groups so that the group is considered as a “point” and the number s in the group is also recorded. Chen (1999) considered estimation for this situation. The cluster density is $D = N/A$, where N is the number of clusters and A the area. The density of individual animals is $D_s = N\mu_s/A$, where μ_s is the mean cluster size. Let y_i denote the (signed) perpendicular distance of the i th cluster from the transect line with \mathbf{z}_i covariates, where z_1 the first element of \mathbf{z} is the cluster size. Assuming the transect lines with a total length of L are allocated randomly in the survey region, we have that y and \mathbf{z} are independent, y is uniformly distributed in $(-W, W)$, and their joint distribution is

$$f_{12}(y, \mathbf{z}) = (2W)^{-1} f_2(\mathbf{z}) I(-W \leq y \leq W),$$

where $I(\cdot)$ is the usual indicator function. For observers $i = 1, 2$, let $g_i(y, \mathbf{z})$ be the probability of detecting a cluster at a perpendicular distance y , and let

$$P_i = \int \int_{-W}^W g_i(y, \mathbf{z}) f_{12}(y, \mathbf{z}) dy d\mathbf{z}.$$

Then, from the theory of weighted distributions,

$$f_i(y, \mathbf{z}) = \frac{g_i(y, \mathbf{z}) f_{12}(y, \mathbf{z})}{P_i} = \frac{g_i(y, \mathbf{z}) f_2(\mathbf{z})}{2P_i W}.$$

If $N(\mathbf{z})$ is the total number of clusters having covariates \mathbf{z} , then $D(\mathbf{z}) = N(\mathbf{z})/(2LW)$ is the local abundance density at \mathbf{z} . Let $f_i(y | \mathbf{z})$ be the conditional density of the sighting distance given \mathbf{z} , and let $\mu_i(\mathbf{z}) = \int_{-W}^W g_i(y, \mathbf{z}) dy$ be the effective search width of the i th observer at \mathbf{z} . Then, from the previous theory,

$$f_i(y | \mathbf{z}) = \frac{g_i(y, \mathbf{z})}{\mu_i(\mathbf{z})}.$$

Recalling that z_1 is the cluster size, if \mathbf{z} is discrete (or continuous and binned) and takes values within a discrete set \mathcal{S} , the abundance densities are

$$D = \sum_{\mathbf{z} \in \mathcal{S}} D(\mathbf{z}) \text{ and } D_s = \sum_{\mathbf{z} \in \mathcal{S}} z_1 D(\mathbf{z}).$$

The probability that at least one observer detects a cluster of (y, \mathbf{z}) is

$$g(y, \mathbf{z}) = g_1(y, \mathbf{z}) + g_2(x, \mathbf{z}) - g_1(x, \mathbf{z})g_2(x, \mathbf{z}),$$

assuming conditional independence of detection between the two observers. The joint effective search width at \mathbf{z} is

$$\mu(\mathbf{z}) = \int_{-W}^W g(y, \mathbf{z}) dy = \mu_1(\mathbf{z}) + \mu_2(\mathbf{z}) - \alpha(\mathbf{z})\mu_1(\mathbf{z})\mu_2(\mathbf{z}),$$

where

$$\alpha(\mathbf{z}) = \int_{-W}^W f_1(y \mid \mathbf{z})f_2(y \mid \mathbf{z}) dy.$$

Let $n(\mathbf{z})$ be the number of distinct sightings made by the two observers. If the transect lines are allocated randomly such that $E[n(\mathbf{z})] = N(\mathbf{z})P_0(\mathbf{z})$, where $P_0(\mathbf{z}) = \mu(\mathbf{z})/(2W)$ is the probability of detecting a cluster of \mathbf{z} , then from Burnham and Anderson (1976), an estimator for $D(\mathbf{z})$ that does not require $g(0, \mathbf{z}) = 1$ is

$$\widehat{D}_0(\mathbf{z}) = \frac{n(\mathbf{z})}{L\widehat{\mu}(\mathbf{z})},$$

where $\widehat{\mu}(\mathbf{z})$ is an estimator of $\mu(\mathbf{z})$. We now have estimators for D and D_s , namely,

$$\widehat{D} = L^{-1} \sum_{\mathbf{z} \in S} \frac{n(\mathbf{z})}{\widehat{\mu}(\mathbf{z})}, \text{ and } \widehat{D}_s = L^{-1} \sum_{\mathbf{z} \in S} \frac{z_1 n(\mathbf{z})}{\widehat{\mu}(\mathbf{z})}.$$

The authors gave methods for estimating $\mu(\mathbf{z})$ (using capture-recapture theory) and $\alpha(\mathbf{z})$ (using kernel estimators of the $f_i(y \mid \mathbf{z})$, as well as asymptotic variance estimators).

8.2.7 Transects in Plots

Buckland et al. (2009) considered the situation where there are plots and the transects are laid out in each plot. The density estimate for plot k is then, from an adaption of (8.4),

$$\widehat{D}_k = \frac{1}{2L_k} \sum_{i=1}^{n_k} s_i \widehat{f}_c(0 \mid \mathbf{z}_i) = n_k \frac{H_k}{2L_k},$$

where L_k is the total transect length, n_k is the number of animal clusters detected on plot k , $\hat{f}_c(y \mid \mathbf{z}_i)$ is the estimated probability density function for the i th detected cluster on plot k being detected at distance y given present, \mathbf{z}_i are the covariates for that cluster, and the sample mean

$$H_k = \frac{1}{n_k} \sum_{i=1}^{n_k} s_i \hat{f}_c(0 \mid \mathbf{z}_i).$$

The authors discussed bootstrapping to obtain variance estimates.

8.3 Design Considerations

The first question to consider is how to choose a line transect, and Strindberg and Buckland (2004) mentioned several methods. Any transect can be chosen at random by having a baseline, choosing a point at random on it, and having the line transect perpendicular in direction to the baseline. If there is more than one transect, they can be parallel but arising from random points or are systematically spaced. At the end of each transect line, it is important to include only animals in the covered area, especially if the animals are sparsely distributed, as too many animals might be recorded—a kind of “end effect.” The total length of the transect lines needs to be determined, and this point is discussed later.

Sometimes, a transect line is not straight because of the terrain, so its length needs to be carefully determined. Roadside counts are sometimes used, though roadside avoidance may occur with some species (e.g., Ward et al., 2004, roe deer). Often, transect lines consist of a series of parallel equidistant lines. One transect method used particularly for a marine environment using an unmanned surface vehicle, called the coarse-fine-stitched method, was given by Sun and Fu (2018). Another question is how to accommodate measurement errors. This is discussed in Sect. 7.2.3, where the theory there for point transects also applies to line transects.

Strindberg and Buckland (2004) noted that for a design that has randomly spaced or systematically spaced lines with a random start, there is a major disadvantage if a ship or aircraft is used as it must travel from one line to the next without searching for the animals of interest. Also, if searching is carried out on these lines, the random design can be compromised as there can be greater effort along the boundaries of the survey region where animal density may be atypical. However, if searching is not carried out on these lines, resources are not used efficiently, especially in the case of shipboard surveys, for which ship time is expensive. Successive survey lines may be tens or even hundreds of kilometers apart and be very long, so that the loss of search effort is unacceptable on cost grounds. As a consequence, continuous zigzag designs with continuous effort are often preferred. Although variance estimation may be biased, systematic designs are more efficient because systematic zigzags and systematic parallel lines are more evenly distributed throughout the survey region

than are randomly spaced parallel lines. Zigzags are discussed in more detail in the next section.

We note that in earlier work, W was usually set to ∞ . Very often this has little effect on the analysis, particularly if the detectability falls off quickly after a distance W . It has the advantage of avoiding truncating the density function $f_c(y)$. For this situation, we have (e.g., Seber, 1982, 29–30, 462) that μ_L is replaced by $\mu_\infty = \int_0^\infty g(y)dy$ and $f_c(0) = 1/\mu_\infty$ that is estimated by \hat{f} . Also, the probability P of sighting an animal from the transect is $2L\mu_\infty/A$ and, given $g(0) = 1$,

$$\hat{N} = n/\hat{P} = nA\hat{f}/2L \quad \text{and} \quad \hat{D} = n\hat{f}/(2L),$$

as in (8.4). Technically, to use this model, we would need to have a single transect running the whole length of the population area and being chosen at random. This ensures that the covered area is the whole population area. However, as we have seen above, we can get around this problem.

8.3.1 Some Designs

Strindberg and Buckland (2004) considered some designs including the zigzag design. Consider a survey strip centered on the transect line of width $2W$, and define the coverage probability $\pi(\mathbf{u})$ as the probability that the sampler covers the point $\mathbf{u} = \mathbf{u}(x, y)$, given a rule for randomizing the location of the sampler. As standard zigzag designs assume uniform coverage, this can lead to substantial bias in abundance estimates if animal density varies appreciably through the study area. It is now shown how to modify the design to take account of uneven coverage probability using a Horvitz-Thompson type of estimator.

Let N be the population size of the surveyed area with surface area A , let L be the total transect length, let n be the number of animals detected, and let p_i be the probability that animal i is observed. Then, from (8.4), N can be estimated by

$$\hat{N} = \frac{An}{2\hat{\mu}_L},$$

where $\hat{\mu}_L$ is the estimated “effective strip half-width.” This estimate can also be obtained as a Horvitz-Thompson-like estimator as follows. If the inclusion probability of animal i is $\pi_c \approx WL/A$, then a Horvitz-Thompson-type estimator is

$$\hat{N} = \sum_{i=1}^n \frac{1}{\hat{p}_i} = \sum_{i=1}^n \frac{1}{(\hat{\mu}_L/W)\pi_c} = \frac{An}{2\hat{\mu}_L L}.$$

If we now allow the probability that the sample strip includes a given animal i to be π_i , and each animal has an individual effective strip half-width μ_i , then N can be estimated by

$$\widehat{N} = \sum_{i=1}^n \frac{1}{\widehat{p}_i} = \sum_{i=1}^n \frac{1}{(\widehat{\mu}_i/W)\pi_i} = \frac{W}{\widehat{\mu}_i} \sum_{i=1}^n \frac{1}{\pi_i},$$

where μ_i may be modeled as a function of the animal's distance from the line and of other covariates. If animals occur in clusters, and the i th detected cluster is of size s_i , then we replace $\sum_{i=1}^n 1/\widehat{p}_i$ by $\sum_{i=1}^n s_i/\widehat{p}_i$ in each of the above expressions. Variances for the various estimators may be estimated by bootstrapping transects.

Strindberg and Buckland (2004) showed how to modify π to allow for a zigzag sampler along with algorithms for three types of zigzag: the equal-angle zigzag sampler, the equal-spaced zigzag sampler, and the adjusted-angle zigzag sampler. The methods were applied to minke whales in the Antarctic. In conclusion, they recommended that equal-angle zigzag samplers not be used, but equal-spaced zigzag samplers are likely to prove satisfactory in most surveys, typically with low bias, even if constant coverage probability is assumed. With sophisticated navigation systems, the adjusted-angle zigzag sampler may be practical and offers a worthwhile improvement over equal-spaced zigzag samplers when the shape of the survey region is irregular.

Buckland et al. (2015, 26) compared the efficiencies of three designs for surveying a simulated population, namely, a systematic sample of parallel lines with a random start, a simple random sample of parallel lines, and a zigzag sample, based on equally spaced waypoints. Using simulation, they found that there was no evidence of bias for the three methods. Although the total distance covered by each method is approximately the same, they see that the distance covered while surveying is much greater for the zigzag design, because there is no off-effort travel from the end of one line to the start of the next. Consequently, the zigzag design gave higher precision for their example. In cases where on-effort surveying is more costly or slower per unit distance than off-effort travel, the gain, if any, of zigzag sampling will be less. They noted that zigzag surveys tend to be especially useful for shipboard surveys conducted in large survey regions and also for aerial surveys when separation between successive lines in a parallel line design would be large.

In addition to zigzag lines, sometimes, we end up with a curved line such as when following a particular path, trail, or roadside. For example, it can be difficult to follow a straight line through dense bush. Hiby and Krishna (2001) considered this problem, assuming either animals are randomly distributed with constant density through the study area or transects are located independent of animal density. If the shortest distance of an animal detected from the site to the transect is measured (which may be hard to do in some circumstances), then under certain conditions, the usual formula for density estimation still holds, namely,

$$D = E[n]f_c(0)/(2L).$$

8.3.2 Choice of Transects

There are two components to sample size: the number of line transects in the design and the number of animals (or animal clusters for clustered populations) detected. Buckland et al. (2001, 232) recommended a minimum of 10–20 replicate lines to allow the reliable estimation of variance of encounter rate. If the population is very patchily distributed, then more lines are needed to ensure that the variability through the study area is adequately represented and estimated. Buckland et al. (2001, 240–241) also suggested that at least 60–80 animals (or animal clusters) should be detected for reliable estimation of the detection function for line transect sampling.

To determine the total length of line to use, one can carry out a pilot survey using lines of total length L_0 (Buckland et al., 2015, 24) that provides sufficient data to estimate density D and its coefficient of variation by \widehat{D} and $cv(\widehat{D})$, respectively, with the target coefficient of variation $cv_t(\widehat{D})$. An estimate of the total length of line needed to achieve cv_t is then

$$L = \frac{L_0[cv(\widehat{D})]^2}{[cv_t(\widehat{D})]^2}.$$

If the pilot data are insufficient for this method, and n_0 animals are detected, we can use

$$L = \frac{L_0}{n_0} \frac{b}{[cv_t(\widehat{D})]^2},$$

with b set equal to 3 Buckland et al. (2001, 242).

8.4 Transects of Random Length

In Sect. 9.2, we consider transects of random length for the line intercept method, following Seber (1979). The extension to distance sampling was discussed by Burnham et al. (2004) who noted that we can either have line transects of fixed length followed by random sample sizes (overall and per line) or else fix the total sample size to be achieved and then traverse the line(s) until that predetermined sample size is obtained. If L is the total transect length and n is the total sample size, we thus have either fixed L and n random or fixed n and L random. They concluded, as shown below, that these two approaches “under some idealized conditions, are not importantly different in their statistical properties. Primarily, field (i.e. applied) considerations dictate the choice between sampling schemes.”

To distinguish between the two cases, the authors used \tilde{L} and \tilde{n} for the random cases, so that $E[\tilde{L} | L] = L$ and $E[\tilde{n} | n] = n$. For the first case with L fixed, we

have from Seber (1982, 462) with μ_L replaced by $\mu_\infty = \int_0^\infty g(y)dy$,

$$\mathbb{E}[\tilde{n}] = 2DL\mu_\infty = 2DL/f_c(0),$$

which leads to the estimator

$$\hat{D} = \frac{\tilde{n}\hat{f}_c(0)}{2L}.$$

Assuming \tilde{n} has a Poisson distribution, then $n \sim \text{Poisson}(2LD/f_c(0))$.

For the second case with random \tilde{L} , it is assumed there is a random starting point for the line and one moves along it until n detections are made. Thus, the first segment is from the starting point on the line to the point perpendicular to the first detection, and there are $n - 1$ random inter-observational segments, each of random length $\tilde{\ell}_i$, which add to \tilde{L} . Here, $\tilde{\ell}_i$ is the distance traveled between points perpendicular to detections $i - 1$ and i , where $i = 0$ is defined to be the starting point. Assuming that the number of objects in any area of size is Poisson with mean aD , then it can be shown that $\tilde{\ell}$ is an exponentially distributed random variable with mean $\mathbb{E}[\tilde{\ell}] = f_c(0)/(2D)$. The probability density function (pdf) of $\tilde{\ell}$ is

$$f_1(\tilde{\ell}) = \frac{2D}{f_c(0)} \exp\left[-\frac{2\tilde{\ell}D}{f_c(0)}\right].$$

Because of the above assumptions, the $\tilde{\ell}_i$ are mutually independent exponential random variables, so that their sum \tilde{L} has a gamma distribution

$$f_2(\tilde{L}) = \frac{[(2D/f_c(0))]^n \tilde{L}^{n-1} \exp[-(2\tilde{L}D/f_c(0))]}{(n-1)!}.$$

It then follows that $\mathbb{E}[\tilde{L} | n] = nf_c(0)/(2D)$, which leads to the estimator

$$\tilde{D} = \frac{n\hat{f}_c(0)}{2\tilde{L}},$$

where $\hat{f}_c(0)$ is computed conditional on n exactly as in the case of fixed L and random \tilde{n} , so that $\hat{f}_c(0)$ is the same estimator in either sampling scheme. Also, for large n , the two estimators have almost identical biases, while the bias for \tilde{D} is reduced when n is replaced by $n - 1$.

Using the delta method, the asymptotic variance of \tilde{D} is

$$\text{var}[\tilde{D}] = D^2 \left[[CV(\tilde{L})]^2 + [CV\{\hat{f}_c(0) | n\}]^2 \right],$$

where CV is the coefficient of variation. For comparable effort and large samples, it transpires that

$$\text{var}[\tilde{D}] \approx \text{var}[\hat{D}].$$

The authors noted that having fixed n and random \tilde{L} is often not a practical design in line transect and discussed this question in detail. For example, when planes or helicopters are used, you cannot set out to fly a random distance; \tilde{L} cannot exceed the fuel capacity of the plane. For most methods of traversing the line(s), the distance to travel must be specified in advance for a number of reasons including allowing an accurate cost estimate for a study and not finishing the study in an inconvenient place or without adequate line coverage.

8.5 Animal Signs

Transect methods can be used to determine population sizes and densities of animal signs or of even dead animals. They provide estimates of relative density for the animals as discussed in Sects. 2.6 and 4.1.3 It was mentioned there that to convert to population density, “correction factors” are needed of the average number of signs per animal. For example, pellet groups (along with information about defecation rates and pellet persistence) can be used for deer.

Marques et al. (2001) using line transect distance methods obtained estimates of deer defecation and length of time to dung decay to convert pellet group density to deer density by geographical block and habitat type. They noted that one difference between direct surveys of animals and indirect surveys of dung is that the latter provide estimates of average abundance over several months, whereas the former usually yield estimates of abundance for the day of the survey. This may provide misleading information on habitat use. Keeping, Burgar et al. (2018) found that using track counts carried out by seasoned local trackers provided better information than transect methods because encounter rates by direct sightings were relatively low.

Urbanek et al. (2012) compared the distance method using pellet groups with aerial censusing with regard to costs and other factors. Anderson et al. (2013), who referred to a number of studies using signs (e.g., of African elephants, deer, badgers, mountain hares), compared the method with spotlighting from road transects and concluded that the latter method may be better in non-forested landscapes because of roadside avoidance. The use of road transects should be discouraged, though in some cases it can be helpful when access is a problem. A variety of signs can be used, for example, waterfowl nests were considered by Anderson and Pospahala (1970). As noted above, transects have also been used for calling surveys such as with birds and frogs (anurans). Castellón et al. (2015) used tortoise burrows and determined occupancy.

8.6 Line Transect Radial Distances

We note that $y_i = r_i \sin \theta_i$, where r_i is the radial distance to the i th observed animal and θ_i is the angle that the radial line makes with the line transect. Unless y_i is measured directly, the obvious question to ask is why not use the r_i instead of the y_i , especially when animals have to be flushed out into the open and θ is not easy to measure. Unfortunately, not all animals flushed are actually seen, and flushed animals will vary with regard to the distance from the observer when they will be flushed. They may also “drive” ahead and move down the line of travel ahead of the observer. Of course, these problems occur irrespective of what distances are measured. We now briefly develop some theory based on Burnham and Anderson (1976, 331). Let y be the right-angle distance and r the radial distance of an animal. We now generalize the detection probability $g(y)$ (with $g(0) = 1$) to

$$g(y, r) = g(y)h(r | y), \quad 0 \leq y \leq W, y \leq r \leq \infty,$$

where $g(y) = \int_y^\infty g(y, r)dr$. One can make a notational change and let $z = y$ when the measurement y is made conditional on the animal being observed. Then using (8.2) with $f_c(y) = f(z)$, the joint density function of z and r (conditional on the animal being detected) is

$$f(z, r) = f(z)h(r | z) = \frac{g(z)h(r | z)}{\mu_L} = \frac{g(z, r)}{\mu_L}, \quad 0 \leq y \leq W, y \leq r \leq \infty,$$

where $\mu_L = \int_0^W g(z)dz$. We also note that $f(z, r) = h(r)f(z | r)$, where $h(r)$ is the unconditional probability density function of the sighting distance and $f(z | r)$ is the conditional density of z given r , for $0 < z < \min\{W, r\}$. Now $f(z)$ can be expressed as

$$f(z) = \frac{g(z)}{\mu_L} = \int_z^\infty h(r)f(z | r)dr.$$

As $g(0) = 1$,

$$f_c(0) = 1/\mu_L = \int_0^\infty h(r)f(0 | r)dr = E_r[f(0 | r)].$$

If $f(0 | r)$ was known, an estimator of $f_c(0)$ would be

$$\hat{f}_c(0) = \frac{1}{n} \sum_{i=1}^n f(0 | r_i).$$

However, $f(0 | r)$ is not specifically known so the focus is on assuming a theoretical or empirical form for $f(z | r)$ to obtain an estimate $\hat{f}(0 | r)$. Since it is difficult to specify W when only radial distances are measured, it is usual to assume $W = \infty$.

The above theory overlaps with point transects, as considered in Sect. 7.2.1.

8.6.1 Circular Flushing Region

If we assume that an animal is flushed into the open as soon as the observer crosses the boundary of a circle of radius r centered on the animal, we find from the line intercept theory of Chap. 9 with $w_i = 2r_i$ that an estimator due to Hayne (1949a) which is

$$\begin{aligned}\hat{D}_H &= \frac{n}{2L} \left(\frac{1}{n} \sum_{i=1}^n \frac{1}{r_i} \right) \\ &= \frac{1}{2L} \sum_{i=1}^n \frac{1}{r_i}\end{aligned}$$

is unbiased, that is, $E[\hat{D}_H | \{r_i\}] = D$, so that $E[\hat{D}_H] = D$, provided the transect is randomly placed. Also,

$$f(z | r) = \frac{1}{r}, \quad 0 \leq z \leq r,$$

so that given r , $\sin \theta = z/r$ has a uniform distribution on $[0, 1]$ and θ has density function

$$f_2(\theta) = \cos \theta, \quad 0 \leq \theta \leq \pi/2,$$

with

$$E[\theta] = \frac{\pi}{2} \text{ radians} = 32.704^\circ.$$

Any substantial departure of the sample mean angle from the above would indicate a model failure. We can also test if the $\sin \theta_i$ come from a uniform distribution. For further discussion, see Seber (1982, 468–470).

The question is now which to use, right-angle or radial distances. There is the question of how to treat an animal at distance d along the transect. For the radial model, we simply use $g(d)$, but with the distance model, we use $g(0) = 1$. The main disadvantage of using r is the need to know $f(0 | r)$, which is not needed to obtain an estimator based on right-angle distances. For the latter, we need to know $f_c(0)$. If y cannot be measured directly, then using $y = r \sin \theta$ may not be accurate,

as θ may not be easy to measure, particularly if the transect is not straight. Clearly, the choice will depend on both the terrain and the animal under observation. Hayes and Buckland (1983) concluded that a nonparametric right-angle distance model will generally result in more reliable estimation than any radial distance model.

8.7 Plant Sampling

We consider plant sampling in Chap. 9 where coverage methods are used for estimating plant cover when plants are not readily distinguishable. If they can be uniquely identified, Buckland et al. (2007) considered using a line transect method. They note that in line transect sampling, substantially more ground can be covered in a given time than for quadrat sampling. The ability to cover a large area of ground with modest resources is an important advantage as plants often have a very patchy distribution.

There are special features involved with plants. For instance, the survey site might be very small so that a large number of sites will be needed. If site-specific abundance or density estimates are required, standard line transect sampling requires that there are of the order of 20 transect lines or more at each site, which can be a problem with a very patchy plant distribution. Also, for small or cryptic species, plants on the line may not be detected. Although in many circumstances quadrat (or plot) sampling as described previously is likely to prove adequate for plant surveys, the more complex methods of this chapter are potentially useful in the following circumstances.

First, if individual plants are easily overlooked, quadrat sampling may require painstaking searching on hands and knees, severely limiting the amount of ground that can be covered. The methods given below in which detection on the transect line can be allowed to be uncertain may prove a more cost-effective option. For example, the authors referred to cowslips, where they estimated that approximately 30% of the cowslips were missed in thick vegetation at distance zero despite very careful searching along the transects by the observers. Second, if the plant distribution is very clumped, or if plants are sparsely distributed, quadrat sampling may give poor precision. We now consider a particular line transect model.

8.7.1 Crossed Design

Buckland et al. (2007) proposed a “crossed design” consisting of two systematic grids of parallel lines perpendicular to each other, north to south (N/S) and east to west (E/W). As usual, detections within a distance W of a line are recorded. Wherever lines from the two grids intersect, the square of side $2W$ that is centered on the point of intersection is surveyed twice, providing “double-observer” mark-recapture data, from which the probability of detection can be estimated without

assuming detection is certain on the line. Because the square is surveyed from each of its perpendicular centerlines, the dependence between detections arising from heterogeneity in the detection probabilities is reduced, making the task of modeling the heterogeneity easier. The advantage of the crossed design over a parallel strip design is that the assumption that all plants within the strip are detected can be tested. This can be done by determining whether there are plants in the intersection squares that are detected in one of the surveys, but not both.

Assuming all plants on the centerline are detected, a detection function is fitted using the perpendicular distances. For the E/W data only, we have the likelihood function (cf. (8.2)), conditional on the sample size n_E , of

$$L_E(\boldsymbol{\theta}_E; \{y_i\}) = \prod_{i=1}^{n_E} f_E(y_i) = \frac{\prod_{i=1}^{n_E} g_E(y_i)}{(WP_E)^{n_E}}, \quad (8.5)$$

where y_i is the distance from the line to the i th detected plant, ($i = 1, \dots, n_E$), and $f_E(y)$ is the probability density function of distance y given the plant is detected (equal to $f_c(y)$ above). Here, $\boldsymbol{\theta}_E$ is the vector of parameters used to specify $f_E(y)$, $g_E(y_i)$ is the probability that a plant at distance y from the line is detected, and

$$P_E = \frac{1}{W} \int_0^W g_E(y) dy$$

is the probability that a plant is detected, given that it is somewhere within the surveyed strip of half-width W . This leads to an estimate of population size $\hat{N}_E = An_E/(a_E \hat{P}_E)$, where A is the population area, $a_E = 2WL_E$ is the covered area corresponding to the E/W lines, and L_E is the total length of E/W lines. It is assumed that plants available for detection are uniformly distributed with respect to distance from the line, a requirement that is achieved on average by appropriate randomization of the survey design. We can obtain a second estimate \hat{N}_N from the data from the N/S lines by simply replacing the subscript E by the subscript N so that the likelihood conditional on the sample size n_N is

$$L_N(\boldsymbol{\theta}_N; \{x_i\}) = \prod_{i=1}^{n_N} f_N(x_i) = \frac{\prod_{i=1}^{n_N} g_N(x_i)}{(WP_N)^{n_N}},$$

where x_i is the distance of the i th detected plant from the N/S lines ($i = 1, \dots, n_N$). Assuming that probability of detection of a plant from one line is independent of its distance from the other line, we can combine the two likelihoods to give

$$L(\boldsymbol{\theta}_E, \boldsymbol{\theta}_N; \{x_i\}, \{y_j\}) = L_N(\boldsymbol{\theta}_N; \{x_i\})L_E(\boldsymbol{\theta}_E; \{y_j\})$$

The above likelihood can be maximized by conducting independent analyses of the E/W data and the N/S data using the software **DISTANCE** to get estimates $\hat{\boldsymbol{\theta}}^E$ and

$\widehat{\theta}^N$ and hence obtain estimates \widehat{P}_E and \widehat{P}_N . If field methods are identical for the N/S and the E/W lines, it is reasonable to assume that $g_N(y) = g_E(y) = g(y)$. This analysis may be achieved in distance by entering the E/W data as one stratum and the N/S data as a second stratum and fitting a pooled detection function across the two strata. Stratum-specific abundance estimates should then be averaged, not summed. A Horvitz-Thompson-like estimator now provides our estimate of abundance:

$$\widehat{N}_1 = \frac{A}{a} \left(\frac{n_E}{\widehat{P}_E} + \frac{n_N}{\widehat{P}_N} \right),$$

with $a = a_E + a_N = 2W(L_e + L_N)$. Here, each intersection square has been surveyed twice, and its size is included in a twice. If plants are markedly non-uniform through the survey region and the number of lines in the design is small, the authors suggest that the following estimator might be preferred:

$$\widehat{N}_2 = 0.5 \left(\frac{n_E}{q_E \widehat{P}_E} + \frac{n_N}{q_N \widehat{P}_N} \right),$$

where q_E is the proportion of plants detected from the N/S lines that fall within the intersection squares (and hence within the E/W strips) and conversely for q_N . Thus, q_E estimates the proportion of objects in the survey region that fall within the E/W strips, and q_N the proportion that fall within the N/S strips. If \widehat{P}_E , \widehat{P}_N , n_E , and n_N are assumed to be independent, the delta method may be used to estimate an approximate variance of \widehat{N}_1 :

$$\widehat{\text{var}}[\widehat{N}_1] = \left(\frac{A}{a} \right)^2 \left[\widehat{\text{var}} \left(\frac{n_E}{\widehat{P}_E} \right) + \widehat{\text{var}} \left(\frac{n_N}{\widehat{P}_N} \right) \right],$$

where

$$\widehat{\text{var}} \left[\left(\frac{n_E}{\widehat{P}_E} \right) \right] \approx \left(\frac{n_E}{\widehat{P}_E} \right)^2 \left[\{\widehat{CV}(n_E)\}^2 \left(1 - \frac{a_E \widehat{P}_E}{A} \right) + \{\widehat{CV}(\widehat{P}_E)\}^2 \right],$$

and similarly for $\widehat{\text{var}} \left(\frac{n_N}{\widehat{P}_N} \right)$. The coefficients of variation $\widehat{CV}(n_E)$ and $\widehat{CV}(\widehat{P}_E)$ may be estimated, for example, using the software **DISTANCE**. A similar expression is available for \widehat{N}_2 , namely,

$$\widehat{\text{var}}[\widehat{N}_2] = 0.25 \left[\widehat{\text{var}} \left(\frac{n_E}{q_E \widehat{P}_E} \right) + \widehat{\text{var}} \left(\frac{n_N}{q_N \widehat{P}_N} \right) \right],$$

where

$$\begin{aligned} \widehat{\text{var}} \left[\left(\frac{n_E}{q_E \widehat{P}_E} \right) \right] \\ \approx \left(\frac{n_E}{q_E \widehat{P}_E} \right)^2 \left[\{\widehat{CV}(n_E)\}^2 (1 - \frac{a_E \widehat{P}_E}{A}) + \{\widehat{CV}(\widehat{P}_E)\}^2 + \{\widehat{CV}(q_E)\}^2 \right], \end{aligned}$$

and similarly for $\widehat{\text{var}} \left[\left(\frac{n_N}{q_N \widehat{P}_N} \right) \right]$. Here, $\widehat{CV}(q_E)$ may be estimated from the between-line sample variance in q_E , weighted by the number of observations contributing to the estimate of q_E from each line. From simulation results, the authors preferred \widehat{N}_1 because of lower variance.

Buckland et al. (2007) discussed the use of bootstrap methods and introduced covariance terms to allow for the case when estimates are not independent. They then turned their attention to the case when there is a non-uniform distribution of objects. Let $\pi_E(y)$, with parameters ϕ_E , be the pdf of distances from the line of all objects within the covered strips, whether or not they are detected, ($0 \leq y \leq W$). Then (8.5) becomes

$$L_E(\theta_E, \phi_E; \{y_i\}) = \prod_{i=1}^{n_E} f_E(y_i) = \frac{\prod_{i=1}^{n_E} g_E(y_i) \pi_E(y_i)}{(W P_E)^{n_E}},$$

where

$$P_E = \int_0^W g_E(y) \pi_E(y) dy.$$

A similar expression was given for $L_N(\theta_N, \phi_N; \{y_i\})$. If objects are uniformly distributed with respect to distance from the line, then $\pi_E = 1/W$, and we get (8.5) again. However, in general, when $\pi_E(y)$ and $g_E(y)$ must both be estimated, the above likelihood is not sufficient, as the two functions always appear as a product. The authors proposed a method to estimate $\pi_E(y)$ using the data from the intersection squares. The same applies to $\pi_N(y)$. They then considered a method to allow for uncertain detection on the trackline using the fact that data from the intersection squares can be regarded as two-sample capture-recapture data, which can help to estimate the detection function. Also, covariates can be used to model heterogeneity.

8.7.2 Time Spent

Garrard et al. (2008) considered the question of what is the appropriate amount of time to spend at survey locations, as distinct from the appropriate number of

repeat visits. They assumed an exponential detection time model $f(t) = \lambda \exp[-\lambda t]$ with mean $\mu_t = 1/\lambda$ for the time that the first individual of a species at a given location was first observed, referred to as the detection time of that species. If T is the duration of the survey and ψ is the probability of occupancy at a site, then

$$f(t | \psi, \lambda) = \begin{cases} \psi \lambda \exp[-\lambda t], & y \leq T \\ \psi \exp[-\lambda t] + (1 - \psi), & y > T. \end{cases}$$

If ψ_i is the occupancy probability for site i , $\text{logit } (\psi_i)$ was expressed as a linear regression. The model was found to perform reasonably well with real data. The probability p of detection is estimated by $\hat{p} = 1 - \exp[-\hat{\lambda}t]$ and can be used to plot detectability curves. This model was then simulated to find the best combinations of parameter values for planning survey effort.

8.8 Acoustic Methods

Some species are difficult to sight and are not easily marked or recaptured but produce readily identifiable sounds, as these include mammals, birds, fishes, amphibians, and insects, and sounds of species yet to be discovered. It is also important to be able to link sounds to species' behavior and the environment and to cue rates. The method is particularly useful in marine surveys discussed in Sect. 8.13 providing an opportunity to use passive acoustic data to estimate animal density. The same is true for other species in difficult environments such as undersea just mentioned, where sound travels better than air, or in dense bush and where animals are difficult to see visually. Marques et al. (2013) gave an extensive review of estimating animal population density using passive acoustics. They proposed four steps in the estimation process.

- (1) Identify calls produced by animals of the target population that relate to animal density, i.e., calls that are produced by a known proportion of the population (e.g., adult males) with some regularity following a mean call production rate given, for example, as the number of calls produced by an individual per day.
- (2) Collect a sample of n calls using a well-designed survey, such as the calls detected in acoustic recordings.
- (3) Estimate the false-positive rate, that is, the rate of incorrectly classifying a detected sound as the call of the target species.
- (4) Estimate the average probability of detecting a call within the search area.

The authors included a long list of references together with instructions and a useful list of conclusions. Pérez-Granados and Traba (2021) carried out an extensive review of locating bird sounds using autonomous recording units (ARUs) for estimating bird densities or bird abundances. Acoustic methods are alluded to elsewhere in this book as they are becoming more widely used, with further examples described below.

The vocal activity rate index is a useful method to infer terrestrial bird abundance with acoustic monitoring (Pérez-Granados et al., 2019). Among the major obstacles for implementing ARUs in monitoring programs are the costs associated with acquiring recorders and the time and expertise required for the interpretation of recordings. However, the recent development of low-cost recorders and advances in computational capabilities to aid in audio interpretation have helped overcome these barriers. The authors mentioned some other difficulties as well. They noted that there is no consensus on how to estimate bird density using ARUs and referred to some recent methods. Darras et al. (2018) devised and tested a method for accurately estimating bird detection distances from sound recordings, using a reference recording of test sounds at different frequencies, emitted from known distances. Metcalf et al. (2022) concluded that high temporal resolution sampling of passive acoustic monitoring-based surveys should be considered the primary method for estimating the species richness of bird communities in tropical forests. Chambert et al. (2018) provided a hierarchical model for analyzing data from an automatic acoustic recording device.

Whitehead (2009) introduced a novel method using a passive acoustic survey where records of vocalizations detected or not detected at regularly spaced listening stations spaced at distance d were made on a line transect. There was a record for each station as to whether vocalizations were detected or not, not the number of vocalizers. The individuals were not assumed to move during a recording. It was assumed that they were randomly and independently distributed at a density of D and could be detected within a distance of W units. They produced audible (to listeners within range W) sounds with probability P during any listening. It was assumed probabilities were uncorrelated among listening stations if an individual was within audible range for ≥ 2 stations. The density of vocalizers is then DP . Assuming a Poisson model with mean $DP\pi W^2$, the probability of recording at least one vocalization at a station is $1 - \exp(DP\pi W^2)$. Thus, if vocalizations were heard at a proportion of \hat{p} stations, then given W and μ , an estimate of D is

$$\hat{D} = -\log(1 - \hat{p})/(P\pi W^2).$$

Whitehead (2009) gave estimators for three situations: (1) W known and P unknown, (2) W unknown and P known, and (3) W and P both unknown. Three variations of the method were considered using simulation, and it was concluded that “the methods did seem to be useful in some circumstances.” Also, “in some circumstances (e.g., night, fog, and thick vegetation), auditory or olfactory detection is more effective than sight, and ranges to animals cannot be easily measured.” Stevenson et al. (2015) developed methods that also used spatial information described in Chap. 14. Marine acoustical methods are described in Sect. 8.13.2.

8.9 Using Presence/Absence Data

In Chap. 4, we considered methods for handling presence/absence data with a main emphasis on occupancy and also along transects. They considered two cases, (i) independent detections and (ii) clustering within the detections of each individual, and developed a model using species detection/non-detection data accounting simultaneously for both dependence and abundance-induced heterogeneity in the detection process.

Suppose species detection surveys are carried out along one or more transects at a number of sampling sites (e.g., forest patches or quadrats), recording the location of each detection. The detection data do not allow individuals to be uniquely identified. If N_i is the number of individuals present at site i , it is assumed that the species detection process at each site results from the superposition of n_i identical independent point processes, one for each individual. As N_i is unknown, it is modeled as a random variable with some probability distribution $\Pr(N_i; \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a vector parameter describing site abundance. The probability of site occupancy is then $\psi = 1 - \Pr(0; \boldsymbol{\theta})$.

Case 1: Independent Detections

We assume that each individual point process can be modeled as a homogeneous Poisson process with rate γ , where γ is the average number of detections per unit length. With independence, we can add the Poisson variables so that, conditional on N_i , we have a Poisson process with rate (mean) γN_i . If d_i is the number of detections at site i , L_i is the total transect length surveyed at site i , and there are S sampling sites, we have the likelihood function, by summing out the distribution of N_i , namely,

$$L(\boldsymbol{\theta}, \gamma) \propto \prod_{i=1}^S \left[\sum_{N_i=0}^{\infty} \left\{ (\gamma N_i)^{d_i} \exp[-\gamma_i N_i L_i] \Pr(N_i; \boldsymbol{\theta}) \right\} \right].$$

The factors $L_i^{d_i}/d_i!$ are omitted as they are not needed for the likelihood. Several distributions can be used for N_i , for example, Poisson(δ), where δ is the average site abundance for the species, the negative binomial, and the Poisson log-normal distributions. Without going into details, for the Poisson distribution, the infinite summation can be avoided using the likelihood (Guillera-Arroita & Lahoz-Monfort, 2012)

$$L(\delta, \gamma) = \prod_{i=1}^S \left\{ \gamma^{d_i} \exp[-\delta + \delta e^{-\gamma L_i}] \sum_{k=0}^{d_i} S(d_i, k) (\delta e^{-\gamma L_i})^k \right\},$$

where $S(d_i, k)$ are Stirling numbers of the second kind (Graham et al., 1988, 243). As this expression involves only finite summations, it provides a convenient computational method.

If the mixing distribution given by $L(\theta, \gamma)$ is a nonparametric distribution with mass only at $N_i = 0$ and $N_i = m$, the model is equivalent to the occupancy model with Poisson detection process presented by Guillera-Arroita et al. (2011), with likelihood

$$L(\psi, \gamma) = \prod_{i=1}^S \left[\psi \lambda^{d_i} \exp[-\lambda L_i] + (1 - \psi) I(d_i = 0) \right],$$

where $I(d_i = 0) = 1$ when $d_i = 0$ (and 0 otherwise), $\gamma m = \lambda$, and the abundance distribution has $\Pr(0) = 1 - \psi$ and $\Pr(m) = \psi$.

Case 2: Clustered Detections

Hines et al. (2010) considered the effect of clustering in spatial models, which can occur with the sampling method used. Temporal or spatial replication can allow inference about the detection process. The authors noted that inference based on spatial replication strictly requires that replicates be selected randomly and with replacement, but the importance of these design requirements is not well understood. They focused on an increasingly popular sampling design based on spatial replicates that are not selected randomly and with replacement, often for logistical reasons, and are expected to exhibit Markovian dependence.

We consider designs with correlated spatial replication and refer to their example for tigers in southwestern India. Tigers are known to use forest roads and trails as travel routes and to mark them intensively with tracks, scent, and scats. It is sensible to survey such trails, being likely travel routes. However, because individual animals walk along trails for distances that can exceed segment length, detection of a sign on one segment will likely translate to an increased probability of detecting a sign on the next segment. This can possibly lead to spatial correlation between detections on adjacent sample fragments. This type of sampling design can be thought of generally as a cluster sampling design.

Suppose that animal presence is modeled as a first-order spatial Markov process. Let S be the number of sample units. It is assumed that in a unit, the survey begins at one end of a survey route (trail) and then proceeds in order along K total segments of the route. The data for each segment consists of 1 for detection and 0 otherwise. Let

ψ = Probability a sample unit is occupied

p = Probability of detection at a segment, given sample unit occupied and
species present on segment

θ = Probability species present on segment, given sample unit occupied and
species not present on previous segment

θ' = Probability species present on segment, given sample unit occupied and
species present on previous segment

Given a detection history $h_j = 01011$, the probability associated with this history under the Markov spatial process model is (suppressing subscript j)

$$\Pr(h_j = 01011) = \psi[(1 - \theta)\theta + \theta(1 - p)\theta'] \times p[(1 - \theta')\theta + \theta'(1 - p)\theta']p\theta'p.$$

As a further example, if there were just two segments, then

$$\Pr(h_j = 00) = (1 - \psi) + \psi[\theta(1 - p)(1 - \theta')p] + (1 - \theta)(1 - \theta)p].$$

For S samples, we then have the likelihood

$$L(\psi, \theta, \theta', p | h_1, \dots, h_s) = \prod_{j=1}^S \Pr(h_j),$$

and a general computing expression for $\Pr(h_j)$ is given as well as further details in the internet appendices of Hines et al. (2010). A new trap response occupancy was used as a rough approximation for data obtained under such a sampling design. Hopefully, the above gives some general idea of the method.

We mention in passing an alternative model of Skaug (2006), which also assumed that the detections from each individual exhibit some degree of clustering and therefore cannot be considered independent (e.g., individuals tend at times to use the route followed by the transect, as in the previous method). One possible way to account for this is to describe the detection process for each individual as a two-state Markov-modulated Poisson process (2-MMPP) in which the intensity is governed by an unobserved two-state Markov process. Guillera-Arroita et al. (2010) also considered this model in which detections take place at two different rates, $\gamma = (\gamma_1, \gamma_2)$, and the interval spent in each state is stochastic and controlled by parameters $\mu = (\mu_{12}, \mu_{21})$, the switching intensities between states.

The next step in the previous model above of Guillera-Arroita et al. (2012) is to incorporate distances between consecutive detections. Following case 1, we assume that we have N_i independent identical 2-MMPP distributions, the sum of which is 2-MMPP. Let R_i be the number of independent transects surveyed in site i , and let $M_{ij|N_i}$ be the likelihood contribution of transect j in site i , given that detections along it are described by the superposition of N_i independent realizations. The likelihood is then

$$L(\boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{\mu}) = \prod_{i=1}^S \left[\sum_{N_i=0}^{\infty} \left\{ \left(\prod_{j=1}^{R_i} M_{ij|N_i} \right) \Pr(N_i; \boldsymbol{\theta}) \right\} \right].$$

The construction of $M_{ij}|N_i$ is described in Guillera-Arroita (2011). If the mixing distribution in the above likelihood is set to a nonparametric distribution with masses at $N_i = 0$ and $N_i = 1$, then the likelihood becomes

$$L(\psi, \lambda, \mu) = \prod_{i=1}^S \left\{ \psi \prod_{j=1}^{R_i} M_{ij} + (1 - \psi) I(d_i = 0) \right\},$$

where $\Pr(0) = 1 - \psi$, $\Pr(1) = \psi$, and $M_{ij} = M_{ij|1}$. With strong clustering, it was found to be more effective to increase the length of the survey rather than the number of sampling sites: doubling the survey length gave better results than a fivefold increase in sampling sites. The proposed models, although developed for data collected along transects, could also be applicable to surveys in which detections are collected over a continuous interval of time, such as camera trap data. Covariates can also be incorporated into the model using a generalized linear model approach.

The MMPP process can also be used to provide a model for availability where some individuals are not available for observation, as, for example, with marine mammals that are submerged. Langrock et al. (2013) developed such a model for availability and probability of detection, given available, when the probability of detection $g(y)$ is unknown at zero distance $y = 0$. To incorporate an MMPP model into a distance survey, it is assumed that surfacing events in marine surveys are generated by an MMPP and then thinned according to the observer's detection process, given a surfacing event at some distance. A surfacing event does not always lead to a sighting, and the probability of a surfacing being sighted depends on the distance from observer to the surfacing. This model was extended by Borchers and Langrock (2015) to accommodate double-observer surveys by introducing a mark-recapture component. Their model also accommodated multiple detections of the same animal at different times.

8.9.1 Log-Linear Model

Barker et al. (2014) used plot replication to build a Bayesian hierarchical model of koala sightings by multiple observers, focusing on two or three observers. They used a log-linear model for incomplete contingency tables to allow for capture (sighting) heterogeneity induced by dependencies between sampling methods. (Such log-linear models for basic capture-recapture for both closed and open populations are described by Seber & Schofield, 2019: Chapter 10.) The usual notation was used with a 1 for sighting and 0 for no sighting giving possible capture histories $H_2 = \{11, 10, 10, 01\}$ for two observers and $H_3 = \{111, 110, 101, 100, 011, 010, 001\}$ for three observers. The unobserved categories 00 or 000 are represented by **0**.

Let y_{hs} denote the number of animals with sighting history h , where h belongs to H_2 or H_3 on plot $s = 1, \dots, S$. Let N_s be the number of animals on plot s , and

let \mathbf{y}_s be the vector of such counts, with y_{0s} the number of animals on plot s that were present but not seen. Then $N_s = n_s + y_{0s}$, where $n_s = \sum_h y_{hs}$. The counts of animals on plots, including the missing count, can be modeled as independent Poisson random variables conditional on plot-specific parameters $\boldsymbol{\xi}_s = \{\xi_{hs}\}$. To include N_s as a parameter explicitly and hence allow hierarchical modeling in terms of abundance, we can condition our model on N_s , which leads to

$$f(\mathbf{y}_s, y_{0s} | N_s, \boldsymbol{\xi}_s) = f_1(\mathbf{y}_s, y_{0s} | n_s, \boldsymbol{\xi}_s) f_2(n_s | N_s).$$

Here, f_1 is a multinomial distribution with index N_s and cell probabilities $\{\pi_{hs}\}_{h \neq 0}$, where

$$\pi_{hs} = \frac{\xi_{hs}}{\sum_{j \neq 0} \xi_{js}}.$$

Also, $n_s | N_s \sim \text{Binomial}(N_s, \psi_s)$, where

$$\psi_s = 1 - \frac{1}{1 + \sum_{j \neq 0} \xi_{js}}$$

is the probability that an animal on the plot is observed by at least one observer. To relax the assumption of independent observers, we can model the multinomial cell probabilities for, say, three observers using a log-linear model having coefficients b_{is} representing plot-specific main effects, with i indexing observer and s indexing plot. Also, we can have two-way interactions between observer i and observer j for each pairing of observers, and because of the missing cell with zero observations, the three-way interaction cannot be fitted. The model takes the form

$$\log(\xi_{hs}) = \begin{cases} b_{1s} + b_{2s} + b_{3s} + b_{12s} + b_{13s} + b_{23s}, & h = 111 \\ b_{1s} + b_{2s} + b_{12s}, & h = 110 \\ b_{1s} + b_{3s} + b_{13s}, & h = 101 \\ b_{1s}, & h = 100 \\ b_{2s} + b_{3s} + b_{23s}, & h = 011 \\ b_{2s}, & h = 010 \\ b_{3s}, & h = 001. \end{cases}$$

Then, taking advantage of the replications, the b_{rs} were modeled as random effects with independent and identical normal distributions $N(\beta_r, \sigma_r^2)$ for $r \in \{1, 2, 3, 12, 13, 23\}$. The authors then modeled the N_s as exchangeable Poisson (λ_s) distributions, where the $\log(\lambda_s)$ can be modeled with exchangeable normal random variables with covariates, namely, $N(\alpha_0 + \boldsymbol{\alpha}' \mathbf{x}_s, \sigma_\lambda^2)$, with covariates \mathbf{x}_s and prior distributions for the parameters. The authors fitted the model by Markov chain Monte Carlo implemented by the software **JAGS**.

8.10 Point or Line Transects?

In this section, we compare point and line counts with or without additional distance measurements. As noted by Buckland (2006), the great value of distance sampling is that it can provide estimates of abundance over very large areas with only modest resources for detections at a site, though field tests are typically conducted on small study sites. In determining whether to use a point or line transect, there are a number of factors that need to be taken into consideration. Obviously, it will depend on the species being investigated and the comparative cost factors in carrying out a survey. For example, Bollinger et al. (1988) considered a population of bobolink birds that was of known size because of intensive sampling and found that the line transect did better, though both seriously overestimated the density because of undetected movement of the birds toward the observer. Line transects took up more counting time, but less overall time because of logistics. In multispecies bird surveys using points, birds are easier to locate and identify when the recorder is standing still and when they can be located to habitat measurements. They are easier to use when walking in difficult terrain.

Watson and Quinn (1997) compared simulated counts using both methods for underwater visual census methods. They found that the transect and point-count methods performed equivalently in terms of bias and variability when fish do not move, the amount of sampled area is the same, and all fish within the sampled area are observed. However, the possible effects of incomplete detectability were not considered.

Buckland (2006) looked at several methods of conducting a point transect survey with regard to birds. The first method was the standard approach of recording any birds detected, together with estimated distances from the point, within a fixed count period. The second method, closely similar to the first, was the “snapshot” method of Buckland et al. (2001). The observer attempts to record the locations of detected birds at a snapshot moment, with time spent before this moment to identify and locate birds, and afterward to confirm locations. The method is a compromise between the ideal of recording bird locations at an instant in time and the need to spend some time at a point to be able to detect and identify birds, so that time-keeping is important.

The third method was based on recording distances to detected cues for a predetermined length of time at each point. A cue is defined as a single burst of song or the start of a songburst if it lasts for a significant length of time. It needs careful definition especially with birds that have a variety of songs and calls; acoustic recordings can be helpful. The fourth method used line transects along with standard distance methods and used parallel equally spaced lines. A fifth method called territory mapping was also used. The assumptions underlying all the methods were discussed in detail, and both simulated and field studies were carried out.

Buckland (2006) came to the following conclusions. Taking account of precision of estimates, line transect sampling was estimated to be the most efficient method for all four species investigated. This is likely to be the case generally in relatively

open habitat where a straight line can be walked without large deviations. However, in more difficult habitats and terrains, line transect sampling may be impractical or workable only by cutting transect lines—in which case, they may prove less efficient than point transect methods.

There was no clear winner among the point transect methods, and although the cue-count method tended to have poor precision and can be problematic with multispecies because of data overload for an observer, it is more robust with respect to departures from some of the assumptions. For example, the method remains valid even if birds are moving large distances during the count. As multiple songbursts from the same bird will often be recorded, thus violating the assumption of independence, distance methods are very robust to such violations, though this is not the case with goodness-of-fit testing and model selection.

Buckland et al. (2008) provided a lot of helpful practical information including strategies when representative sampling is difficult. They came to the following additional conclusions. For the same effort, line transect surveys have the following advantages: they generate more detections, because time is lost in moving between points for point transect sampling; they tend to have lower bias and higher precision, given the same number of detections and good-quality data, essentially because a higher proportion of detections is close to the observer than for point transects. Hence, point transect sampling should be considered only if it is believed that the quality of line transect data might be severely compromised. For example, random line transects may be difficult to navigate because of the terrain. It could be easier to get to a random point than to travel along a random line. This problem can occur in multispecies surveys, where the observer risks being “swamped” with records (e.g., birds). Also, it is easier to concentrate on detecting and recording everything if the observer does not have to also navigate along a line, and it is easier to record habitat at a point than along a line, making it simpler to investigate habitat associations. It is also easier to model variation in density through the survey region using point transect data.

In both methods, distances should be accurately measured. Buckland et al. (2001, 132, 176) noted that if all distances are overestimated by 10%, then line transect estimates of abundance are biased low by about 9% and point transect estimates are biased low by about 17%. If all distances are underestimated by 10%, then line transect estimates are biased high by about 11%, and point transect estimates by about 23%. If we have errors in our estimated distances but the mean error is zero (i.e., the estimates are unbiased), line transect estimates are usually approximately unbiased, whereas point transect estimates tend to be biased high. If the mean error is zero so that distance estimates are unbiased, line transect estimates are usually approximately unbiased, whereas point transect estimates tend to be biased high.

Considering just counts only, Moore et al. (2004) carried out a field comparison of two methods for estimating detectability and abundance for birds, a point-count removal (time-to-detection) method, where sampling is over several time intervals and once an individual is observed, it is then treated as being “removed,” and the double-observer method (considered in the next section). They found that when detection probabilities were relatively high for individual observers,

the 2 methods yielded similar estimates of density for nearly all 17 species modeled. However, when true detection probabilities for observers were relatively low, removal estimates of detectability and density were biased high and low, respectively, perhaps because of the effect of low-detection probability on the removal estimator or smaller sample sizes associated with less-skilled observers. In general, they considered removal modeling a more desirable approach than double-observer modeling. It required half as many observers and allowed more sources of variation in detectability to be modeled, as well as estimates abundance or density of the true population of birds.

On the other hand, double-observer modeling estimates only the abundance of the “apparent” population (i.e., those birds that are visually or audibly conspicuous). For species that vocalize infrequently or are otherwise elusive, the apparent population may be significantly smaller than the true population. However, double-observer modeling is more robust to violations of the assumption of population closure and may outperform removal methods when data are collected by less-experienced observers.

Golding and Dreitz (2016) also compared the same two methods for five bird species. The comparison depends on the criteria used as, for example, there are trade-offs in the amount of effort required for each method based on the field surveys. As already noted, only one observer is required for point-count method, but two are required for the double-observer method. The authors found that for tracking changes in abundance of multiple songbird species in open habitats, the double-observer method is suggested as an alternative. It provided more precise information for a similar amount of effort compared to the point-count method. However, point-count survey methods may provide better estimates in more closed habitats or be the better method to use, given logistical and financial constraints.

8.11 Method of Double Observers Using Distances

The key assumption made for many terrestrial species is that $g(0) = 1$ so that all the animals in the observer’s path are detected. This assumption may not be tenable as there are two main biases that may lead to violation of this assumption: availability bias and perception bias. Availability bias occurs when animals are missed because they are not available for detection, a common problem with long-diving species as in marine surveys of many whale and dolphin species where the animals can be underwater when the vessel passes over them. Perception bias occurs when animals are available for detection but are missed for other reasons, such as cryptic surfacing behavior, poor weather conditions, or observer fatigue. For shipboard surveys, there are typically two visual platforms on the same survey vessel (double-platform methods). Ways around this problem of $g(0) \neq 1$ using the method of double observers introduced in Sect. 8.11 are now considered in more detail. The method can be applied to both line transects and point counts under the title of MRDS (mark-recapture distance sampling) described below.

8.11.1 Using Mark-Recapture and Distance

The method we now consider, also called a mark-recapture line transect (MRLT) survey, and can be used with two independent observers when $g(0) \neq 1$ or if animals can be hard to see such as polar bears in white terrain. We first consider some of the earlier work on this topic before looking at the problem of some possible dependence between observers. Alpizar-Jara and Pollock (1996) evaluated the performance of three variations of the Lincoln-Petersen estimator for line transects, namely, the overall estimator, the stratified estimator, and the general stratified estimator, using simulation. However, possible dependence of observers could severely bias the population size estimate due to heterogeneity. This is discussed later in Sect. 8.11.3.

Alpizar-Jara and Pollock (1999) gave a slight modification of their previous model in which two observers are replaced by two different sampling techniques. If the probabilities from the two methods are independent, the problem of heterogeneity can be neglected (Seber, 1982, 81–89). The first sample is the usual marking sample, or one can utilize those with natural identifying marks, as with some animals. The second one is a traditional line transect survey with a random placement of lines, where the perpendicular distances of individuals from the transect of all individuals sighted are recorded, as well as noting those that are marked (resighted). Let n_{11} be the number of those objects marked during the first sampling occasion and also seen on the second occasion, n_{01} the number unmarked observed during the line transect survey, and n_{10} the number of marked not seen during the transect survey. Then $n_1 = n_{11} + n_{10}$ is the number marked in the first sample, $n_2 = n_{11} + n_{01}$ the number detected in the line transect survey, and $n_\omega = n_{11} + n_{10} + n_{01} = n_1 + n_2 - n_{11}$ the total number detected during the two sampling occasions.

We recall some of the above notation. Let W be the usual half-width of the transect of length L , θ the vector parameter of the detection function $g(y) = g(y | \theta)$ for the transect, and $g(0)$ the probability of detecting an individual on the transect line. Let $g(y | \theta) = g(0) \times g^*(y | \theta^*)$, where $\theta = (g(0), \theta^*)$. It is assumed that $g^*(0 | \theta^*) = 1$. Let

$$\mu = \int_0^W g(y) dy, \text{ the so-called effective strip half-width}$$

$f_2(y | \theta) = g(y) / \mu$, the density function of the perpendicular distances, y_i

p_1 = the probability of an animal being marked in the first sample

p_2 = the probability of an animal being sighted in the second sample

$p_\omega = p_1 + p_2 - p_1 p_2$, the probability of detecting an animal in either the first or second sample or both

N_ω = total number of animals in the covered area $A_\omega = L \times 2W$

The usual assumptions apply, namely, the population is closed, all animals are equally likely to be captured in the first sample, no marks lost or overlooked, independence between marked animals and sightings, and detections independent. This allows a combined likelihood of the Petersen method (Sect. 12.2) and line transect method, namely, $L = L_1 L_2$, where

$$\begin{aligned} L_1 &= f(n_{11}, n_{10}, n_{01} \mid N_\omega; p_1, p_2(\boldsymbol{\theta})) = \\ &\quad \binom{N_\omega}{n_{11}, n_{10}, n_{01}} [p_1 p_2(\boldsymbol{\theta})]^{n_{11}} [p_1 \{1 - p_2(\boldsymbol{\theta})\}]^{n_{10}} [(1 - p_1) p_2(\boldsymbol{\theta})]^{n_{01}} \\ &\quad \times [(1 - p_1) \{1 - p_2(\boldsymbol{\theta})\}]^{N_\omega - n_{11} - n_{10} - n_{01}} \\ L_2 &= f(\{y_i\} \mid n_2; \boldsymbol{\theta}) = \prod_{i=1}^{n_2} f_2(y_i \mid \boldsymbol{\theta}) = \prod_{i=1}^{n_2} \frac{g(y_i)}{\int_0^W g(y) dy}, \end{aligned}$$

where

$$p_2(\boldsymbol{\theta}) = \int_0^W g(y \mid \boldsymbol{\theta}) dy / W = \mu / W.$$

Using the full likelihood $L = L_1 L_2$ and an appropriate model for $g(y \mid \boldsymbol{\theta})$, we can obtain an estimator of $g(0)$ and the maximum likelihood estimator $\widehat{N}_{FL} = n_\omega / \widehat{p}_\omega$ of N_ω , where

$$\widehat{p}_\omega = \widehat{p}_1 + \widehat{p}_2(\widehat{\boldsymbol{\theta}}) - \widehat{p}_1 \widehat{p}_2(\widehat{\boldsymbol{\theta}})$$

and $n_\omega = n_1 + n_2 - n_{11}$. Since N_ω is a random variable, we have

$$\text{var}[\widehat{N}_{FL}] = E[\text{var}(N_\omega \mid N_\omega)] + \text{var}[E(\widehat{N}_\omega \mid N_\omega)].$$

To find this, a distribution needs to be given for N_ω , such as a Poisson process. If just L_1 is used with p_2 constant, we obtain the usual Petersen-Lincoln estimate \widehat{N}_{LP} , while if L_2 is used, the line transect estimator \widehat{N}_{LT} can be found. The authors compared all three estimators and found that \widehat{N}_{FL} performed the best as well as providing an estimate of $g(0)$. Further theoretical and computational details are given by (Alpizar-Jara & Pollock, 1999).

Manly et al. (1996) developed a theory that allowed for group size g . If $P_i(y, g)$ is the probability of sighting a group of size g at distance y from the line by observer i , this can be expressed as a logistic function of covariates. Assuming the key assumption of independent observers, the probability of a group of size g being seen by at least one observer, given the animal is in the covered region, is (using $P(A \cup B) = P(A) + P(B) - P(A \cap B)$)

$$P(g) = \frac{1}{W} \int_0^W [P_1(y, g) + P_2(y, g) - P_1(y, g)P_2(y, g)] dy. \quad (8.6)$$

A likelihood function can then be constructed, namely, $L = L_1 L_2 \cdots L_n$ for n groups seen, where L_i can take one of the three cases depending on whether the group is seen by just observer 1, just observer 2, or both observers. Abundance estimates are also given.

A variation on the above method, discussed by (Borchers et al., 1998), used the Horvitz-Thompson estimator and a slight variation on the capture-recapture model. It is called the dual platform model and allows the two observer teams (platforms) to do different things. Let n be the number detected. If P_j is the probability of detecting the j th animal ($j = 1, \dots, N_L$), an unbiased estimate of N_L is $\sum_{j=1}^n 1/P_j$, assuming the P_j are known. If P_j is modeled as some function of a parameter vector θ using, for example, a logistic function with covariates for the $g_i(y)$, we can then estimate N_L by $\sum_{j=1}^n 1/\widehat{P}_j$. Here, $\widehat{P}_j = P_j(\widehat{\theta})$, and $\widehat{\theta}$ is some estimator of θ .

Quang and Becker (1997) developed a model for one- and two-sided surveys, with independent observers, with distance classes, and with a possible blind strip as follows, using a transect of total length L . Let $W = h - d_0$, where h is the width of the strip and d_0 the width of the blind strip. If $g_i(y)$ is the detectability function for observer i ($i = 1, 2$), the probability that animal located at distance y is sighted by observer 1 only is $g_1(y)[1 - g_2(y)]$, for observer 2 only is $g_2(y)[1 - g_1(y)]$, and for both observers is $g_1(y)g_2(y)$. The detection functions g were modeled by truncated Weibull distributions. It was assumed that all the animals are available for detection, whether detected or not, and are uniformly distributed over the sampling region, giving $\pi(y) = 1/W$ ($d_0 \leq y \leq h$). Then it transpires that q_{1k} , the probability that animal is sighted by observer 1 only between the distances d_{k-1} and d_k , is

$$q_{1k} = \int_{d_{k-1}}^{d_k} g_1(y)[1 - g_2(y)]\pi(y) dy = \frac{1}{W} \int_{d_{k-1}}^{d_k} g_1(y)[1 - g_2(y)] dy.$$

Similar expressions apply to q_{2k} for observer 2 and q_{12k} for both observers. Let n_{1k} , n_{2k} , and n_{12k} be the corresponding observed counts in the distance classes, let $n_T = \sum_k (n_{1k} + n_{2k} + n_{12k})$, and let $q_T = \sum_k (q_{1k} + q_{2k} + q_{12k})$. Then the array

$$(\{n_{1k}\}, \{n_{2k}\}, \{n_{12k}\}, N_L - n_T)$$

is multinomially distributed with probabilities

$$(\{q_{1k}\}, \{q_{2k}\}, \{q_{12k}\}, 1 - q_T),$$

giving a likelihood that is a function of the Weibull parameters and of N_L , the number of animals in the region of detection (also called the covered region). Maximum likelihood estimates of the parameters can then be obtained giving \widehat{N}_L . An estimate of N , the total population size on area A is $\widehat{N} = \widehat{N}_L A/a$, where a is the covered area LW .

Borchers et al. (1998) developed a comprehensive extension of the model of (Quang & Becker, 1997), which included the other models as special cases. They

gave an extensive discussion of various types of mark-recapture line transect models and added the additional feature of assuming that N_L is Binomial(N, P_a), where $P_a = a/A$. The authors considered binned (distance classes) and unbinned data and also used covariates incorporated in a more general way, including, for example, the perpendicular distance, the group size if the unit in the study is a group rather than an individual, the sea state (in maritime surveys), the observer index, platform index, climate, and so forth. The covariates can be used for heterogeneity in the sighting probabilities. Further details are given by Borchers et al. (1998).

Unfortunately, using two observers does not necessarily solve the problem of $f_c(0) \neq 1$. For example, Smolensky and Fitzgerald (2010) applied the method to dune-dwelling lizards and were able to compare it with the actual numbers by using total removal sampling. They found that the method greatly underestimated densities as it only accounted for visibility bias and did not account for lizards that were inactive and beneath the surface during surveys. Only active lizards that were missed by the first observer and detected by the second observer were quantified. The theory only applies when all animals are available for sighting.

8.11.2 Allowing for Heterogeneity

Borchers et al. (2006) developed the following model for mark-recapture distance sampling (MRDS). Let N be the number of animals in the survey region of area A , and let N_c be the number in the region searched on the survey, the covered region. Let y represent the usual perpendicular distance, \mathbf{z} a vector of other covariates that affect detection probability, and $g_j(y_i, \mathbf{z}_i)$ the probability that an object at y with variables \mathbf{z} is detected by observer j ($j = 1, 2$). Let $\boldsymbol{\theta}$ be the unknown parameter vector of $g_1(y_i, \mathbf{z}_i)$ and $g_2(y_i, \mathbf{z}_i)$ (with some, none, or all parameters shared between observers). It is assumed that y and \mathbf{z} are independent of each other and that they are independent between animals. The probability density function (pdf) of the perpendicular distance of an object (detected or not) is $\pi(y)$. Animals are assumed to be uniformly distributed in two-dimensional space and with the usual truncation of observations at distance W , $\pi(y) = 1/W$ for line transects and (from Sect. 7.2.1 with $y = r$) $2y/W^2$ for point transects. The pdf of \mathbf{z} is $f(\mathbf{z}; \boldsymbol{\phi})$, with some unknown parameter vector $\boldsymbol{\phi}$. (We temporarily suspend notational dependence on $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ for convenience.)

If n animals are detected, we have the mark-recapture component of the overall likelihood function

$$L_{\omega}(\boldsymbol{\theta}) = \prod_{i=1}^n \frac{\Pr(\omega_i \mid y_i, \mathbf{z}_i)}{g_+(y_i, \mathbf{z}_i)},$$

where ω_i is the observed capture history for detected animal i (one of (1, 0), (1, 1), (0, 1)) and $g_+(y_i, \mathbf{z}_i)$ is the probability that an object at distance y_i with variables

\mathbf{z}_i is detected by at least one observer. Assuming independent detection, and given (y_i, \mathbf{z}_i) ,

$$g_+(y_i, \mathbf{z}_i) = g_1(y_i, \mathbf{z}_i) + g_2(y_i, \mathbf{z}_i) - g_1(y_i, \mathbf{z}_i)g_2(y_i, \mathbf{z}_i)$$

and

$$\Pr(\boldsymbol{\omega}_i \mid y_i, \mathbf{z}_i) = \prod_{j=1}^2 g_j(y_i, \mathbf{z}_i)^{\omega_{ij}} [1 - g_j(y_i, \mathbf{z}_i)]^{1-\omega_{ij}}.$$

The authors called this the assumption the “full independence” (FI) assumption.

The next component of the likelihood is given by

$$L_{y|\mathbf{z}}(\boldsymbol{\theta}) = \prod_{i=1}^n \frac{g_+(y_i, \mathbf{z}_i)\pi(y_i)}{g_+(\mathbf{z}_i)},$$

where

$$g_+(\mathbf{z}_i) = \int_0^W g_+(y_i, \mathbf{z}_i)\pi(y)dy$$

is the probability that at least one observer detects animal i . There is an identifiability problem here as a common constant can be introduced in both the numerator and denominator, which cancels out. It is therefore necessary to treat the detection probability at distance $y = 0$ (or some other point) as known. Let $g_+ = \int g_+(\mathbf{z})f(z)d\mathbf{z}$, where the integral sign stands for a multiple integral or summation if necessary.

The \mathbf{z} component is

$$L_{\mathbf{z}}(\boldsymbol{\phi}, \boldsymbol{\theta}) = \prod_{i=1}^n \frac{g_+(\mathbf{z}_i)}{g_+}.$$

The final component of the likelihood is

$$L_n(N_c, \boldsymbol{\theta}, \boldsymbol{\phi}) = \binom{N_c}{n} g_+^n (1 - g_+)^{N_c - n},$$

the only component containing N_c . The full likelihood is

$$L(N_c, \boldsymbol{\theta}, \boldsymbol{\phi}) = L_n(N_c, \boldsymbol{\theta}, \boldsymbol{\phi}) L_{\mathbf{z}}(\boldsymbol{\phi}, \boldsymbol{\theta}) L_{y|\mathbf{z}}(\boldsymbol{\theta}) L_{\omega}(\boldsymbol{\theta}).$$

In multiple covariate distance sampling (MCDS), estimators typically estimate $\boldsymbol{\theta}$ from $L_{y|\mathbf{z}}(\boldsymbol{\theta})$ and then estimate N_c using a Horvitz-Thompson-like estimator

(Marques & Buckland, 2003)

$$\widehat{N}_c = \sum_{i=1}^n \frac{1}{\widehat{g}_+(\mathbf{z}_i)}.$$

If the detection probability at distance $y = 0$ is known (usually 1), the mark-recapture likelihood component $L_\omega(\boldsymbol{\theta})$ is unnecessary because conventional distance sampling methods can be used. Conversely, if the detection probability is not known at any distance, then this component is essential, and the authors give three possible options to proceed further, namely: (1) estimation of $\boldsymbol{\theta}$ (and hence $\widehat{g}_+(\mathbf{z}_i)$) from $L_\omega(\boldsymbol{\theta})$ alone, followed by the use of a Horvitz-Thompson-like estimator to estimate N_c ; (2) estimation of $\boldsymbol{\theta}$ from $L_{y|\mathbf{z}}(\boldsymbol{\theta})L_\omega(\boldsymbol{\theta})$, followed by the use of a Horvitz-Thompson-like estimator to estimate N_c ; and (3) estimation of $\boldsymbol{\theta}$ and N_c from the full likelihood $L(N_c, \boldsymbol{\theta}, \phi)$.

8.11.3 Levels of Independence

Point independence (PI), discussed in Sect. 7.7.1, assumes that there is no unmodeled heterogeneity (i.e., that there is independence) at a given point, typically at the observation point. (cf. Burt et al., 2014 for a helpful discussion of this problem.) Borchers et al. (2006), referred to above, gave an example where the model FI (full independence) is wrong and introduced a PI (point independence) example. They also discussed interval estimation, model selection methods, and goodness-of-fit diagnostics. Extensive details are given in their paper. The difference between FI and PI is discussed further by Burt et al. (2014) for line transects. They noted that although PI is less restrictive than FI, PI is untenable if there is responsive movement before detection and, in this case, it may be preferable to assume FI.

Buckland et al. (2010) also considered the question of dealing with heterogeneity when using double observers, but allowing for some level of dependence between the two observers, thus generalizing the previous section. The usual uniform perpendicular distance assumption was used (from a random transect) where they diagnosed dependence by (a) modeling the shape of observer j 's detection function $g_j(y)$, ($j = 1, 2$), (b) modeling the conditional probability $g_{j|k}(y)$ that observer j detects an animal at y given that observer k detected it ($j = 1, 2, k = 3 - j$), and (c) modeling the covariance in the observers' detection probabilities as a function of y using a function $\delta(y)$ defined below.

In practice, the full independence assumption ($g_{j|k}(y) = g_j(y)$) cannot necessarily be made at each distance, as at greater distances, only the most detectable animals tend to be recorded and those that are detected by one observer are therefore more likely to be detected by the other observer. Although we can anticipate less dependence between detections on the line than at greater distances, unless detection on the line is certain, it seems possible that some dependence remains. Now if $g_{12}(y)$

is the probability that an animal is seen by both observers, then having independent observers implies that $g_{12}(y) = g_1(y)g_2(y)$ and the probability that the animal is seen by at least one observer is then

$$g_+(y) = g_1(y) + g_2(y) - g_1(y)g_2(y).$$

This is the theory of the previous section. In relaxing the full independence assumption, the authors introduced $\delta(y)$, a measure of dependence between the detection probabilities so that $g_{12}(y) = \delta(y)g_1(y)g_2(y)$, where $\delta(y)$ is related to the covariance $\sigma_{12}(y)$ between the detection probabilities $g_1(y)$ and $g_2(y)$ as

$$\sigma_{12}(y) = [\delta(y) - 1]g_1(y)g_2(y).$$

We then have

$$g_+(y) = g_1(y) + g_2(y) - \delta(y)g_1(y)g_2(y) \quad (8.7)$$

and $g_{j|k}(y) = \delta(y)g_j(y)$ ($j = 1, 2, k = 3 - j$). A number of alternative expressions for $\delta(y)$ are given, which demonstrates that $\delta(y)$ measures the discrepancy between the conditional detection functions $g_{j|k}(y)$ derived from the mark-recapture data and the unconditional detection functions $g_j(y)$, which are derived from distance sampling data with the requirement that $g_j(y^*)$ is known for some y^* . We recall that with distance sampling with a single observer, the standard assumption is $g_j(0) = 1$. With double observers, this often untenable assumption can be replaced with the assumption of full independence, namely, $\delta(y) = 1$ for all y , or with point independence $\delta(y^*) = 1$ at a specified y^* , usually $y^* = 0$ (Laake & Borchers, 2004). It transpires that fitting full independence models to data requires a functional form for $g_j(y)$ and for point independence and requires a model for $g_{j|k}(y)$. Neither require a model for $\delta(y)$.

The authors, Buckland et al. (2010), considered relaxing the assumption of full independence by imposing the following restrictions on $\delta(y)$.

- (1) $\delta(y) \leq u(y)$, where $u(y) = \min\{1/g_1(y), 1/g_2(y)\}$, which ensures that $g_{j|k}(y) \leq 1$.
- (2) $\delta(y) \geq l(y)$, where $l(y) = \max\{0, [g_1(y) + g_2(y) - 1]/[g_1(y)g_2(y)]\}$, which ensures that $g_+(y) \leq 1$.

With the above restrictions, we have the useful result that if $g_j(y) \rightarrow 1$ ($j = 1, 2$), then $\delta(y) \rightarrow 1$. If we define $\delta_0(y) = [\delta(y) - l(y)]/[u(y) - l(y)]$, it is restricted to the unit interval and can be represented by an appropriate functional form such as a logistic, in particular

$$\log\left(\frac{\delta_0(y)}{1 - \delta_0(y)}\right) = \alpha + \beta y + \log\left(\frac{1 - l(y)}{u(y) - 1}\right),$$

giving three levels of independence, namely,

- (1) Setting $\alpha = 0$ specifies point independence at $y^* = 0$.
- (2) Setting $\alpha = \beta = 0$ specifies a full independence model.
- (3) If $\beta = 0$ and $\alpha \neq 0$, we have a model with constant dependence for all y .
- (4) If $\alpha \geq 0$ and $\beta \geq 0$, then we are restricted to a model with either independence or positive dependence.

This gives us a whole range of methods for limiting independence. We now derive the likelihood L for the above double-observer method, which is a product of separate likelihoods (described more fully above), namely,

$$L = L_n L_{\mathbf{z}} L_{y|z} L_{\omega},$$

where n is the total number of animals detected by at least one observer; \mathbf{z} are any observation-specific covariates; $y|z$ represents the conditional distribution of the distance y , given covariates \mathbf{z} ; and ω represents the capture-recapture data. Just two components of the full likelihood, namely, $L_{y|z}$ and L_{ω} , can be used to avoid making distributional assumptions about n and \mathbf{z} , as estimation is not robust to failure of such assumptions. Inference is then conditional on n and \mathbf{z} , using a design-based approach to allow for variation in n . This means $L_{y|z}$ incorporates the assumption of a uniform distribution of animals perpendicular to transect lines and reduces to L_y if there are no covariates, which we assume for convenience, so that $L = L_n L_y L_{\omega}$, but we do not need L_n because of the uniform distribution assumption. Hence,

$$L_y = \prod_{i=1}^n f_+(y_i) = \prod_{i=1}^n \frac{g_+(y_i) f_Y(y_i)}{\text{E}[g_+(y_i)]},$$

where $g_+(y_i)$ is given by (8.7), $f_+(y)$ is the probability density function (pdf) of the detection distance y of an animal detected by at least one observer, $f_Y(y)$ is the unconditional pdf of distance y in the population (whether detected or not), and from Laake and Borchers (2004, 114)

$$\text{E}[g_+(y)] = \int_0^W g_+(y) f_Y(y) dy.$$

In this case, it is assumed that $f_Y(y) = 1/W$. Also,

$$L_{\omega} = \prod_{i=1}^n \frac{\text{Pr}(\omega_i)}{g_+(y_i)},$$

where

$$\begin{aligned}\Pr(\omega_i = (1, 0) | y_i) &= g_1(y_i)[1 - g_2(y_i)\delta(y_i)], \\ \Pr(\omega_i = (0, 1) | y_i) &= g_2(y_i)[1 - g_1(y_i)\delta(y_i)], \\ \Pr(\omega_i = (1, 1) | y_i) &= g_1(y_i)g_2(y_i)\delta(y_i).\end{aligned}$$

The likelihoods for full, point, and limiting independence only differ in the definition of $\delta(y)$. However, if the full independence assumption holds, then it is only necessary to use L_ω . With the point independence assumption, L_ω and L_y can be maximized independently using models for $g_{j|k}(y)$ and $g_j(y)$, which separate into the two respective likelihood components. The authors modeled the detection functions using a logistic approach, namely,

$$\text{logit}(g_j(y)) = \lambda_{0j} + \lambda_{1j}y \quad \text{for } j = 1, 2.$$

Certain diagnostic checks are suggested by the authors, Buckland et al. (2010), as when fitting limiting independence models, the Hessian matrix is sometimes nearly singular due to high correlation between the estimates of $g_j(y)$ and $\delta(y)$ at $y = 0$. To estimate abundance, the likelihood conditional on n can be maximized when models are given for $g_1(y)$, $g_2(y)$, and $\delta(y)$. Then bringing covariates back in, $E[g_+(\mathbf{z}_i)]$ can be estimated giving

$$\begin{aligned}\widehat{N}_c &= \sum_{i=1}^n \frac{1}{\widehat{E}[g_+(\mathbf{z}_i)]} \\ &= \frac{n}{\widehat{E}[g_+]}, \text{ if covariates are absent,}\end{aligned}$$

and $\widehat{N} = A\widehat{N}_c/a$, where $a = 2WL$ and L is the total length of the transect. Details of variance estimation are given as well as including the case for clusters. Covariates can be introduced by replacing y by y, \mathbf{z} in the above theory. In addition to using simulation, the methods were applied to a shipboard survey of minke whales using a covariate of sea state. The theory adopts readily to point-count methods.

Roy et al. (2021) developed a hierarchical Bayesian model for dependent double observers for estimating waterfowl breeding pair abundance from helicopters. Front and rear pairs of observers were used differently. During the survey, the primary observers reported to the secondary observers all flocks of birds they observed in front of an imaginary line between the observers. The secondary observers confirmed the observations reported by the primary observers and identified additional flocks of birds that were unreported by the primary observers once the flock was past the imaginary line. Observations from each pair were treated as a single observer, and one pair of observers were assigned the “primary” role ($a = 1$) and the other pair assigned the “secondary” role ($a = 2$). This dependence between counts of primary and secondary observers was modeled using a multinomial distribution with

index 1, namely, for $a = 1, 2$,

$$\mathbf{y}_{af} \sim \text{Multinomial}(1, \boldsymbol{\theta}_{af}),$$

where \mathbf{y}_{af} is an f -dimensional vector representing flock f that has been seen in the survey area by team a at various sites and $\boldsymbol{\theta}_{af}$ is the vector of probabilities that team a observed flock f . For any given flock, the elements of $\boldsymbol{\theta}_{1f}$ can be expressed in terms of p_{1ij} , the probability of the primary observer detecting species j on plot i . Similarly, the elements of $\boldsymbol{\theta}_{2f}$ were defined as the probability that the second observer detected the flock given primary observer did not detect, namely, $(1 - p_{1ij})p_{2ij}$. These probabilities were then expressed as logistic models with covariates of seat position and experience. The data consisted of the total number of flocks observed in plot i for species j . Further modeling details were given by the authors Roy et al. (2021).

8.11.4 Point Independence and Capture-Recapture

Fewster and Pople (2008) compared three mark-recapture distance sampling (MRDS) methods for eastern gray kangaroos. Two methods were based on the assumption of full independence between observers in the mark-recapture component, and this appeared to introduce more bias in density estimation than it resolves through allowing uncertain trackline detection. Both of these methods gave lower density estimates than conventional distance sampling, indicating a clear failure of the independence assumption. The third method, discussed above and termed point independence, appeared to perform very well, giving credible density estimates and good properties in terms of goodness of fit and percentage coefficient of variation. The model applied is described in Sect. 7.7, but with clusters, using the likelihood

$$L = L_n L_{\mathbf{z}} L_{y|\mathbf{z}} L_{\omega},$$

where now $L_{\mathbf{z}}$ models the distribution of detected cluster sizes. The point independence method uses only two of the likelihood components $L_{y|\mathbf{z}} L_{\omega}$, omitting the more controversial components $L_{\mathbf{z}}$ and L_n , which has some advantages and disadvantages. Let $p_o(y, z)$ be the probability that observer o detects a cluster of size z located at distance y from the transect line, for observers $o = 1, 2$. The authors proposed four models for this probability. Bootstrap methods were used for variance estimation.

Burt et al. (2014) provided a summary and overview of the above methods showing how conventional distance methods can be suitably linked to capture-recapture (MRDS methods). They focused particularly on model options, and we recall from previous discussions the FI (full independence) and PI (partial independence) models, which were compared by the authors. Here, the MR component allows

the probability of detection on the trackline to be estimated in an MRDS model. Unbiased estimation using PI requires that duplicate detections are independent along the trackline only, whereas FI requires that detections are independent at all distances. Although it is less restrictive than FI, PI is untenable if there is responsive movement before detection, and, in this case, it may be preferable to assume FI.

Different observer configurations have been mentioned, and, with two observers (or two observer platforms), two search configurations are possible. These are searchers being independent of each other or having a one-way independence. The authors also mentioned a trial configuration where the function of one of the observers (whom we will call observer 2 or the “tracker”) is to simply generate trials for the other observer (observer 1, also called the “primary” observer). The idea is that the tracker searches a region sufficiently far ahead that animals are unlikely to have reacted to the observer’s presence before being detected. Searching far ahead introduces a separation between the times animals are first available for observer 2 and first available for observer 1 to detect. This helps reduce correlation due to availability and will remove it altogether if the time separation is sufficiently long. This can also help reduce bias due to responsive movement.

The methods described above are available as an **R** (**R** Core Team 2012) package called *mrds* (Laake et al., 2013), and so can be accessed directly via the **R** environment, but this does require some understanding of **R**. However, the **R** package can now be accessed via **DISTANCE** (Thomas et al., 2010), and this provides an accessible route to these analysis options for those not familiar with **R**.

Becker and Christ (2015) introduced a PI model for the MRDS method with particular emphasis on aerial surveys, which often have a blind strip of width W_b (Quang & Becker, 1996). It is also assumed that there is independent sighting data between observers only at some distance $y = y^*$. Aerially collected distance sampling data can have a unimodal shape and have been successfully modeled with a gamma detection function. Covariates in gamma detection models cause the apex of detection to shift depending upon covariate levels, making this model incompatible with the PI assumption when using double-observer data.

An alternative detection model is the split-normal distribution, which is essentially two half-normal distributions that share a common mode (μ) but have different variances, namely, $\exp[-(y_i - \mu)^2 / 2\sigma_{1i}^2]$ for $W_b \leq y_i < \mu$ and $\exp[-(y_i - \mu)^2 / 2\sigma_{2i}^2]$ for $y_i \geq \mu$, with $\mu = y^*$. The $\log(\sigma_i)$ ($i = 1, 2$) were expressed as linear models in some covariates. The overall model consisted of four component likelihoods, namely, a distance component, a mark-recapture component, a binomial component relating to population size for the covered region, and a covariate component, as described above. Becker and Quang (2009) previously used a gamma-shaped detection function that depended on the observer, the group of animals, and covariates.

Conn et al. (2012) also used the PI method and applied a hierarchical modeling framework for multiple-observer line transects. Here, abundance intensities can be modeled as a function of habitat covariates, making it easier to extrapolate to unsampled areas. Unobserved animals and their covariates were modeled using a reversible jump Markov chain Monte Carlo algorithm, and observer detections were

modeled via a bivariate normal distribution on the probit scale, with dependence induced by a distance-dependent correlation parameter.

8.12 Aerial Censusing

The previous section also applies to aerial censusing, though there can be some additional problems such as animal movement (Fewster et al., 2008), which can interfere with the uniform assumption for the animal perpendicular distance. A pioneering paper by Cook and Jacobson (1979) proposed applying the double-observer method to groups using quadrat sampling. They used two observers designated as primary and secondary observer where the primary observer behaves as if he or she was the only observer present. The secondary observer confirmed all sightings by the primary observer, and records only those groups that he or she detected that were missed by the primary observer. Another pioneering paper using two aircraft was described by Hiby and Lovell (1998) for dealing with missed observations on the trackline.

Manly et al. (1996) used aerial censusing with a helicopter to estimate the abundance of groups of polar bears (*Ursus maritimus* Phipps) off the coast of Alaska. Polar bears occur in low densities across extensive areas of polar ice, which presents several logistical and statistical problems, including the detection of white bears on a white substrate, as mentioned in the previous section. Detection probabilities were expressed as functions of distance (y), group size (G), and possibly other covariates. One difference they used was to have two pairs of observers, one each side of the aeroplane, with two positions for each pair—front- and rear-seat positions. A curtain was used between members of each pair of observers to insure independence of sightings. Also, the internal communication system was constructed so that an observer could announce sightings to the data recorder without alerting the other member of the pair.

Additional features were incorporated such as individuals in an observer pair were rotated so that equal amounts of time were spent in the front and rear seats. It was assumed that the distance of groups of animals from the transect line has a uniform distribution over the interval $(0, W)$. Let $g_j(y, G)$ be the probability of a group of size G at distance y from the line being sighted from the j th seating position, given the group was in the covered area. It was assumed that

$$\text{logit}(g_j(y, G)) = \alpha_{0j} + \alpha_{1j}y + \alpha_{2j}G \quad (8.8)$$

for $G = 1, 2, \dots$, $j = 1, 2$, and $0 \leq y \leq W$. Other covariates could be added if needed. In what follows, it is assumed that the group is in the covered region.

The probability of a group of size G being sighted in the small interval $(y, y+dy)$ from the line from position 1 only is

$$\Theta_1(y, G) = (1/W)g_1(y, G)(1 - g_2(y, G))dy,$$

from position 2 only

$$\Theta_2(y, G) = (1/W)(1 - g_1(y, G))g_2(y, G)dy,$$

and from both positions

$$\Theta_{12}(y, G) = (1/W)g_1(y, G)g_2(y, G)dy.$$

Then, the total probability that a group of size G is detected is by summing the three probabilities and integrating over possible values of y

$$\begin{aligned} P(G) &= \int_0^W [\Theta_1(y, G) + \Theta_2(y, G) + \Theta_{12}(y, G)]dy \\ &= (1/W) \int_0^W [g_1(y, G) + g_2(y, G) - g_1(y, G)g_2(y, G)]dy, \end{aligned} \quad (8.9)$$

which corresponds to (8.4). The likelihood for an observed set of data is the joint probability of the distances from the line and the type of observation (seen from position 1 only, seen from position 2 only, or seen from both positions), conditional on groups being sighted. Given that that n groups are sighted, the likelihood function for the data is then

$$L = L_1 L_2 \cdots L_n,$$

where L_i is the likelihood for the i th group. There are three possibilities for L_i . If the i th group (of size G_i) is seen from position 1 only, then

$$L_i = \Theta_1(y_i, G_i)/P(G_i),$$

and division by $P(G)$ makes the likelihood conditional on the group being sighted. Similarly, if the i th group is seen from position 2 only, then

$$L_i = \Theta_2(y_i, G_i)/P(G_i),$$

and if the i th group is seen from both positions, then

$$L_i = \Theta_{12}(y_i, G_i)/P(G_i).$$

This means that the L_i chosen in the product L will depend on the source of the data. Here, L is now a complicated function of the regression parameters given in (8.8) and can be maximized numerically using approximations for the above integrals to obtain maximum likelihood estimates of the parameters. These can then be used to obtain population size estimates, which we now consider.

8.12.1 Population Size

We begin with N_L animals in the covered population being sighted with different probabilities, and the n animals sampled have sighting probabilities of π_1, \dots, π_n . An unbiased estimator of N_L is given by the Horvitz-Thompson estimator described by Thompson (2012, 49), namely,

$$\widehat{N}_L = \sum_{i=1}^n 1/\pi_i. \quad (8.10)$$

If the π_i can be estimated, we can get an approximately unbiased estimate of N_L . Manly et al. (1996) presented two methods.

The first method uses estimates of the α regression parameters in Eq. (8.8) as follows. With a slight change in notation, let P_{ij} be the sighting probability for group i and position j , and use (8.8) with observations, namely,

$$\text{logit}(P_{ij}) = \alpha_{0j} + \alpha_{1j}y_i + \alpha_{2j}G_i.$$

We then use the regression estimates in the above equation to get the estimates \widehat{P}_{ij} . Now the probability of being seen from at least one of the positions is $P_{i1} + P_{i2} - P_{i1}P_{i2}$ so that using (8.8), an estimator of N_L is

$$\widehat{N}_1 = \sum_{i=1}^n G_i / (\widehat{P}_{i1} + \widehat{P}_{i2} - \widehat{P}_{i1}\widehat{P}_{i2}).$$

An estimate of the number of groups of size G can also be obtained by summing $1/\widehat{P}_{i1} + \widehat{P}_{i2} - \widehat{P}_{i1}\widehat{P}_{i2}$ for the observed groups of this size.

The second method makes use of (8.9), which gives the probability of a group of size G being observed during the survey. If from (8.9) we have an estimate \widehat{P}_G , then we have the following estimator of N , namely,

$$\widehat{N}_2 = \sum_{i=1}^n G_i / \widehat{P}_{G_i}.$$

If n_G is the number of sighted groups of size G , then Gn_G/\widehat{P}_G is an estimate of the total number of animals of size G , and

$$\widehat{M}_G = n_g/\widehat{P}_G$$

is an estimate of the number of groups of this size. With covariates \mathbf{z}_i , we use $\widehat{P}_{G_i}(\mathbf{z}_i)$ in \widehat{N}_2 .

The theory for mark-recapture line transect sampling in the previous section applies to aerial censusing. Some history of this development and support for the idea of combining distance sampling with capture-recapture is given by Laake et al. (2008) and applied to a population of wild horses.

8.12.2 Some Examples

As aerial censusing is widely used, we consider a selection of examples using multiple observers. Ridgway (2010) used the method for double-crested free-ranging cormorants in coastal regions of Lake Huron. There were two observers seated on each side of the plane, and birds were observed through markers on wing struts, using distance bands, but with an unobservable area beneath the plane. Ten transect lines and their boundaries were mapped in a GIS and connected to an on-board geographic positioning system. Observations of cormorant detections in distance categories, estimates of cluster size for each detection, and whether birds were flying, on the lake surface, or loafing on shore (land) were recorded on portable tape recorders and later transcribed. The half-normal detection function was used, namely,

$$g(y) = \exp[-y^2/(2\sigma^2)],$$

with $\log(\sigma)$ expressed as a linear model with covariates including group size and the behavioral category of detected cormorants (on water, land, or flying).

Thomsen et al. (2006) applied the method to the harbor porpoise (*Phocoena phocoena*) as numbers had been declining. Multiple observers were used at different positions on the aeroplane, and observation distances y were calculated from $y = v \tan(90^\circ - \phi)$, where v is the altitude in meters and ϕ the angle of declination measured with a clinometer. For the detection distance function $g(y)$, they used a hazard-rate key function with a simple polynomial series expansion (cf. Sect. 7.2.4). Densities were calculated for each flight as $\widehat{D} = nGL/(2\mu_L g(0))$, where n is the number of sightings, G the average group size, L the total transect length, μ_L the effective strip half-width, and $g(0)$ the probability of detection at distance zero. An important assumption generally made is $g(0) = 1$. However, in cetacean surveys, this is not the case, because observers sometimes miss animals present (=perception bias) and diving individuals are unavailable for detection (=detection bias). To

estimate $g(0)$, the authors used a mark-recapture method combined with published diving data for harbor porpoises.

Gómez de Segura et al. (2006) used aerial censusing for the endangered loggerhead turtle in Spanish Mediterranean waters. The usual distance formula (cf. (8.4)) was used, namely,

$$\widehat{D}_L = \frac{n\widehat{f}}{2L\widehat{g}},$$

as, in this example, $g(0)$ was much lower than 1, because on the trackline, a portion of the turtles are diving and hence unavailable for detection and observers may fail to detect animals, although available, because of bad weather conditions. The only available data on juvenile loggerhead diving behavior was based on satellite tracking, so that $g(0)$ was estimated to be equal to the mean proportion of time that turtles spent at the surface (35.1%). Witt et al. (2009) discussed aerial surveying of the world's largest leatherback turtle rookery looking at additional ground surveys, tidal effects, and nesting effort. Aars et al. (2009) proposed a method for polar bears that corrected for areas not surveyed.

Aerial and shipboard surveys have been combined. Hammond et al. (2002) used this approach to carry out a very extensive program to estimate the abundance of harbor porpoise and other cetaceans in the North Sea and adjacent waters, allowing for $g(0) < 1$ on the transect line, and shipboard surveys allowed for animal movement in response to the survey platform. The porpoise, *P. phocoena*, and other small cetaceans, is a difficult species for such surveys because its small size and undemonstrative behavior at the surface make it hard to detect except in good conditions. Because of this and the aim of obtaining precise abundance estimates, the survey intensity was greater than is typical. Stratified blocks were used with zigzag paths for both types of surveying, along with multiple observers. Two aircraft were also used and school sizes were noted. Surveys have also been combined.

8.12.3 Some Model Variations

Sometimes, distance sampling and mark-recapture are not viable methods for some surveys. Butler et al. (2007) found this to be the case with wild turkeys in the Texas Rolling Plains, USA, where fixed-wing surveys were carried out. Their objectives were (1) to determine effects of flock size, distance from observer, and vegetative cover type on the detectability of wild turkey flocks and on errors in counting flock size, (2) to evaluate the accuracy and precision of wild turkey abundance estimates, and (3) to examine power to detect trends in population change with fixed-wing aerial surveys. They hypothesized that flock detectability and errors in counting flock size were functions of distance, vegetative cover type, and flock size. Simulation methods were also used.

Survey blocks were located in areas with radiotagged wild turkeys and representative cover. One observer was positioned in the right front seat, and the other was positioned in the back left seat. Streamers were placed on the struts, and there was a mark on the window to help observers categorize distances into distance groups. Also, a strip of about 100m wide directly below the observers was obstructed by the aircraft (50 m on either side of the transect). Prepositioned decoy flocks were used with GPS locations along with radiotagged wild turkeys to help evaluate the applicability of using turkey decoys to develop detectability models. Logistic regression was used with flock detectability. The total population size N was estimated by a modified Horvitz-Thompson estimator

$$\widehat{N} = \sum_{i=1}^m \frac{(y_i / h_i)}{g_i},$$

where m is the total number of flocks detected, y_i is the number of turkeys counted in flock i (flock size), h_i is the percentage error of the count of individuals in flock i (calculated from the model-averaged predictions of the linear regression models that did not include flock size), and g_i is the probability of detection for flock i (from the model-averaged predictions of the logistic regression models that did not include flock size).

The findings were as follows: (1) Flock detectability was primarily influenced by flock size and vegetative cover, and errors in counting flock size were primarily influenced by size of flock; (2) simulations suggested that the fixed-wing aerial survey technique may underestimate abundance by 10–15%, and such bias occurred because the model-averaged estimates of the number of individuals in detected flocks were increasingly underestimated as flock size increased; and (3) the continued use of decoys was recommended to further refine and verify detectability models, evaluate other potential factors affecting detectability, and evaluate the applicability of the technique in other ecoregions.

Alisauskas and Conn (2019) studied how a distance covariate influenced the probability of double-observer detections for birds counted during a helicopter survey in Canada's central Arctic using a fixed width transect. Two observers, one behind the other but visually obscured from each other, counted birds in an incompletely shared field of view to a distance 200 m. Guided by semi-transparent marks on an aircraft window, each observer assigned detections to one of the five 40 m distance bins, and information was recorded on taxonomic group, wing-flapping behavior, and group size.

Two estimation procedures were used: double-observer mark-recapture method (MR) without distance information and double-observer mark-recapture with a categorical distance covariate (MRD). The authors were concerned that heterogeneity in the detection probability could affect MR models due to unmodeled distance information, as more distant birds are less likely to be detected by both observers.

With regard to the mathematical model, let y_{io} be a binary indicator for whether or not observer o detects waterfowl group i , and let d_{io} denote the distance bin

recorded by observer o to group i ($i = 1, \dots, n$). Here, d_{io} is only defined when $y_{io} = 1$. Detection histories were included in the analyses if (a) $y_{i1} = 1$ and $d_{i1} \in \{1, 2, 3, 4, 5\}$ (i.e., observer 1 detects the animal in distance bins 1–5) or (b) $y_{i1} = 0$, $y_{i2} = 1$, and $d_{i2} \in \{1, 2, 3, 4, 5\}$ (i.e., observer 1 misses the animal but observer 2 detects it in bins 1–5). Let $p_{io} = 1 - q_{io}$ be the probability of detecting group i by observer o . The distance used for a waterfowl group i was set to $d_i = y_{i1}d_{i1} + (1 - y_{i1})d_{i2}$ (i.e., giving preference to the first observer's distance determination). Then, if $p_i^* = 1 - q_{i1}q_{i2}$, we can use the capture-recapture model conditional on an individual to be observed by at least one observer with cell probabilities for group i such that

$$y_{io} \sim \text{Bernoulli}(p_{oi}/p_i^*) \quad \text{where} \quad \text{logit}(p_{io}) = \mathbf{x}_i' \boldsymbol{\beta}.$$

Here, the \mathbf{x}_i are covariate values, and the model can be used to estimate $\boldsymbol{\beta}$. Estimates of the number of waterfowl groups (G_s) and of the total abundance of species s (N_s) are given, respectively, by

$$\widehat{G}_s = \sum_{i \in \Omega_s} 1/\widehat{p}_i^* \quad \text{and} \quad \widehat{N}_s = \sum_{i \in \Omega_s} g_i/\widehat{p}_i^*,$$

where Ω_s is the set of waterfowl groups detected by at least one observer that are assigned to species s and g_i is the number of birds in the i th waterfowl group. However, the main emphasis of the study was to identify the factors that affected detectability. Observations were limited to species categories which had 20 or more detection histories. Alisauskas and Conn (2019) concluded that MRD models were preferred to MR models for Arctic birds because of smaller standard errors and were therefore more able to determine population size trends. Also, distance procedures are insensitive to moderate heterogeneity in the detection probability owing to a “pooling robustness” property discussed in Sect. 8.2.4.

Williams et al. (2017) considered adapting aerial imaging when there is imperfect detection. One way of mimicking what happens on the ground with replicate surveys is to take multiple images and use overlaps to determine detectability. If $N(A, t)$ is the count of the true number of animals in area A at time t , then the animals can be regarded as points forming a point process with intensity $\lambda(\mathbf{s}, t)$, which varies over space and time, where $\mathbf{s} = (s_1, s_2)'$ describes the location of a point (e.g., latitude and longitude). If $N(A, t)$ is an inhomogeneous Poisson point process, with distribution Poisson ($\lambda(A, t)$), then

$$\phi(A, t) = \Pr(N(A, t) > 0) = 1 - \exp[-\lambda(A, t)].$$

A set of n aerial images are taken at locations \mathbf{s}_i ($i = 1, \dots, n$) capturing $y(A_i, t)$ counts on bounded subareas A_i taken at time t . It is assumed that the aerial images have sufficient resolution so that the observer detection probability is 1. Let $p(A_i, t)$

be the probability that an animal is available to be counted on an image. Then

$$y(A_i, t) \sim \text{Binomial}(N(A_i, t), p(A_i, t)),$$

which reduces to

$$I_{y(A_i, t) > 0} \sim \text{Bernoulli}(\phi(A_i, t)),$$

where $I(\cdot)$ is the usual indicator, when occupancy is a state variable of interest. The authors expressed $\lambda(s, t)$ in terms of a log-linear model of covariates. Temporal replication is available when two subareas overlap. Further details are given by Williams et al. (2017) who applied the method to sea otters.

Laake et al. (2008) noted that altitude and flight speed affect detection probability and they typically vary during the course of most aerial surveys. An increase in altitude will make objects appear smaller but will increase the time available for viewing a given area. An increase in flight speed will decrease the time available to observe a given area and lead to an increase in visibility bias.

One approach is to include altitude and speed and other factors as covariates in the scale parameter of the detection function. However, variation in survey altitude causes additional difficulties when the aircraft has flat windows creating a “blind” strip beneath the aircraft not visible to the observer, and its width changes as the altitude varies. Similar problems occur when survey altitude varies and distances are collected in bins (intervals), as is the case with strut markers. The authors developed an unbiased estimator based on vertical angles α rather than distance, where $\alpha = \tan^{-1}(x/H)$, x is the horizontal distance, and H is the survey altitude. The estimator for population density then becomes

$$\widehat{D} = n \widehat{f}_\alpha(0)/(2LH) = n/(\widehat{P}2LW),$$

where $\widehat{f}_\alpha(0)$ is the probability density function for angles evaluated at 0° .

For groups of size s_i for the i th group at surveying height H_i , the density estimator is

$$\widehat{D} = \sum_{i=1}^n s_i \widehat{f}_{i\alpha}(0)/(2LH_i),$$

where

$$\widehat{f}_{i\alpha}(0) = \left[\int_{H_i \tan \alpha_0}^{H_i \tan \alpha_k} g(x - H_i \tan \alpha_0) dx \right]^{-1}$$

Here, α_0 is the angle out to the edge of the blind strip, and α_k is the angle out to the outermost bin when there are k bins. Mathematical details are given in their appendix.

8.12.4 Supplementary Ground Counts

In Sect. 3.2.3, we considered the method of double sampling where a rapid survey is carried out and then an accurate subsample is made to estimate the detection probability. Sometimes, we can do the same thing, but using two very different kinds of sample. For example, one can carry out an aerial survey and then do more accurate ground counts on segments of the aerial survey. For example, Smith (1995) in a large survey of waterfowl in North America said that a subsample of segments is surveyed annually by ground crews and these ground counts are used to compute visibility correction factors (VCFs) that adjust aerial counts for incomplete detection, employing a combined ratio estimator within a double sampling estimation framework.

There is also another method for the second subsample. Rugged, remote terrain in the ESA (Eastern Survey Area) precludes the use of ground counts for the estimation of detection rates. Instead, helicopters can be used to conduct slow, hovering counts on a subsample of survey segments designed to flush birds or otherwise cause bird movement and aid in detection (Koneff et al., 2008). These helicopter counts were used in lieu of ground counts, again in a double sampling framework. Koneff et al. (2008) also used a double-observer method. Lubow and Ransom (2009) used natural markings of feral horses (along with photography) to provide a naturally marked population of known size to investigate heterogeneous sightability, which was found to be substantial. Camera methods are described in more detail in Sect. 14.6.

Mizel et al. (2018) developed a time-removal model for the long-term monitoring of cliff-nesting raptors (e.g., golden eagle territories in the Denali National Park and Preserve) combining aerial and ground counts. Using a helicopter, an observer surveyed nests within all nesting territories in late April or early May (i.e., after most clutches were assumed to have been completed) and recorded the presence or absence of an incubating eagle, eggs, or eggshells. The ground survey included visually searching the area for eagles engaged in territorial behavior, such as nest building, territorial defense, or courtship, and generally lasted more than 4 hours or until they observed a behavior indicative of occupancy. They conducted a subset of surveys entirely on foot, which occasionally included multiple observers. Only the first aerial visit to nesting territories was used, and they censored (post hoc) all ground surveys at 2 hours to specify a time-removal model, i.e., once seen effectively “removed.”

A state-space framework was used describing the probability of each nesting territory existing in one of the three mutually exclusive states: unoccupied, occupied without breeding, and occupied with breeding. In their approach, they assumed that the territorial population is closed, but condition on the territory’s state at the time of the survey. The parameters of the state detection matrix were the probability of detecting occupancy without breeding during the aerial survey, the probability of detecting occupancy without breeding (nonbreeding) during the ground survey for some time interval, and a binary indicator of whether the nest in which eggs were laid was surveyed. Further details are given in their paper. The key difference is

having two completely different methods (aerial and ground) for the repeat surveys and using a removal approach.

8.12.5 Digital Observations Using Cameras

As computers are replacing people, it is not surprising cameras can be used instead of observers in aerial censusing using, for example, drones and two digital cameras (Schroeder et al., 2020). These can perhaps be mounted on one aircraft, possibly unmanned, which covers the same area with a short time delay between them. Animal movement between the passage of the cameras introduces uncertainty in individual identity, so individual capture histories are unobservable and are treated as latent variables.

One of the problems with camera pairs is that individual identity is uncertain, so compiling detections into individual capture histories is prone to error. Stevenson et al. (2019) used a cluster approach to deal with this problem, generalizing the methods of Tanaka et al. (2008) and Fewster et al. (2016) using two observers/cameras separated by a time lag to account for the unavailability of animals, for example, due to diving. The method has wider applications as it involves parent/offspring and sibling relationships, here porpoise pods, and uses the parameters of Neyman-Scott point processes, which generate clustered point patterns, as well as a Palm likelihood.

Using the approach of Schroeder et al. (2020), Borchers et al. (2022) obtained the likelihood for mark-recapture line transects without capture histories by automatically enumerating all possibilities within segments of the transect that contain ambiguous identities, instead of attempting to decide identities in a prior step. They called this method “latent capture-history enumeration” or LCE. Two coordinates were used for location: the forward coordinate along the transect line and the transverse coordinate perpendicular to the line. An “observer,” which can also refer to a camera, “passes over” an animal at the instant that their forward coordinates coincide, regardless of the animal’s transverse coordinate at that instant. It is assumed that the observer speed exceeds animal speed, so the time at which each observer passes over each animal is well defined.

To arrange for a time delay between the two cameras, there can be one forward-pointing, and the other rear-pointing. These can be engineered so that the rear-pointing camera records the same area as the forward-pointing camera after a time delay of several seconds. This separation generates data with which one can model the availability cycle, as long as there is a chance that the availability status of an animal changes between the passage of the two cameras. Any animal movement between the two observers is modeled as a Brownian motion such that the animal’s displacement over time t follows a bivariate normal distribution with mean $(0, 0)$ and covariance matrix $\Sigma(t) = \sigma^2 t \mathbf{I}_2$, where \mathbf{I}_2 is the 2×2 identity matrix.

Borchers et al. (2022) modeled the availability of an animal as a two-state Markov process with transition probabilities over an interval of time t that induces

dependence between the two cameras. As animals may move into or out of the detection zone between the passage of the two observers, the survey area is considered to constitute a wider strip of width $b > W$ on either side of the line, where the buffer b is chosen such that there is negligibly small probability that animals beyond b at the passage of the first observer will be within the searched strip w at the passage of the second. The buffered strip of width $2b$ therefore covers all animals that may be exposed to detection.

Let y_0 is the signed distance of an animal to the right of the transect line when the first observer passes overhead and y_t be this distance time t later. It is assumed that y_0 has a uniform distribution on $(-b, b)$ independently for all animals which implies that $y_t - y_0$ is $N(0, \sigma^2 t)$. We have just sketched some of the details that make the method different from previous approaches, and the reader is referred to Borchers et al. (2022) for mathematical details.

Digital cameras can be used in a number of different ways. Conn et al. (2014) used a double sampling aerial procedure consisting of infrared imagery to detect animals and independent, high-resolution digital photography to provide information on species composition and thermal detection accuracy. The combination was applied to a survey of four ice-associated seal species in the eastern Bering Sea. The authors suggested a need to balance model complexity with the richness of the data set. For example, highly parameterized models can lead to spuriously high predictions of abundance in areas that are not sampled, especially when there are large gaps in spatial coverage. They recommended “that ecologists employ double sampling when enumerating animal populations with automated detection systems to estimate and correct for detection errors. Combining multiple data sets within a hierarchical modeling framework provides a powerful approach for analyzing animal abundance over large spatial domains.” As the model is complex, with a large number of parameters and random variables, and this book is already too large, we shall not give details.

8.13 Extension of Model-Based Methods

8.13.1 Shipboard Acoustic Surveys

Acoustic methods have been used for a long time in estimating densities of marine life, and their use has increased dramatically. For example, Brierley et al. (2004) used single-target echo detections even for jellyfish. Marques et al. (2009) used fixed passive acoustic sensors, as visual line transect methods have a number of disadvantages for surveying cetaceans: they can only be performed during daylight hours and are strongly dependent on good weather conditions; they do not work well for species that spend long periods of time underwater; and they are expensive to do well and have restricted temporal coverage. On the other hand, some cetacean species make frequent and characteristic vocalizations, which has led

to an increasing use of passive acoustic methods. Using acoustics, the proportion of animals not detected by a visual observer team was estimated by Rankin et al. (2020). They distinguished between full and point independence.

Acoustic line transect methods still require an expensive survey platform and have restricted temporal coverage. Also, boat movement can affect animal movement, and Palka and Hammond (2001) reviewed some methods for correcting for movement. They developed a method that uses animal orientation data to determine whether, and at what distance, cetaceans are responding to the survey ship.

Marques et al. (2009) noted that fixed acoustic sensors (either anchored or buoyed) can be deployed to record data over long periods or transmit it to shore via cables. Their methods involved cue counting, and population density can be estimated by

$$\hat{D} = \frac{n_c(1 - \hat{c})}{K\pi W^2 \hat{P} T \hat{r}},$$

where W is the distance away from the hydrophones beyond which cues are assumed to not be detected, \hat{P} is the estimated average probability of detecting a cue made within distance W , \hat{r} is the estimated cue production rate, \hat{c} is the estimated proportion of false-positive detections, and K is the number of replicate sensors used. Here, $n_c(1 - \hat{c})$ corresponds to the number of detected cues that were actually from the target species, $T\hat{r}$ corresponds to the number of cues produced by an average animal during the recording time T , and $a = K\pi W^2$ is the covered area sampled. If $\hat{n} = n_c(1 - \hat{c})/(T\hat{r})$, an estimate of the number whales detected, then \hat{D} takes the simple intuitive form

$$\hat{D} = \frac{\hat{n}}{a\hat{P}}.$$

If CV denotes coefficient of variation, then from the delta method, and assuming independence of the various random components,

$$\text{var}[\hat{D}] \approx D^2 \left\{ CV(n_c)^2 + CV(\hat{r})^2 + CV(\hat{c})^2 + CV(\hat{P})^2 \right\}.$$

The variance in n_c can be estimated from the empirical variance of the cue counts n_{cj} ($j = 1, \dots, K$) over K hydrophones, namely,

$$\widehat{\text{var}}[n_c] = K \sum_{j=1}^K \frac{[n_{cj} - (n_c/K)]^2}{K - 1}.$$

Further details are given by the authors. Marques et al. (2011) used the same approach for North Pacific right whales, but with the difference of having a minimum left-truncation distance W_1 , say, because distance cannot be reliably

measured to close-by detections. This means that the effective area is now $\pi(W^2 - W_1^2)$ instead of πW^2 . For an earlier overview of passive acoustic methods, see Mellinger et al. (2007).

Barlow and Taylor (2005) used a combined acoustic and visual survey to study sperm whales. Of all cetaceans, sperm whales (*Physeter macrocephalus*) are the most amenable to acoustic and survey methods, as they are long, deep divers, and hence hard to detect visually, but produce loud echolocation clicks. The whales were detected acoustically using a hydrophone array towed at 15 km/h and 100 m depth. The hydrophone array was towed for 14,500 km, and locations were estimated acoustically for 45 distinct sperm whale groups. Acoustic techniques substantially increased the number of sperm whales detected on this line transect survey by increasing the range of detection and allowing nighttime surveys; however, visual observations were necessary for estimating group size. Two groups of observers were used. Population size N for area A was estimated as

$$\widehat{N} = \frac{An\widehat{E}(s)\widehat{f}(0)}{2L\widehat{g}(0)},$$

where n is the number of visual or acoustic detections, $E[s]$ is the expected group size estimated by its average, $f(0)$ is the sighting probability density at zero perpendicular distance, $g(0)$ is the probability of detecting a sperm whale on the transect line, and L is the length of the transect. Further details of estimation are given by the authors.

As noted by Hedley and Buckland (2004) and mentioned above, there has been a move to use a model-based approach to estimate density. This is in contrast to relying on unbiased estimation from a design-based approach using randomized transect locations. It is important to have adequate spatial coverage of the survey area with sufficient spatial spread to provide representative coverage of the survey area. This can lead to a more systematic rather than a random approach to setting up line transects. We see this particularly in maritime surveys if the spatial coverage of the nonrandom data is adequate. This leads to abundance estimation from sighting data collected from so-called platforms of opportunity (e.g., ferries, merchant navy vessels, oceanographic survey vessels, etc.).

Hedley and Buckland (2004) commented that it is often prohibitively expensive to hire a ship, so that many marine surveys are conducted from these types of platforms. This leads to collecting large quantities of nonrandom data at a cost substantially lower than that of a properly designed survey. However, this approach has two disadvantages: model misspecification can lead to estimation bias, and more sophisticated analysis methods are needed than for a design-based analysis. Their model is discussed below in Sect. 8.13.5 on spatial models.

Shipboard and aerial surveys have been combined. Hammond et al. (2002) used this approach to carry out a very extensive program to estimate the abundance of harbor porpoise and other cetaceans in the North Sea and adjacent waters, allowing for $g(0) < 1$ on the transect line, and shipboard surveys allowed for animal

movement in response to the survey platform. The porpoise, *P. phocoena*, as well as other small cetaceans, is a difficult species for such surveys because its small size and undemonstrative behavior at the surface make it hard to detect except in good conditions. Because of this and the aim of obtaining precise abundance estimates, the survey intensity was greater than is typical. Stratified blocks were used with zigzag paths for both types of surveying, along with multiple observers. Two aircraft were also used and school sizes were noted.

8.13.2 Other Marine Surveys

The two most popular methods for estimating cetacean abundance are line transect surveys with related distance sampling methods and mark-recapture methods using photo-identification data. As already noted, shipboard surveys can be expensive, but small-boat surveys can be cost-effective (Williams & Thomas, 2009). However, they do have some obvious limitations such as needing to be close to the shore, having small teams of observers, and having a narrow field of view.

Katsanevakis (2007) described for the first time a method for estimating marine benthic fauna in a lake using a density surface modeling approach combining distance sampling with a geographic information system (GIS). It was used to estimate the population density of the endangered fan mussel *Pinna nobilis*. Eighteen line transects, perpendicular to the depth contours and extending from the shore (zero depth) to a depth of 30 m, were defined randomly in the lake. Each line transect was defined with a nylon line that was deployed using a diving reel while scuba diving. The line was marked with a water-resistant paint marker every meter and divided into 4-m segments, with water-resistant numbered signs. After deploying the line, all *P. nobilis* individuals within 2 m (=W) from the line were counted.

For each observed individual, the following variables were recorded: the corresponding segment in which it was found, the perpendicular distance from the line, and the shell width. The location of an individual was defined as the midpoint of the section of the shell by the seabed plane. In addition, at the midpoint of each segment of the line transect, the exact depth was measured using the electronic depth meter of a diving computer. Two detection functions $g(y)$ were considered, the half-normal ($\exp[-y_2/(2\sigma^2)]$) and the two-parameter hazard-rate function ($1 - \exp[-y/\sigma]^{-b}$). In two of the models, shell size was used as a covariate because of its possible effect on detectability. Two spatial covariates were used in modeling the density surface, the depth d , and the variable x representing the distance from the channel connecting the lake with the open sea, when moving clockwise along the shoreline. A Horvitz-Thompson-like estimator was used to estimate the number in a segment. Further details are given by the authors.

Shipboard surveys have been used to estimate sea bird densities from birds on the water, as described, for example, by Ronconi and Burger (2009). Their density estimator was

$$\widehat{D} = \frac{n\widehat{E}[s]}{2WL\widehat{P}} \frac{1}{\widehat{g}(0)},$$

where n is the number seen; s is the group size; P is the average probability of detection, which can be estimated from **Distance** software; L is the transect length; and $g(0)$ is the probability of detection at distance zero. Here, $E[s]$ can be estimated in a number of ways.

Irigoyen et al. (2018) discussed underwater visual censuses (UVC) as they are a nondestructive method and are commonly used in shallow and clear waters worldwide, e.g., studying reef fish. There can be some problems such as the physiological constraints of scuba diving, individual response of fishes (either attraction or escape), and possible estimation bias with cryptic, elusive, hidden, or/and small-sized fishes.

Choosing appropriate transects is not always straightforward, and the authors proposed a method called the tracked roaming transect, an UVC method designed to maximize transect length (and thus the surveyed area) with respect to diving time invested in monitoring. It was specifically designed to survey commercial and charismatic, medium- and large-size fish species commonly found at low densities (such as groupers or sharks) and/or threatened species. The method was based on a GPS density survey designed to asses reef fish spawning aggregations and other experiences of diver-towed GPS techniques. We will not describe the method as there are detailed technicalities.

An extensive and general review of marine spatial management was given by Katsanevakis et al. (2011).

8.13.3 Allowing for Movement

Animal movement can be two types, due to either the observer's presence or the movement on the animal's accord independent of the observer, though most of the focus has tended to be on the first type. Glennie et al. (2015) noted that movement independent of the observer caused bias in the estimated probability of detection and this bias combined with the bias arising from the movement of animals into the encounter region. They used simulation to study the effects.

Methods that allow for an observer effect that include specific survey techniques such as searching further along line transects to see animals before they respond or remaining at point transects long enough for animals to resume normal behavior are recommended to mitigate bias in density estimates (Buckland et al., 2004). Double-observer methods have also been used. An extension of a previous model by Alisauskas and Conn (2019) to allow for movement and measurement error was

given by Conn and Alisauskas (2018). Glennie et al. (2021) considered just the second type of movement by combining distance sampling survey data with animal telemetry data.

8.13.4 Full Model-Based Method

Conventional distance sampling is a mix of design-based and model-based methods, where models are proposed for the detection function. Buckland et al. (2016) considered using a general distribution for the distance y of an animal (irrespective of whether it is seen or not) on the grounds that full model-based methods offer greater flexibility. Also, a conditional likelihood is used, conditional on the number n of animals detected. The full model approach replaces the design-based approach that uses a uniform distribution. We recall from Eq. (8.1) that P_L , the probability that an animal is detected, given that it is in the area $a = L \times 2W$ of possible detection, is

$$P_L = f_X(1) = \int_0^W f_{X|Y}(1 | y, \boldsymbol{\theta}) f_Y(y) dy = \int_0^W g(y | \boldsymbol{\theta}) f_Y(y) dy, \quad (8.11)$$

where $f_Y(y)$ (the density function of the distance y irrespective of whether an animal is detected or not) is now assumed to be some distribution rather than being determined by the random placement of the transect, a design feature leading to the uniform distribution. We condition on the unknown parameter vector $\boldsymbol{\theta}$ as we consider a Bayesian approach in the next section; P_L is a function of $\boldsymbol{\theta}$. The density function of y given that the animal is detected is now, by Bayes' theorem,

$$f_c(y | \boldsymbol{\theta}) = f_{Y|X}(y | 1, \boldsymbol{\theta}) = \frac{f_{X|Y}(1 | y, \boldsymbol{\theta}) f_Y(y)}{f_X(1)} = \frac{g(y | \boldsymbol{\theta}) f_Y(y)}{P_L}. \quad (8.12)$$

A natural model for n is the binomial distribution

$$f(n) = \binom{N_L}{n} (\psi P_L)^n (1 - \psi P_L)^{N_L - n},$$

where N_L is the population size in the study (covered) region and ψ is the presence probability that an animal within the covered region is on one of the surveyed plots.

The full likelihood for the data is now

$$L(n, \{y_i\} | \boldsymbol{\theta}) = f(n) L(\{y_i\} | \boldsymbol{\theta}, n) = f(n) \prod_{i=1}^n [g(y_i | \boldsymbol{\theta}) f_Y(y_i) / P_L].$$

If we use grouped distance data, where the y_i are defined by cutpoints $c_0 = 0, c_1, \dots, c_J = W$, then we simply replace $L(\{y_i\} | n, \theta)$ by $L(\{m_j\} | n, \theta)$ given by

$$L(\{m_j\} | n, \theta) = \frac{n!}{\prod_{j=1}^J m_j!} \prod_{j=1}^J f_j(\theta)^{m_j},$$

where m_j is the number of detections in distance interval j , with $\sum_{j=1}^J m_j = n$, and

$$f_j(\theta) = \int_{c_{j-1}}^{c_j} g(y | \theta) f_Y(y) dy / P_L \quad (8.13)$$

is the probability of being detected in the j th interval. The full likelihood is now

$$L(n, \{m_j\} | \theta) = f(n)L(\{m_j\} | n, \theta).$$

Maximum likelihood or Bayesian methods of estimation can now be used.

If we use a Poisson model for n , we can have the spatial distance sampling model using a nonhomogeneous Poisson process model given by Hedley and Buckland (2004) and considered in the next section. Dorazio (2012) discussed a similar model that made use of presence-only data discussed in Sect. 4.12. Buckland et al. (2016) showed how the general model can be used to generate a number of models using distance data combined with covariate models, capture-recapture models, and plot-based models using plot counts (where the term “plot” refers to either a circular area from point counts sampling or a strip from line transect sampling). Using random effects was also considered. The key question is the specification of the unknown $f_Y(y)$.

8.13.5 Spatial Models

Although spatial models are considered in Chap. 14, we consider them here in the context of distance sampling. Hedley and Buckland (2004) considered a spatial approach to distance sampling using minke whales as an example. Let $D(u, v)$ represent the animal density at location (u, v) , and let $g(x, y)$ be the probability of detecting an animal at coordinates (x, y) . Here, x is the distance along the transect recorded in intervals $d_j \leq x \leq d_j + b_j$, and y is the usual perpendicular distance from the line ($-W \leq y \leq W$), though above we have that the distribution is folded, so that $0 \leq y \leq W$. The D and g operate very differently; D runs along the line, but typically varies very little within a strip in the y -direction. However, g varies largely in the y -direction rather than the x -direction.

Let

$$\mu(A) = \int_A D(x, y)g(x, y)dxdy,$$

where the integration is over the whole population of area A . The joint density function of the number of detections n and their locations is

$$f(\{(x_i, y_i)\}, n) = \exp[-\mu(A)] \prod_{i=1}^n D(x_i, y_i)g(x_i, y_i)/n!, \quad n = 1, 2, \dots,$$

and equals $\exp[-(\mu(A))]$ when $n = 0$. In their first method, they assume $D(x, y) \equiv D(x)$ independent of y , $g(x, y) \equiv g(y)$ independent of x , $g(0) = 1$, $g(-y) = g(y)$ ($0 \leq y \leq W$), and x and y are independent. They then used the interdetection distances $l_i = x_i - x_{i-1}$ along the line with the value at the start of survey effort as $x_0 = 0$ and the x -value corresponding to the end of survey effort as x_{n+1} ; then l_1 is the distance surveyed until the first detection, and l_{n+1} is the distance surveyed after the last detection. They obtained a marginal likelihood for θ (the spatial parameters for the density surface $D(x)$) and β (the parameters for the detection function $g(y)$), based on the y_i and l_i data. Variation in detectability with x can be modeled by introducing a vector of covariates \mathbf{z} , so that the detection function is $g(y, \mathbf{z})$.

The authors suggested two models when direct maximization is not feasible: using generalized linear modeling (GLM) or generalized additive modeling (GAM) software (e.g., Miller et al., 2013). The first applied a link function to the $E[l_i]$ and used covariates. The second divided up a transect into T small contiguous sampling units or “segments” each of (approximately) equal length, the length of each segment being such that the geographic location did not change appreciably within a segment. The length of the i th segment was denoted by l_i and the number of animals detected within it by n_i ($i = 1, \dots, T$). Then $\log E[n_i]$ is modeled in terms of the l_i , the estimated probability of detection, and covariates. This second method, based on just counts, is appropriate for strip transect counts of, for example, seals hauled out on ice. Here, when animals are numerous and close together, it is not practical to attempt to estimate distances along the line between successive animals. Instead, photographs can be taken, and the number of animals in each photograph, after deleting any overlap, is counted later. Variances are estimated using bootstrap methods.

Yuan et al. (2017) used a spatiotemporal distance model for blue whales, an endangered species, using a shipboard survey. As we know, the probability of detecting an animal typically decreases with distance from the observer. From a modeling perspective, this results in a so-called “thinned” spatial point process with the intensity function scaled by the detection probability. They used a log-linear detection function model based on a semiparametric piecewise quadratic and considered a stochastic partial differential equation approach to model spatial structure in density that is not accounted for by explanatory variables. It allowed the simultaneous fitting of detection and density models and permitted the prediction

of density at an arbitrarily fine scale. The focus of the study was on how density changes continuously in space with respect to available explanatory variables, and across years (22 years), using Bayesian inference.

Johnson et al. (2010) also used a spatial model-based distance method described as follows. The locations $\mathbf{s} = (s_x, s_y)$ of all individuals in an area A , say \mathcal{S}^+ , are a realization of a point process with intensity function $\lambda(\mathbf{s}, \boldsymbol{\beta}) = \exp[\mathbf{x}(\mathbf{s})' \boldsymbol{\beta}]$, where $\mathbf{x}(\mathbf{s})$ is a vector of environmental covariates with $x_1(\mathbf{s}) \equiv 1$. We now need to allow for the fact that A is not surveyed in its entirety and individuals are not detected with certainty. This can be accomplished by viewing the line transect sampling procedure as thinning the original location process \mathcal{S}^+ to obtain the locations of observed individuals $\mathcal{S} = (\mathbf{s}_1, \dots, \mathbf{s}_n)$.

For a given realization of locations \mathcal{S}^+ , one thins the process by retaining each point with probability $q(\mathbf{s}_i; \boldsymbol{\eta})$, say, and discards the rest to obtain the subset \mathcal{S} . The resulting intensity function of the thinned point process \mathcal{S} is $q(\mathbf{s}; \boldsymbol{\eta})\lambda(\mathbf{s}; \boldsymbol{\beta})$. To determine $q(\cdot)$, it is assumed without loss of generality that A is surveyed in disjoint regions $C_k \subset A$ ($k = 1, \dots, K$) with straight-line transect corridors of width $2W_k$. The detection function is assumed to take the form

$$g(\mathbf{s}; \alpha_k, \gamma) = \exp[-\{y_k(\mathbf{s})/\alpha_k\}^{1/\gamma}],$$

where, for $\mathbf{s} \in C_k$, $y_k(\mathbf{s})$ is the usual perpendicular distance from a location \mathbf{s} to the transect centerline; otherwise, $g(\mathbf{s}; \alpha_k, \gamma) \equiv 0$. The function is defined for every location in A , leading to the thinning function,

$$q(\mathbf{s}; \boldsymbol{\eta}) = \sum_{k=1}^K g(\mathbf{s}; \alpha_k, \gamma),$$

where $\boldsymbol{\eta} = (\alpha_1, \dots, \alpha_K, \gamma)'$. Here, $\gamma = 0.5$ for a Gaussian distribution, $\gamma = 1$ for an exponential distribution, and $\gamma \rightarrow 0$ for a uniform distribution. We can now define a full likelihood for distance sampling by multiplying $\lambda(\mathbf{s}, \boldsymbol{\beta})$ by $q(\mathbf{s}, \boldsymbol{\eta})$ to obtain the likelihood for the observed data \mathcal{S} , namely,

$$\begin{aligned} L(\boldsymbol{\theta}; \mathcal{S}) &= \sum_{k=1}^K \sum_{j=1}^{n_k} [\mathbf{x}(\mathbf{s}_j)' \boldsymbol{\beta} - \{y_k(\mathbf{s}_j)/\alpha_k\}^{1/\gamma}] \\ &\quad - \sum_{k=1}^K \int_{C_k} \exp[\mathbf{x}(\mathbf{u})' \boldsymbol{\beta} - \{y_k(\mathbf{u})/\alpha_k\}^{1/\gamma}] d\mathbf{u}, \end{aligned}$$

where n_k is the number of animals detected in C_k and $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\eta}')'$. The above model generalizes the likelihood given by Equation (2.9) in Hedley and Buckland (2004). Details of maximum likelihood estimation, abundance information, and overdispersion are given by Johnson et al. (2010).

8.13.6 Adaptive Sampling

Adaptive sampling was introduced briefly in Sect. 2.8. The key idea is that sampling is increased in the neighborhood of a sampling unit if some criterion is triggered. Pollard and Buckland (1997) and Pollard et al. (2002) used a shipboard adaptive method in which survey effort is increased when areas of high animal density are located. In particular, they advocated an approach where the sampling effort is a function of the degree in which the vessel is ahead or behind schedule. The effort is doubled when the trigger is activated, giving an effort factor of two. As the method conditions on the effort, it is not design unbiased. The method was applied to harbor porpoises (*Phocoena phocoena*).

Normally, increasing the effort in areas of higher animal density would lead to abundance overestimation. To compensate for this, the adaptive method weights the data inversely proportional to the effort factor so that each section of the transect is weighted in proportion to the length of straight-line (nominal) effort throughout that section. Many tracking designs for increasing the effort are possible such as a zigzag pattern, which has the advantage of not crossing itself and having no gaps in the trackline so that no search effort is lost in raveling from one transect to the next. Also, the track is easily followed, which is important for shipboard surveys. In addition, the increase in effort is directly related to the length and angle of the zigzags and thus can be fixed at any value greater than 1 by changing either or both of these factors.

To set up the model, each transect is divided into a number of subtransects or “legs” where one leg is deemed to start whenever there is a change in effort. Within a leg, there can be a straight line or some other shape such as a zigzag. Let L be the total line length with L' its “nominal length,” that is, the straight line length rather than, say, the zigzag length. Similarly, ℓ is the length of a leg and ℓ' its nominal length. Working with groups (e.g., schools), let n be the number of groups detected in a leg, so that the encounter rate per unit length for the leg is $e = n/\ell$. The group size is s , the animal density is D , and $f_c(0)$ is the usual density function for the perpendicular distance to the school evaluated at zero distance. Let i refer to the transect ($i = 1, \dots, k$) and j refer to the leg ($j = 1, \dots, m_i$). The standard transect assumptions are:

- (1) The probability of detection on the line, $g(0)$, is one.
- (2) The probability of group detection is independent of the group size.
- (3) There is no responsive movement of animals in advance of detection, and any nonresponsive movement is slow relative to the speed of the observers.

We also add some specific assumptions for the above method, namely:

- (4) The expected encounter rate for an adaptive track is the same as the expected encounter rate for the corresponding nominal track.
- (5) The expected group size for an observation on an adaptive track is the same as the expected group size for an observation when following the corresponding nominal track.

- (6) Conditional on the location of the actual (as distinct from the nominal) track, each observation is an independent event, i.e., the probability of an observation is only a function of its perpendicular distance from the actual line (although the position of the line itself may depend on past observations).

The effort factor λ is defined to be the ratio of the actual effort to the nominal effort, so that

$$\lambda_{ij} = \ell_{ij}/\ell'_{ij}$$

is the effort factor for the j th leg of the i th transect. Suppose additional effort is triggered so that we need to calculate the effort factor as a function of the remaining effort available. Let $L_E(t)$ (measured in units of time or distance) be the total excess effort remaining at time t . It is the total effort available at the start of the survey less the actual effort used to time t less the nominal effort to complete the survey (without any further adaptive effort). This description generally sets up the model basics, and the reader is referred to Pollard et al. (2002) for the comprehensive mathematical details. The key thing here is the innovative application of adaptive sampling. Further information on adaptive sampling combined with distance sampling is given by Moradi (2019) who extended the method to airborne geophysics studies.

8.14 Bayesian Methods

Although a few Bayesian methods have been referred to above, they have generally been slow to enter the distance methodology. Oedekoven et al. (2014) developed a model that included both point and line intercept methods. We continue with the notation used above. If y_i refers to the i th measurement and there are n detections, then the likelihood function for the y_i was expressed as

$$L_y(\boldsymbol{\theta}) = \prod_{i=1}^n f_c(y_i \mid \boldsymbol{\theta}) = \prod_{i=1}^n [g(y_i \mid \boldsymbol{\theta}) f_Y(y_i)/P_L], \quad (8.14)$$

or

$$L(\{m_j\} \mid \boldsymbol{\theta}) = \left(\frac{n!}{\prod_{j=1}^J m_j!} \right) \prod_{j=1}^J f_j(\boldsymbol{\theta})^{m_j},$$

for grouped data with J distance intervals and m_j the number observed in the j th interval. Using the term “site” as a more general area, n_{jpr} is defined to be the number of counts at visit r to line or point p at site j as a Poisson random variable with $E[n_{jpr}] = \lambda_{jpr}$. To adjust counts for imperfect detection out to distance W , the

effective area sampled, $v(\boldsymbol{\theta})$, is defined as the area beyond which as many animals are seen as are missed within. Then the density D_{jpr} is given by

$$D_{jpr} = \lambda_{jpr}/v(\boldsymbol{\theta}). \quad (8.15)$$

For line transects, $v(\boldsymbol{\theta}) = 2L_p \int_0^W g(y | \boldsymbol{\theta}) dy$, where L_p is the length of the line surveyed; for point transects, $v(\boldsymbol{\theta}) = 2\pi \int_0^W yg(y | \boldsymbol{\theta}) dy$. These definitions for $v(\boldsymbol{\theta})$ are given for the case where all detections are pooled in a global detection function. However, when modeling heterogeneity, the effective area may vary between lines (points), and the global v becomes v_{jpr} . If we use the covariate model with K covariates,

$$D_{jpr} = \exp[\beta_0 + b_j + \sum_{k=1}^K x_{kjpr} \beta_k],$$

then from (8.15),

$$\lambda_{jpr}(\boldsymbol{\beta} | \boldsymbol{\theta}) = \exp[\beta_0 + b_j + \sum_{k=1}^K x_{kjpr} \beta_k + \log v(\boldsymbol{\theta})].$$

Here, $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_K, \sigma_b)'$, and b_j is the random effect for site j , where $b_j \sim N(0, \sigma_b^2)$. However, in cases where lines (points) follow a random survey design and each line (point) is surveyed only once, the random effect term may be omitted. Using the above model for λ_{jpr} , we have the count model likelihood expressed as

$$L_n(\boldsymbol{\beta} | \boldsymbol{\theta}) = \prod_{j=1}^J \int_{-\infty}^{\infty} \left(\prod_{p=1}^{P_j} \prod_{r=1}^{R_j} \frac{(\lambda_{jpr})^{n_{jpr}} \exp[-\lambda_{jpr}]}{n_{jpr}!} \cdot \frac{\exp[-b_j^2/(2\sigma_b^2)}{\sqrt{2\pi\sigma_b^2}} db_j \right). \quad (8.16)$$

Finally, using (8.14) and (8.16), the overall likelihood is given by

$$L_{n,y}(\boldsymbol{\theta}) = L_y(\boldsymbol{\theta}) L_n(\boldsymbol{\beta} | \boldsymbol{\theta}).$$

Uniform priors were placed on all the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ for which bounds were chosen in preliminary analyses. Generally, this involved adding/subtracting two times the standard errors from the maximum likelihood estimate of the full model; however, bounds were extended if a parameter value reached either of these. If the parameter had natural bounds, e.g., zero for the lower bound of the random effect standard deviation, these were adopted. A range of detection functions were fitted as well as seven multiple covariate models. A Metropolis-Hastings updating

algorithm was used so that different prior distributions for the parameters are readily implemented. A reversible jump Markov chain Monte Carlo (RJMCMC) algorithm allowed for model uncertainty to be incorporated. Computational details were given by Oedekoven et al. (2014).

Oyster et al. (2018) developed a Bayesian hierarchical model to estimate moose abundance and incorporated habitat use (in the form of availability as a function of canopy closure) into a detection model within a mark-recapture distance sampling framework. Their model of availability used a latent density surface employing habitat with data obtained from 17 adult female moose wearing Global Positioning System (GPS) collars. Distance sampling data was obtained from helicopter surveys.

8.15 Three-Dimensional Line Transects

Burnham et al. (2004) gave some theory for distance sampling in three dimensions as, for example, with fish under water. The observer might (a) traverse a “line” and record detection distances in two dimensions perpendicular to the line of travel or (b) remain at a point and record data in three dimensions within the sphere centered at the point.

Case (a)

Focusing on just the line, we assume that a line of length L is randomly placed in three dimensions and the density D is the number of objects per unit volume. All distances r are recorded for all objects detected out to perpendicular distance W , and counting takes place in a cylinder of volume $v = \pi W^2 L$ rather than a strip of area $a = 2WL$. Let P_v be the average probability of detecting an object in the sampled cylinder of volume v , and let $g(r)$ be the probability of detecting an object given it is present at a perpendicular distance r , $0 \leq r \leq W$. Since the objects are assumed to be uniformly distributed in space within the cylinder, the probability density function (pdf) of the radial distance r for a randomly specified object (before the detection process) is $\pi(r) = (2\pi r)/(\pi W^2)$ (a random circumference inside a circle of area πW^2). The unconditional detection probability is then $P_v = E[g(r)]$, namely,

$$P_v = \int_0^W \pi(r)g(r) dr = \frac{1}{\pi W^2} \int_0^W 2\pi r g(r) dr,$$

which is identical to the unconditional detection probability in point transects.

If n is the number of individuals seen, then to find $E[n]$, we let v_ε be a small volume in the cylinder centered at distance r and position ℓ along the line ($0 \leq \ell \leq L$). Then Dv_ε is the expected number of objects in volume $v_\varepsilon = 2\pi r dr d\ell$, and the

expected count of these objects is then $g(r)Dv_\varepsilon$. Hence, summing (integrating) over the small volumes, we have

$$\mathbb{E}[n] = \int_0^L \int_0^W g(r) D 2\pi r dr dl = LD \int_0^W 2\pi r g(r) dr = D\pi W^2 L P_v.$$

An estimator of D is

$$\widehat{D} = n/(\pi W^2 L \widehat{P}_v) = n/(L \widehat{\mu}),$$

where

$$\mu = \pi W^2 P_v = \int_0^W 2\pi r g(r) dr.$$

The pdf of distance r to a detected object is

$$f(r) = \frac{2\pi r g(r)}{\mu} = \frac{rg(r)}{\int_0^W rg(r) dr}.$$

This result is identical to $f_c(r)$ for point transects in Sect. 7.2.1 and can be proved using the same theory. In fact, slight modifications of point transect theory suffice as a complete theory for line transect sampling in three dimensions. In particular,

$$f'(r) = \frac{2\pi g(r)}{\mu} + \frac{2\pi r g'(r)}{\mu},$$

so if $g'(0)$ is finite (say equal to 0) and $g(0) = 1$, then $h(0) = f'(0) = 2\pi/\mu$ and

$$\widehat{D} = n\widehat{h}/(2\pi L),$$

where \widehat{h} is an estimate of $h(0)$. This is the same as \widehat{D}_W for point transects, with L replaced by k , the number of points selected. In fact, all the theory for point transects applies to line transect sampling in three dimensions if we replace k by L . Thus, estimation of $h(0)$ or P_v could be done using a program such as **Distance** and treating the detection distances, r_i ($i = 1, \dots, n$), as point transect data. Using the delta method,

$$\text{var}[\widehat{D}] = D^2 \left[\{CV(n)\}^2 + \{CV(\widehat{h})\}^2 \right], \quad (8.17)$$

where CV is the coefficient of variation. If the above theory is extended to groups or clusters of size g , then the density estimator is

$$\widehat{D}_g = (n\widehat{h}\widehat{\mathbb{E}}[g])/(2\pi L).$$

Case (b)

We have linked the three-dimensional line transect to a two-dimensional point transect. However, we now consider three-dimensional point transect sampling in its own right and focus on just the k points. The volume of the sphere surrounding a point is $v = (4\pi W^3)/3$, so that the expected sample size of detections at k random points is $E[n] = kDvP_v$, and $\pi(r)$ the pdf of the radial distance for a randomly selected object in the sphere is $\pi(r) = (4\pi r^2)/(4\pi W^3/3)$. This leads to

$$P_v = \int_0^W \pi(r)g(r) dr = \frac{3}{4\pi W^3} \int_0^W 4\pi r^2 g(r) dr$$

and

$$E[n] = kD(4/3)\pi W^3 P_v = kD \int_0^W 4\pi r^2 g(r) dr = kD\mu_1, \text{ say.}$$

Then

$$\widehat{D} = \frac{n}{k\widehat{\mu}_1} = \frac{n}{k(4\pi W^3/3)\widehat{P}_v},$$

and

$$f_c(r) = \frac{4\pi r^2 g(r)}{\mu_1}, \quad 0 \leq r \leq W.$$

Taking second derivatives, we get

$$f'_c(r) = \frac{8\pi g(r) + 16\pi r g'(r) + 4\pi r^2 g''(r)}{\mu}.$$

Hence, if $g(0) = 1$ and both $g'(0)$ and $g''(0)$ are finite (say zero as then the estimators have better properties), then $d(0) = f'_c(0) = 8\pi/\mu$ and

$$\widehat{D} = (n\widehat{d})/(8\pi k),$$

where \widehat{d} is an estimate of $d(0)$. Using the delta method,

$$\text{var}[D] = D^2 \left[\{CV(n)\}^2 + \{CV(\widehat{d})\}^2 \right]. \quad (8.18)$$

The authors extended the theory to clusters.

An Example

The question now is how to estimate $f_c(r)$ using some model for the detection function $g(r)$. The authors gave examples for both cases, and we consider just one of

them briefly for case (a), namely, $g(r)$ is the half-normal distribution with $W = \infty$, namely,

$$g(r) = \exp[-r^2/(2\sigma^2)]$$

so that $g(0) = 1$ and

$$f_c(r) = \sqrt{\frac{2}{\pi\sigma^2}} g(r), \quad 0 \leq r < \infty, \sigma > 0,$$

with $f_c(0) = \sqrt{2/(\pi\sigma^2)}$. Now, from Seber (1982, 462) with μ replaced by $\mu_\infty = \int_0^\infty g(y)dy$, we have

$$\text{E}[n] = 2LD\mu_\infty = 2LD/f_c(0).$$

Assuming n has a Poisson distribution, then $n \sim \text{Poisson}(2LD)/f_c(0))$. Setting $T = \sum_{i=1}^n r_i^2$ and then omitting constants and collapsing terms where possible, the likelihood is

$$L(D, \sigma) = \Pr(n) \prod_{i=1}^n f(r_i) \propto \exp \left[- \left(LD\sigma\sqrt{(2\pi)} + \frac{T}{2\sigma^2} \right) \right] D^n.$$

From the above equation, the maximum likelihood estimators are

$$\hat{\sigma} = \sqrt{T/n} \quad \text{and} \quad \hat{D} = \frac{1}{L} \sqrt{\frac{n}{2\pi T}} = \frac{n\hat{f}_c(0)}{2L}.$$

Then, by (8.17), the authors showed that

$$\begin{aligned} \text{var}[\hat{D}] &= D^2 \left[\{CV(n)\}^2 + \{CV(\hat{f}_c(0))\}^2 \right] \\ &= D^2 \left[\frac{1}{\text{E}[n]} + \frac{1}{2\text{E}[n]} \right]. \end{aligned}$$

The authors discussed profile likelihood intervals for D . In commenting generally on precision, they recommended that reliable, precise abundance estimates from distance data require minimum sample sizes around 100. Coefficients of variation of around 20% are often adequate for management purposes; the results presented indicated minimum sample sizes of 40–70 in this circumstance.

8.16 Summary

This is the second of two chapters which have a very similar basic mathematical theory. They both involve measuring distances to observed individuals or objects including animal signs. These distances can be from a point chosen by some scheme, as in the previous chapter, or from a point on a line in this chapter, where the perpendicular distance y from the object to the line is measured. There is incomplete detectability, so that some objects may not be observed (negative error) or misidentified for a given species (positive error). The main focus is on estimating population density.

The line or replicate lines are generally referred to as line transects, and distances are measured out to some distance W or any distance ($W = \infty$). The radial distance r to the object and the angle this distance makes with the transect are sometimes also measured, and theory is given for this situation. Since random placement can lead to uneven coverage of the study area, there is more focus on a model-based approach where various probabilities of detection functions are used with particular shape properties. The theory is extended to clusters of objects such as with schools of fish.

Estimation of the encounter rate as well as systematic sampling, pooling robustness, the use of multiple observers, and covariates are discussed. Several practical issues are considered such as the choice of transect(s), the total transect length, the type of transect such as a road or track or zigzag lines, and the number of transects and observation points on a transect. Theory is also given for transects of random length, which can occur with irregular-shaped study areas.

Plants; acoustic methods; the use of presence/absence data, with both independent and dependent detections; and log-linear models are given special mention. Model-based methods are extended to shipboard acoustic surveys and other marine surveys. Aerial censusing is considered. Spatial models are described as well as adaptive and Bayesian methods.

Chapter 9

Line Intercept Methods



9.1 Line Intercept

The line intercept method/transect (LIT) has been used by plant ecologists for many years to estimate plant density. The method is simple to describe and involves locating a line transect or transects at random and measuring the proportion of the line(s) intercepted by the objects (e.g., plants) and the number of objects intercepted. It is particularly useful when individual plants are not readily distinguishable so that one of its main uses has been to estimate canopy coverage of shrubs and other low-growing vegetation (cf. ways of measuring plant canopies and the terms used for them by Wilson, 2011).

The method was introduced in forestry by Warren and Olsen (1964) who coined the term “line intersect” sampling, and it has been used to rapidly estimate the volume of logging waste produced during clear-felling operations (see De Lima et al., 2020, for recent extensions). It has been used generally for downed woody debris (e.g., Woodall et al., 2012) and in streams, where there is an interest in such things as forest ecosystems, carbon dynamics, and wildfire behavior (see Fraver et al., 2018, for further references).

A very different type of application was given by Sidle et al. (2012) who used a LIT estimator to study the extent of prairie dog colonies in a large geographic area, backed up by a ground survey, to distinguish between inhabited and uninhabited burrows. Another type of application of the LIT method is to particles under a microscope such as food fragments to determine various proportions of food matter in an animal’s diet (Seber & Pemberton, 1979). It has also been used for assessing animal damage (Thomas et al., 2013) and has applications to subjects other than ecology such as with particles in a plane section of a metal sample.

Another major area of application is to objects on the sea bed. For example, Loya (1972) used the method to study coral, where it is found to be more efficient for information recorded for time spent underwater than with quadrat sampling. A coral

reef constitutes a very complex community of the marine environment as it involves several thousand species that occupy various ecological niches, a bit like the land version of a tropical forest. For the purpose of the method, an individual coral was defined as any colony growing independent of its neighbors (i.e., whenever an empty space was recorded between two adjacent colonies). When an individual colony was clearly separated by the death of intervening parts, the separate parts were considered as one individual. When point locations are made on a line, the length of the intercept under the point can also be measured. The method is often combined with the point-distance method.

9.1.1 Comparison of Methods

Leujak and Ormond (2007) listed six commonly used coral reef survey methods, mapping of 1 square meter quadrats (MAP), line point transect (LPT, see Sect. 9.2.4), line intercept transect (LIT), video sampling (VIDEO), and 1 square meter photo-quadrats analyzed by point sampling or by outlining coral colonies (PHOTS), and compared them for their accuracy and precision in recording benthic substrate cover and for their time- and cost-efficiency. Sampling was carried out along a 50 m transect at Ras Um Sid, South Sinai, Egypt. For abundant substrate categories (>10% cover), few significant differences were recorded between the results obtained by the different methods. Detailed mapping of quadrats (MAP) along the entire length of the transect served as the baseline against which the results of other methods were evaluated. Compared to this, VIDEO provided the most accurate cover estimates of major benthic substrate categories. Both line transect methods, LIT and LPT, overestimated the cover of some categories, while PHOTP and PHOTS underestimated them. The methods were compared with regard to time- and cost-efficiency as well as the purpose of using a particular method. Differences in accuracy were contributed to differences in contour effect, proportion of substrate sampled, view angle, image resolution, observer bias, and data analysis.

Urbina-Barreto et al. (2021) compared LIT with three methods derived from underwater photogrammetry. They concluded that LIT was “the least time-consuming and most effective for species-level taxonomic identifications but is the most limited method in terms of data outputs and representativeness of the ecosystem.” Facon et al. (2016) compared LIT and LIP for coral reefs using nonspecialist observers and found LPT supplied an almost equivalent level of information as LIT while reducing the time spent underwater and thus reducing the cost.

9.1.2 General Theory

A particle can be any shape, not necessarily convex, but is assumed to be a connected set of points (Buck, 1965, 29). Let N be the number of particles in a population

region \mathcal{R} of area A . If the region is rectangular of width W , and the line(s) of total length L run through it, any transect can be chosen at random by having a baseline, choosing a point at random on it, and having the line transect perpendicular in direction to the baseline. As some lines might fall too close together, an alternative procedure would be to consider a grid of equally spaced lines at a suitable distance apart and choose a random subset of the lines. Alternatively, for a nonrectangular region, we can choose the center of the transect at random and then independently choose its orientation with respect to some fixed baseline; the orientation could be fixed or else chosen at random. Another useful measurement is the maximum width w of the intersecting plant measured perpendicular to the transect line. It can also be described as the maximum distance between two tangent lines to the plant parallel to the transect.

We first consider some general theory from Lucas and Seber (1977). We assume that the clusters or individual particles are small compared to L , the length of the transect, and that L is small compared to the size of \mathcal{R} . A particle is said to be in \mathcal{R} if its “center,” that is, its center of gravity, lies in \mathcal{R} . Two boundary effects will be ignored. First, a particle with its center in \mathcal{R} may not be wholly in \mathcal{R} so that only part of it can intersect the transect. Second, the method of choosing the transect could lead to part of the transect lying outside of \mathcal{R} . However, in practice, both effects will be negligible if \mathcal{R} is large.

Two measures of density are of interest: first, the average number of particle centers per unit area, $D_C = N_C/A$, where N_C is the number of centers in \mathcal{R} , and second, the total proportion of the region covered by the particles, D_A , or area coverage. In this latter case, any overlapping plants can be treated as a single cluster. In the course of measuring the lengths of the individual intercepts, we can follow the convention that we choose one end of the transect at random and complete partial intercepts at this end; partial intercepts at the other end are ignored. We begin with the design whereby the transect is located at random by choosing its center at random and then independently choosing its orientation with respect to some fixed baseline.

9.1.3 Estimation of Coverage

Let y_i be the length of the intercept of the i th individual or clumped particle (cluster) intersecting the transect ($i = 1, \dots, m$), and let M be the number in the population. We now prove that

$$\widehat{D}_A = \sum_{i=1}^m y_i/L = Y/L, \quad \text{say}$$

is an unbiased estimate of D_A .

Let θ ($0 \leq \theta < \pi$) be the orientation of a noncircular particle of area a and perimeter c . For example, θ could be the angle between the transect and a well-

defined chord of the cluster, such as the longest chord, or some such convention. Let $g(\theta)$ be the probability density function of θ . Then we can write $w = w(\theta)$, since w will be a function of θ and of the shape of the cluster. Since θ is independent of the position of the transect,

$$P(\theta) = Lw(\theta)/A$$

is the probability (conditional on θ) that the particle intersects the transect, and the unconditional probability is

$$P = E[P(\theta)] = \frac{L}{A} \int_0^\pi w(\theta)g(\theta)d\theta = \frac{LC}{A}, \text{ say.} \quad (9.1)$$

If $g(\theta | \text{intersection})$ is the conditional probability density function of θ , given that the cluster intersects the transect, then using Bayes' rule, we find that

$$g(\theta | \text{intersection}) = P(\theta)g(\theta)/P = w(\theta)g(\theta)/C. \quad (9.2)$$

Further, suppose that u is the perpendicular distance from the transect to the “upper” of the two tangents which define the distance $w(\theta)$. Then, since the transect is randomly placed, u has a uniform distribution on $[0, w(\theta)]$ and

$$E[y | \theta, \text{ intersection}] = \frac{1}{w(\theta)} \int_0^{w(\theta)} ydu = \frac{a}{w(\theta)},$$

where a is the area of the particle. Using (9.2), we have

$$E[1/w(\theta) | \text{intersection}] = \int \frac{1}{w(\theta)} g(\theta | \text{intersection})d\theta = \int g(\theta)d\theta/C = \frac{1}{C}$$

and

$$E[y | \text{intersection}] = E_\theta E[y | \theta, \text{ intersection}] = a/C. \quad (9.3)$$

Let M be the number of clusters in the population. If we now define the random variable $z = y$ with probability P and $z = 0$ with probability $1 - P$, then

$$E[z] = PE[y | \text{intersection}] = La/A, \quad (9.4)$$

where a is the area of the particle and

$$E[\widehat{D}_A] = E \left[\sum_{i=1}^M z_i/L \right] = \sum_{i=1}^M a_i/A = D_A,$$

so that \widehat{D}_A is unbiased.

If the transect is randomly oriented, then

$$g(\theta) = 1/\pi, \quad (0 \leq \theta \leq \pi), \quad (9.5)$$

but the θ_i may not be mutually independent. If we can assume that the clusters are randomly distributed and randomly oriented, then the u_i are mutually independent, the θ_i are mutually independent with density (9.5), the z_i are mutually independent, and

$$\text{var}[\widehat{D}_A] = \frac{1}{L^2} \left[\sum_{i=1}^M \left\{ E[z_i^2] - \frac{L^2}{A^2} a_i^2 \right\} \right]. \quad (9.6)$$

When the particles are circles and the distribution is dilute, that is, there is no overlapping, an alternative estimate based on the diameters of the intersect circles can be obtained. If a circle of diameter x has intercept y , then from Kendall and Moran (1963, 88),

$$E[(y | x) = \frac{1}{4}\pi x \quad \text{and} \quad E[y^{-1} | x] = \frac{1}{2}\pi/x, \quad (9.7)$$

and the first equation leads to

$$\widetilde{D}_A = \sum_{i=1}^m \frac{1}{4}\pi x_i / L, \quad (9.8)$$

which is also unbiased.

If we further assume that the particles are convex and dilute, then we have the well-established result on convex figures (Blaschke, 1949; Kendall & Moran, 1963, 58) that $C = c/\pi$, where c is the perimeter of the particle. Also, from Crofton's second theorem on convex figures (Kendall & Moran, 1963, 65) and (9.2),

$$E[y^3 | \text{intersection}] = \int_0^\pi y^3 w(\theta) d\theta / c = 3a^2/c,$$

so that $E[z^3] = 3a^2 P/c = 3La^2/(A\pi)$. Hence,

$$\widehat{\text{var}}[\widehat{D}_A] = \frac{1}{L^2} \sum_{i=1}^m \left(y_i^2 - \frac{L\pi}{3A} y_i^3 \right)$$

is an unbiased estimate of (9.6). Since Ly_i/A will be small, the second term of the above equation will generally be negligible. Using the same assumptions and a similar technique, an unbiased estimate of $\text{var}[\widetilde{D}_A]$ can also be found as follows.

We define $X = x$ with probability $P = Lx/A$ and $X = 0$ with probability $1 - P$. Then $E[X^3] = Px^3 = (E[X])^2 A/L$, and we find that

$$\frac{\pi^2}{16L^2} \sum_{i=1}^m \left(x_i^2 - \frac{L}{A} x_i^3 \right)$$

is an unbiased estimate of $\text{var}[\tilde{D}_A]$.

9.1.4 Estimation of Particle Density

For the j th particle (not cluster) intersecting the transect, let w_j be the maximum perpendicular distance between tangents to the j th individual particle that are parallel to the transect (i.e., the maximum width perpendicular to the transect) for $j = 1, \dots, n$. In what follows, we use what is effectively known as a size-biased sampling so that bigger objects are more likely to be sampled. We begin by showing that

$$\hat{D}_C = \sum_{j=1}^n w(\theta_j)^{-1}/L$$

is an unbiased estimate of $D_C = N_C/A$, where N_C is the number of particle centers in \mathcal{R} . Let I_j take the value of $w(\theta_j)^{-1}$ with probability $P(\theta_j) = Lw(\theta_j)/A$ and 0 with probability $1 - P(\theta_j)$. Then

$$\begin{aligned} E[\hat{D}_C] &= E_\theta[E(\sum_{j=1}^N I_j/L)] \\ &= E_\theta[\sum_{j=1}^N w(\theta_j)^{-1} P(\theta_j)/L] \\ &= E_\theta[N_C/A] = D_C. \end{aligned}$$

Arguing as in the previous section, we find, for randomly distributed particles, that

$$\widehat{\text{var}}[\hat{D}_C] = \frac{1}{L^2} \left(\sum_{j=1}^n w_j^{-2} \right) - \frac{\hat{D}_C}{A}$$

is an unbiased estimator of $\text{var}[\hat{D}_C]$.

When the particles are circles and the distribution is dilute, we can set $w_j = x_j$ and use the second equation of (9.7), to get Fullman's (1953) unbiased estimator

$\tilde{D}_C = \sum_{j=1}^n 2y_j^{-1}/(\pi L)$. It is usually quoted as being unbiased for random distributions when, in fact, it is unbiased for any distribution and a random transect. Some other estimates are also given by Lucas and Seber (1977).

Eberhardt (1978) also discussed the above method in the context of flushing out animals by assuming that an animal has a surrounding circular area which, if intercepted, causes the animal to flush into the open. Using the line intercept theory above, Hayne's (1949a) density estimator can be shown to be unbiased under certain conditions.

The assumption of a random distribution is not usually met in nature so that the variance estimates may be of limited usefulness. Also, the union of overlapping convex particles is no longer convex, though perhaps the conglomerate particle could be considered as two convex adjacent particles on either side of the common chord of intersection. In practice, it would be preferable to obtain variance estimates from replication, that is, take s transects or else divide the transect length L into s segments of length L/s . In the latter case, estimates based on adjacent segments may be correlated so that a variance estimate such as that given by Seber (1982, 6–7) might be appropriate.

Comparing the various estimates for circular particles, we note that the reciprocals y_i^{-1} will be sensitive to measurement errors and boundary irregularities for small chords. This together with the fact that $\text{var}[\hat{D}_C]$ does not exist for random circles throws doubt on the usefulness of \hat{D}_C . These misgivings are reinforced by the paper of Watson (1971), which shows up the inadequacies of \hat{D}_C for the case of planes intersecting random spheres. However, \hat{D}_C will be more robust because w_i will be less sensitive than y_i to any kind of irregularity.

In the above theory, we investigated two-dimensional planar density by looking at intercepts on a one-dimensional line. Clearly, this idea can be taken up a dimension to look at planar sections of three-dimensional objects as in sampling rocks for mineral deposits such as quartz or taking cross sections of human tissue to detect the extent of damage to the tissue. A first sample gives a slice of the object, and then we can use a line intersect method on the slice. This leads to the subjects of petrography, quantitative metallography, and stereology.

9.1.5 Kaiser's Generalization

Kaiser (1983), who reviewed the topic, has generalized the above theory as follows. Suppose there are n particles intercepted. For particle i , let x_i ($i = 1, \dots, N$) be some fixed characteristic of an individual and y_i be a variable that depends on the transect sampled. For instance, x_i could be the area of the particle and y_i the length of its intersection, as in the above theory.

Let some arbitrary direction in the plane be denoted by $\theta = 0$. As with the above methodology, a point is chosen at random that forms the midpoint of a transect of length L with random direction θ , which has the density function a uniform

distribution on $[0, \pi]$. By having a random transect, we do not need the particles to be randomly distributed.

Our interest is focused on estimating $\lambda_x = \sum_{i=1}^N x_i/A$. At first, we ignore problems of the edge effect leading to boundary overlap. Let $t_i = 1$ if the i th object is intersected with probability $P_i(\theta) = Lw_i(\theta)/A$ and 0 otherwise (previously $z_i = t_i y_i$). Kaiser proved that

$$\mathbb{E}[t_i | \theta] = \Pr(t_i = 1 | \theta) = \frac{Lw_i(\theta)}{A}, \quad (9.9)$$

and, taking the expectation over θ ,

$$\mathbb{E}_\theta[w_i(\theta)] = \int_0^\pi w_i(\theta)g(\theta)d\theta = C.$$

This leads to

$$\begin{aligned} \mathbb{E}[t_i] &= \frac{L\mathbb{E}_\theta[w_i(\theta)]}{A}, \\ &= LC/A. \end{aligned} \quad (9.10)$$

If

$$\hat{\lambda}_1 = \frac{1}{A} \sum_{i=1}^n \frac{t_i y_i}{\mathbb{E}[t_i y_i]} x_i = \frac{1}{A} \sum_{i=1}^N \frac{t_i y_i}{\mathbb{E}[t_i y_i]} x_i, \quad (9.11)$$

then

$$\mathbb{E}[\hat{\lambda}_1] = \sum_{i=1}^N x_i/A = \lambda_x.$$

Also, we have the estimator

$$\hat{\lambda}_2 = \frac{1}{A} \sum_{i=1}^n \frac{t_i y_i}{\mathbb{E}[t_i y_i | \theta]} x_i, \quad (9.12)$$

for which

$$\mathbb{E}[\hat{\lambda}_2 | \theta] = \sum_{i=1}^N x_i/A = \lambda_x \quad \text{and} \quad \mathbb{E}[\hat{\lambda}_2] = \lambda_x.$$

If the distribution of θ is degenerate, then $\hat{\lambda}_1 = \hat{\lambda}_2$.

Returning to the problem of possible edge effect, we see that for particles within a distance $\frac{1}{2}L$ of the boundary of \mathcal{R} , (9.9) will not hold for all θ . One way to make it hold when \mathcal{R} is convex is to bring into \mathcal{R} any portion of the transect which lies outside of \mathcal{R} by continuing the transect a perpendicular distance d ($> \max\{w_i(\theta)\}$) that no particle can be intercepted twice by the same transect) away from the portion of the transect already in \mathcal{R} . Using similar arguments, Kaiser showed that $E[t_i | \theta]$ then still holds.

By properly specifying the shape of the particles and properly choosing y_i and the distribution of θ in the above two estimates of λ , Kaiser gives examples to show that all previous estimators up till 1983 are special cases of the above theory. We consider just three. Firstly, if the particles are of arbitrary shape and we set $y_i = 1$ for all i , then

$$\hat{\lambda}_1 = \frac{1}{L} \sum_{i=1}^n \frac{t_i x_i}{C_i},$$

and

$$\hat{\lambda}_2 = \frac{1}{L} \sum_{i=1}^n \frac{t_i x_i}{w_i(\theta)}.$$

Then setting $x_i = 1$ for all i , $\hat{\lambda}_2 = \hat{D}_C$, the estimate of particle density given previously. For arbitrary x_i , we have an estimator of McDonald (1980). Secondly, we find from the previous theory that $E[t_i y_i | \theta] = La_i/A = E[t_i y_i]$ so that using (9.11) and (9.12),

$$\hat{\lambda}_1 = \hat{\lambda}_2 = \frac{1}{L} \sum_{i=1}^n t_i y_i = \hat{D}_A,$$

our coverage estimator given above. Thirdly, suppose the articles in \mathcal{R} are the projections onto \mathcal{R} of three-dimensional objects resting on \mathcal{R} . If we wish to estimate $V/A = \sum_{i=1}^N v_i/A$, where v_i is the volume of the i th particle, then we find that $E[z_i | \theta] = Lv_i/A = E[t_i y_i]$ and

$$\hat{\lambda}_1 = \hat{\lambda}_2 = \frac{1}{L} \sum_{i=1}^n t_i y_i$$

is an unbiased estimate of V/A . Kaiser gives further examples including those based on “needles” used in logging.

9.2 Transects of Random Length

Suppose, as usual, we set up a baseline of length δ running between two vertical perpendicular tangents to the population area. The line transect is now considered as running the full length of the population area perpendicular to the baseline, and its position is determined from a random point on the baseline segment at distance u . This means that L has random length and u has a uniform distribution $1/\delta$, $0 \leq u \leq \delta$ (cf. Seber, 1979). Also, θ is not involved in the theory.

9.2.1 Coverage

We recall from Sect. 9.1.3 the estimate of coverage, namely,

$$\widehat{D}_A = \sum_{i=1}^m y_i / L = Y / L.$$

When L is random, this estimator is now biased. If T transects are used, we can use a jackknife method to obtain an estimator with much smaller bias. Let D_{-t} ($t = 1, \dots, T$) be the estimate of D_A with the same form as \widehat{D}_A , but based on $T - 1$ transects, with the data from transect t omitted. Define

$$\widehat{D}_{At} = T\widehat{D}_A - (T-1)D_{-t}, \quad \text{and} \quad \widetilde{D}_A = \sum_{t=1}^T \widehat{D}_{At} / T.$$

The variance of \widetilde{D}_A can be estimated by

$$s_A^2 = \sum_{t=1}^T (\widehat{D}_{At} - \widetilde{D}_A)^2 [T(T-1)],$$

and $(\widetilde{D}_A - D)/s$ is approximately distributed as the t -distribution with $T - 1$ degrees of freedom.

If we can assume that the clusters are randomly distributed and randomly oriented, and a cluster intercepts the transect, then the length of the intercept will not be affected by the length of the transect. Hence, $E[\widehat{D}_A | L] = D_A$ and \widehat{D}_A is unconditionally unbiased. Further variations are given by Barabesi and Marcheselli (2008) with a focus on replicate lines.

9.2.2 Particle Density

Let N be the number of particles in region \mathcal{R} . Let w_j be the usual distance defined above for object j intercepting the transect, and let $P_j = w_j / \delta$ be the probability

that the transect intersects the particle. We define I_j to take the value of $1/w_j$ with probability P_j and 0 with probability $1 - P_j$. If $Z = \sum_{j=1}^n w_j^{-1} = \sum_{j=1}^N I_j$, then

$$\mathbb{E}[I_j] = w_j^{-1} P_j = 1/\delta,$$

and $\mathbb{E}[Z] = N/\delta = DA/\delta$. Here, Z can be used to estimate N .

If L was constant, we could use $\widehat{D}_C = \sum_{j=1}^n w_j^{-1}/L = Z/L$ from the previous theory. When L is random,

$$\mu_L = \mathbb{E}[L] = (1/\delta) \int_0^\delta L du = A/\delta,$$

and we can use the delta method (A.4) to obtain

$$\mathbb{E}[\widehat{D}_C] \approx \frac{\mu_Z}{\mu_L} \left(1 + \frac{\sigma_L^2}{\mu_L^2} - \frac{\sigma_{LZ}}{\mu_L \mu_Z} \right) \quad (9.13)$$

$$\begin{aligned} &= D(1 + C_L^2 - \rho C_L C_Z) \\ &= D(1 + b), \quad \text{say,} \end{aligned} \quad (9.14)$$

where μ_Z and σ_Z^2 are the mean and variance of Z , σ_{LZ} and ρ are the covariance and correlation of L and Z , and C_L and C_Z are the respective coefficients of variation. We have that \widehat{D}_C is now biased, but the bias will be small if C_L and C_Z are small.

In practice, one could take measurements on T transects giving the pairs (Z_t, L_t) ($t = 1, \dots, T$) and the estimator

$$\overline{D}_C = \frac{1}{T} \sum_{t=1}^T (Z_t/L_t) = \frac{1}{T} \sum_{t=1}^T \widehat{D}_{Ct},$$

with variance estimate

$$\widehat{\text{var}}[\overline{D}_C] = \sum_{t=1}^T (\widehat{D}_{Ct} - \overline{D}_C)^2 / [T(T-1)].$$

However, a better estimator is the ratio estimator

$$\begin{aligned} \widehat{D}_r &= \sum_{t=1}^T Z_t / \sum_{t=1}^T L_t \\ &= \sum_{t=1}^T L_t \widehat{D}_{Ct} / \sum_{t=1}^T L_t. \end{aligned} \quad (9.15)$$

Using the delta method again, we now find that

$$\mathbb{E}[\hat{D}_r] \approx D(1 + b/T),$$

where the terms neglected are of order T^{-2} if the pairs (Z_t, L_t) are mutually independent or approximately so. Here, b can be eliminated using the jackknife method as described above for the coverage estimator.

Up till now, we have made no assumption about the distribution of the particle centers; only that the line transect is randomly located. However, if we can assume that the objects are randomly distributed and randomly oriented, then \hat{D}_C is now an unbiased estimate of D_C . Since the particle is randomly placed, the probability that it lies in a given area will depend on the size of the area. Then, for a given transect of length L , $P_j = Lw_j/A$ is the probability that the particle is in the area to be intercepted. Let I_j take the value w_j^{-1} with probability P_j and 0 with probability $1 - P_j$. Then

$$\mathbb{E}[Z | L] = \mathbb{E}\left[\sum_{j=1}^N I_j | L\right] = \sum_{j=1}^N w_j^{-1} L w_j / A = NL/A = DL, \quad (9.16)$$

and since the I_j are mutually independent, we see that

$$\begin{aligned} \text{var}[Z | L] &= \sum_{j=1}^N \text{var}[I_j] \\ &= \sum_{j=1}^N w_j^{-2} P_j (1 - P_j) \\ &= \frac{L}{A} \sum_{j=1}^N \left(\frac{1}{w_j} - \frac{L}{A}\right) \\ &\approx \frac{L}{A} \sum_{j=1}^N w_j^{-1} = kL, \quad \text{say,} \end{aligned} \quad (9.17)$$

since a random distribution of particles implies that the I_j are mutually independent, and in practice, Lw_j is a lot less than A .

If we now have T transects, we see from (9.16) and (9.17) that we can fit a (conditional) weighted least squares regression line of Z_t on L_t through the origin using weights $1/L_t$. Our least squares estimator is again the ratio \hat{D}_r (8.15) with variance estimate

$$v = \hat{\sigma}^2 / \sum_t L_t,$$

where

$$\begin{aligned}\hat{\sigma}^2 &= \sum_{t=1}^T (Z_t - \hat{D}_r L_t)^2 / [L_T(T-1)] \\ &= \sum_{t=1}^T L_t \left(\frac{Z_t}{L_t} - \hat{D}_r \right)^2 / (T-1).\end{aligned}$$

The linearity (or nonlinearity) of the data (Z_i, L_i) will throw some light on the assumption of a random distribution of particles.

9.2.3 Method of Kaiser

We now choose a point at random in \mathcal{R} and run a transect through this point in the direction θ generated independently with an arbitrary distribution. As before, the transect runs entirely across \mathcal{R} , so that its length, L , is a random variable. Once again, we set up a baseline perpendicular to the transect, and it now runs between two parallel tangents to \mathcal{R} that are parallel to the transect. The perpendicular distance between them is $\delta(\theta)$ as above, but it now depends on θ . Let T denote the left-hand tangent to \mathcal{R} . We now measure three perpendicular distances to T : $u(\theta)$ the distance from the transect and for particle i the distance r_i from the particle's left-hand tangent and s_i from its right-hand tangent so that $w(\theta) = s_i(\theta) - r_i(\theta)$, the maximum width of the particle in the direction of the transect. Kaiser (1983) developed the following theory.

If the set of all transects is given by the position $\{(\theta, u) : 0 \leq \theta < \pi, 0 < u < \delta(\theta)\}$, let $L(\theta, u)$ be the length of a transect in that position. Let $f(u | \theta)$ be the conditional density function of u given θ . Then, given θ , the transect will have u in the interval $(u, u + du)$ if the random point chosen in \mathcal{R} is in a strip of area $L(\theta, u)du$ about the transect. Thus, the density function of u given θ is $f(u | \theta) = L(\theta, u)/A$, for $0 < u < \delta(\theta)$ and zero otherwise. We therefore generate $\theta \in [0, \pi)$ and then generate u with $0 < u < \delta(\theta)$ from the density $f(\theta, u)/A$. In order to distinguish from Kaiser's theory in Sect. 9.1.5, we follow his suggestion and use a * for moments with the model here. Now, from the previous theory, we recall that y_i is the length of the intercept of the i th particle by the transect and, given L , $t_i = 1$ if the i th particle is intersected with probability $P(\theta_i) = Lw(\theta_i)/A$ and 0 otherwise. Then, from the above theory,

$$E[t_i y_i | \theta] = E[y_i | \theta, t_i = 1] L w_i(\theta) / A.$$

Since, given θ and a fixed-length transect, u has a uniform distribution on $(0, w(\theta))$,

$$w_i(\theta) E[y_i | \theta, t_i = 1] = \int_0^{w_i(\theta)} y_i \{ \theta, r_i(\theta) + u \} du = I(\theta), \quad \text{say.} \quad (9.18)$$

Now,

$$\begin{aligned}
E^* \left[\frac{At_i y_i}{L} \mid \theta \right] &= \int_0^{\delta(\theta)} \left\{ \frac{At_i(\theta, u)y_i(\theta, u)}{L(\theta, u)} \right\} f(u \mid \theta) du \\
&= \int_0^{\delta(\theta)} \left\{ \frac{At_i(\theta, u)y_i(\theta, u)}{L(\theta, u)} \right\} \left\{ \frac{L(\theta, u)}{A} \right\} du \\
&= \int_{\{u \mid t_i(\theta, u)=1\}} y_i(\theta, u) du \\
&= \int_{r_i(\theta)}^{s_i(\theta)} y_i(\theta, u) du = I(\theta).
\end{aligned}$$

Hence,

$$E^* \left[\frac{t_i y_i}{E[t_i y_i \mid \theta]} \mid \theta \right] = E^* \left[\frac{t_i y_i}{I(\theta)(L/A)} \mid \theta \right] = I(\theta)/I(\theta) = 1.$$

From (9.12), we have

$$\hat{\lambda}_2 = \frac{1}{A} \sum_{i=1}^n \frac{t_i y_i}{E[t_i y_i \mid \theta]} x_i$$

so that, since we can replace n by N (because of t_i),

$$E^*[\hat{\lambda}_2] = \lambda_x = \sum_{i=1}^N x_i / A.$$

By taking expectations over θ in the above, we have from (9.11) that

$$\hat{\lambda}_1 = \frac{1}{A} \sum_{i=1}^n \frac{t_i y_i}{E[t_i y_i]} x_i = \frac{1}{A} \sum_{i=1}^N \frac{t_i y_i}{E[t_i y_i]} x_i$$

and $E^*[\hat{\lambda}_1] = \lambda_x$. This means that with a slight change in the sampling protocol, the estimates \hat{D}_A and \hat{D}_C given above for coverage and particle density are now unbiased. Care is needed, however, as with the change in the method of locating the line transect, it could turn out that the line could be very short because of the choice of random point combined with a random θ leading to being a narrow part of the region \mathcal{R} . With the previous method, we choose a baseline of maximum width, depending of course on access to the region. This may decrease the probability of an edge effect to be dealt with. A slight modification of Kaiser's design was given by Barabesi and Marcheselli (2008).

9.2.4 *Line Point Intercept/Transect*

This method involves measuring species frequencies by recording all species hit by a vertical rod placed at predetermined systematic intervals along the line or a grid. It can be effected by spatial autocorrelation, depending on the distance between points. It is essentially an occupancy problem (Chap. 4), particularly useful for estimating species richness (Chap. 5). There does not seem to be any theoretical background for the topic.

9.3 Summary

This chapter is related to the previous chapter in that a line transect or transects are used, except that different measurements are taken, namely, (1) the proportion of the line(s) intercepted by the objects (e.g., plants) and (2) the number of objects intercepted, for any shaped object. It is particularly useful when individual plants are not readily distinguishable when they overlap and is commonly used to estimate canopy coverage of shrubs and other low-growing vegetation. The theory depends on how transects are located as using some type of random location avoids the need to assume that particles are randomly distributed. In the past, the two measurements required slightly different theory, but more recently have been incorporated under the same general model. The theory is extended to transects of random length.

Chapter 10

Removal and Change-in-Ratio Methods



10.1 Basic Models

The so-called removal method is a special case of a method due to De Lury (1947) and was studied further by Moran (1951), Zippin (1956, 1958), and Seber (1982, Section 7.2), along with an extensive survey of the subject by De Rivera and McCrea (2021). Removal sampling has been used for a variety of species including insects, birds, small mammals, freshwater fish, and human populations. The method involves successively removing samples of sizes n_i ($i = 1, 2, \dots, s$) from a closed population, initially of size N . It is particularly useful in dealing with invasive species (e.g., Davis et al., 2016, feral swine) and with culling overabundant species (e.g., Chee & Wintle, 2010) and has been used for translocations, which is the movement of protected species to prevent their extinction prior to the development of land. Translocations involve capture, relocation and release, or introduction of species from one area to another habitat. A related topic is mitigation-driven translocations that are implemented in response to legislation or governmental regulation, with the intent of reducing a development project's effects on animals or plants inhabiting a site (Germano et al., 2015).

The removal method is referred to elsewhere in this book such as in occupancy modeling, distance methods, and catch-effort methods, for example. It is also a special case of a capture-recapture model in which recaptures are ignored (i.e., effectively “removed” from the population); the case $s = 2$ is considered in detail in Seber (1982, 323). It should be noted that the trapping effort used in catching untagged animals may fall off significantly as tagged animals take up more and more of the traps. The removal method has also been incorporated into other models involving open populations such as that given by Matechou et al. (2016) and robust models (Pollock, 1982) that use primary and secondary sampling periods (e.g., Zhou et al., 2018, lizards, allowing for temporary emigration).

One approach is to use a removal sequence after a capture-recapture experiment to obtain a combined estimator. This means combining a capture recapture model like M_t of Sect. 12.5.1 with one like that described below by multiplying the two probability functions together, but using a common N . For simple random sampling with fixed sample sizes, one could also use the hypergeometric model (12.10) for the mark-recapture model. Various assumptions about the various probabilities involved can be incorporated for the two phases of the census. This is considered in Sect. 10.1.3.

The removal method was initially described in the context of trapping animals with the key assumption being that the trapping effort (capture efficiency) is constant throughout the trapping period so that the probability of catching an individual animal p_i on the i th trapping occasion is a constant p , considered in the next section. Let n_i be the size of the i th removal. We begin with $p_i (=1-q_i)$ varying from sample to sample due to variation in removal effort so that the joint distribution of the $\{n_i\}$ for s removals is

$$\begin{aligned} f(\{n_i\}) &= \prod_{i=1}^s \binom{N - x_i}{n_i} p_i^{n_i} q_i^{N - x_{i+1}} \\ &= \frac{N!}{(\prod_{i=1}^s n_i!) (N - x_{s+1})!} p_1^{n_1} (q_1 p_2)^{n_2} \cdots (q_1 q_2 \cdots q_{s-1} p_s)^{n_s} \\ &\quad \times (q_1 q_2 \cdots q_s)^{N - x_{s+1}}, \end{aligned} \tag{10.1}$$

where $x_i = \sum_{j=1}^{i-1} n_j$ ($i = 1, \dots, s+1$), the cumulative sum of the previous sample sizes. As we have too many parameters to estimate, we can then use a logistic regression approach such as

$$\text{logit}(p_i) = \alpha + \beta z_i,$$

for example, where z is a covariate (or one of several). In a trapping experiment, if w_i is the number of traps available for the i th removal and θ is the constant probability of capture in a single trap, then $p_i = 1 - (1 - \theta)^{w_i}$.

10.1.1 Constant Probability of Removal

A special case assumes the p_i are all equal to p (Zippin, 1956). Then

$$f(\{n_i\}) = \frac{N!}{\prod_{i=1}^s n_i! (N - x_{s+1})!} p^{x_{s+1}} q^{sN - \sum_{i=1}^{s+1} x_i}.$$

The maximum likelihood estimates for \widehat{N} and \widehat{p} are given by

$$\widehat{N} = x_{s+1}/(1 - \widehat{q}^s)$$

and

$$\frac{\widehat{q}}{\widehat{p}} - \frac{s\widehat{q}^s}{(1 - \widehat{q}^s)} = \frac{\sum_{i=1}^s (i-1)n_i}{x_{s+1}} (= R, \text{ say}),$$

which has a unique solution \widehat{q} in the range $[0,1]$ for $0 \leq R \leq (s-1)/2$ (Seber, 1970, Appendix). When $R > (s-1)/2$, $\widehat{q} > 1$ and the above method is not applicable, the experiment “fails.” This condition for failure can be rewritten as

$$\sum_{i=1}^s (s+1-2i)n_i < 0. \quad (10.2)$$

The probability of this happening will generally be small for large N and will decrease as s increases. For large N , \widehat{N} and \widehat{p} are asymptotically unbiased with asymptotic variances

$$V[\widehat{N}] = \frac{N(1-q^s)q^s}{(1-q^s)^2 - (ps)^2q^{s-1}}$$

and

$$V[\widehat{p}] = \frac{(qp)^2(1-q^s)}{N[q(1-q^s)^2 - (ps)^2q^s]}.$$

A standard goodness-of-fit test for the above model is

$$T = \sum_{i=1}^s (n_i - E_i)^2/E_i,$$

where $E_i = \widehat{N}\widehat{p}\widehat{q}^{i-1}$ ($i = 1, \dots, s$). This is asymptotically distributed as χ_{s-2} when the initial model is valid.

If p varies with size, then separate size classes can be used if there are enough individuals. Confidence intervals for N are described by Bedrick (1994) and Hirst (1994). One extension of the above, given by Liu et al. (2003), is when there are two populations being sampled (e.g., male and female) and the ratio of the population sizes is known. It is extended by You and Liu (2006) to the proportional trapping removal described in Sect. 10.1.2.

Seber (1982, 312–314) discussed the question of designing the above experiment to achieve a certain coefficient of variation. He gave a table from Zippin (1956) looking at different coefficients of variation as they relate to N and the proportion of the population that needs to be captured. He noted that a serious limitation of removal methods is that a relatively large proportion of the population must be captured, say greater than 0.3, especially when $N < 200$, in order to obtain reasonably accurate estimates. An exceptional situation is when it is desirable to reduce a population sharply, as in the case of a crop-damaging rodent. Three special cases $s = 3, 2, 1$ are considered below.

In concluding this section, it should be noted that removal sampling can be quite labor-intensive and in extreme cases may alter the habitat in attempts to reduce cover to find individuals. These factors combine to limit the number of sites that can be sampled and often result in large variances in derived population estimates. High amounts of variance in estimates result in low statistical power to detect population trends (Dodd & Dorazio, 2004). The design of a removal experiment is considered in Sect. 10.5.

10.1.2 Trap Reduction

Another variation on the removal method is when an allowance is made for trap reduction, as in many cases a trap is not emptied immediately so that the number of available traps reduces. Earlier methods were given by Good et al. (1979). Liu and Yip (2002) considered the following model referred to as the proportional trapping-removal method in continuous time using the following assumptions.

- (1) Each trap can catch at most one animal, and the capture time is recorded. The caught animal remains in the trap till the end of the experiment and can be regarded as having been removed from the population.
- (2) The animals behave homogeneously, i.e., the risks of capture are the same.
- (3) There is no interference from animals of another species.

We give just an overview in what follows, as there is deeper underlying theory. Let N be the population size from which n animals are caught in the time period $[0, T]$ at times $0 < t_1, \dots, t_n < T$ and m traps are used. Given the above assumptions, the hazard rate for any animal to be caught at time $t \in [0, T]$ is proportional to the number of available traps. Let $n_i(t) = 1$ if the i th animal has been caught up to time t and 0 otherwise. Define $n(t) = \sum_{i=1}^N n_i(t)$, the total number of captures up to time t . Then

$$\eta_i(t) = n_i(t) - \int_0^t \lambda[m - n(s-)][1 - n_i(s-)]ds$$

is a \mathcal{F}_t^* martingale. Here, $1 - n_i(s-)$ indicates if the i th animal is at risk just before time s ; λ is a constant baseline hazard, which can be interpreted as hazard rate if

there are only one trap and only one animal (see Good et al., 1979); and $\mathcal{F}_t^* = \sigma\{n_i(s) : 0 \leq s \leq t, i = 1, \dots, N\}$. Summing over i , we obtain an \mathcal{F}_t martingale, $\eta(t) = \sum_{i=1}^N \eta_i(t)$, i.e.,

$$\eta(t) = n(t) - \int_0^t \lambda[N - n(s-)][m - n(s-)]ds,$$

where $\mathcal{F}_t = \sigma\{n(s) : 0 \leq s \leq t\}$. Here, σ represents a sigma field generated by the processes up to t .

Let $s_1 = t_1$, $s_i = t_i - t_{i-1}$ ($i = 2, \dots, n$) and $s_{n+1} = T - t_n$. Equating $\eta(t)$ to its mean, namely, zero, we obtain

$$n = \lambda \sum_{i=1}^n (N - i + 1)(m - i + 1)s_i,$$

which is essentially a score function. The likelihood function $L(\lambda, N)$ is then obtained, and the maximum likelihood estimate of N is the solution of

$$\frac{n}{N - C} = \sum_{i=1}^n \frac{1}{N - i + 1},$$

where

$$C = \frac{\sum_{i=1}^{n+1} (i - 1)(m - i + 1)s_i}{\sum_{i=1}^{n+1} (m - i + 1)s_i}.$$

The authors give conditions for the existence and uniqueness of \hat{N} . The estimate of λ is given by

$$\hat{\lambda} = n / [\sum_{i=1}^{n+1} (\hat{N} - i + 1)(m - i + 1)s_i].$$

An expression is given for the asymptotic covariance matrix for $(\hat{\lambda}, \hat{N})$.

The above method was extended to contaminated data (disturbance from non-target animals) by Liu and Yip (2002) and to the case when there are two classes (e.g., male and female) by You and Liu (2006).

10.1.3 Combined with a Mark Release

Yip and Fong (1993) modified the above removal method where a known number of marked (or tagged) animals is released into the population before the experiment

commences. The removal probabilities are allowed to be time dependent and are assumed to be the same for marked and unmarked animals, and there is homogeneous mixing. Let

$$\tau = \text{total unmarked population size}$$

$$D = \text{number of marked initially released}$$

$$n_i = \text{number removed in sample } i \ (i = 1, \dots, s)$$

$$m_i = \text{number marked removed in sample } i$$

$$u_i = \text{number unmarked removed in sample } i$$

$$M_i = \sum_{k=1}^{i-1} m_k, \text{ total marked removed before sample } i$$

$$U_i = \sum_{k=1}^{i-1} u_k, \text{ total unmarked removed before sample } i$$

$$N_i = M_i + U_i$$

$$N = D + \tau$$

$$p_i = \text{probability an individual is caught in sample } i$$

The likelihood function, conditional on the n_i , is

$$L(\tau) = \prod_{i=1}^s \frac{\binom{D - M_i}{m_i} \binom{\tau - U_i}{u_i}}{\binom{\tau + D - N_i}{n_i}},$$

with maximum likelihood estimate

$$\hat{\tau} = D \frac{U_{s+1}}{M_{s+1}},$$

which is similar to the Petersen estimate (Sect. 12.2). Assuming the marked and unmarked populations are independent, an approximate variance is given by

$$\text{var}[\hat{\tau}] = \tau \left(1 + \frac{\tau}{D}\right) \left(\frac{1 - Q_s}{Q_s}\right),$$

where $Q_s = 1 - \prod_{i=1}^s (1 - p_i)$ and p_i is estimated by $\hat{p}_i = n_i / (\hat{\tau} + D - N_i)$. A test for homogeneous mixing is given.

Skalski and Robson (1982) developed a similar model with random n_i , but the probability of removal is assumed to be constant and equal to p . Let N be the size

of the initial population with D marked individuals. Defining $t_2 = \sum_{i=2}^s (i-1)n_i = \sum_{i=1}^{s+1} x_i$, then, from Sect. 10.1.1, with $x_i = \sum_{j=1}^{i-1} n_j$ and $x_{s+1} = r$, the total number removed, the removal component is

$$L_1 = L(\{n_i\} \mid N) = \frac{N!}{\prod_{i=1}^s n_i!(N-r)!} p^r q^{sN - \sum_{i=1}^{s+1} x_i}.$$

The authors found the distribution of the n_i conditional on the sufficient statistics, which can be used for sequential goodness-of-fit tests.

The likelihood for the combined model is

$$L = L_1 L_\tau,$$

and the authors discussed the computation of the maximum likelihood estimates and their asymptotic variances. We end up with three estimators, removal, Petersen, and combined estimators. If the underlying assumptions hold, we could expect the combined estimator to be the most efficient. However, introducing extra assumptions can be problematic. The authors provided hypothesis tests for the equality of the probability of capture of the marked and unmarked (i.e., homogeneity of the m_i and u_i sequences) and a test for the constancy of p . These need to be carried out if the combined estimator is to be used.

Gatz and Loar (1988) endeavored to provide a guide for testing the assumptions of both the Petersen and the removal methods and for making adjustments for violations of the assumptions. However, a major problem is that a substantial amount of data is needed to adequately test for the validity of each model and such data is not always available as numbers are frequently small when removals are incorporated. Van Poortena et al. (2017) commented that because of the difficulties in assessing the underlying assumptions, their findings suggested that “including marked fish in the removal process may unknowingly reduce accuracy and precision of initial abundance estimates and that this type of experimental design should be avoided in many instances.”

10.1.4 Regression Methods

When the probability of capture, p , is constant and x_i is the accumulated removal up to just before the i th removal, then (Hayne, 1949b)

$$\mathbb{E}[n_i \mid x_i] = p(N - x_i).$$

If the plot of n_i versus x_i is approximately linear, we have some qualitative evidence as to the stability of p . Since

$$\mathbb{E}[n_i] = Npq^{i-1},$$

and taking logarithms, we can also consider

$$\text{E}[y_i] \approx \log(pN) + (i - 1) \log q,$$

where $y_i = \log n_i$. Both models are weighted regression models because of unequal variances, e.g., using weights inversely proportional to $(\hat{N} - x_i)$. For further details, see Seber (1982, 325–326). Regression catch-per-unit-effort models are described in Sect. 10.4.

10.1.5 Variable Catchability

Variable catchability can commonly occur with removal methods, an example being size selectivity. Below, we shall consider three cases when there is variable catchability, where it is assumed that the j th individual ($j = 1, \dots, N$) has a constant probability of p_j of being caught in a sample and the p_j are a random sample from some density function $f(p; \phi)$ with parameter vector ϕ . Then, if n_i is the i th removal and $r = \sum_{i=1}^s n_i$,

$$f(\{n_i\}) = \frac{N!}{\prod_{i=1}^s n_i!(N-r)!} \left(\prod_{i=1}^s \pi_i^{n_i} \right) \pi_{s+1}^{N-r},$$

where, from Burnham and Overton (1978, 40),

$$\begin{aligned} \pi_i &= \text{E}[p(1-p)^{i-1}] = \int_0^1 p(1-p)^{i-1} f(p; \phi) dp \quad i = 1, 2, \dots, s, \\ \pi_{s+1} &= 1 - \sum_{i=1}^s \pi_i = 1 - \sum_{i=1}^s \text{E}[p(1-p)^{i-1}] = \int_0^1 (1-p)^s f(p; \phi) dp. \end{aligned}$$

We note that this generalized removal model with $p_j = p_h$ and $n_i = u_i$ is the one used for model M_{bh} discussed in Sect. 12.5.7. The methods used there apply here when the emphasis is on estimating N .

Van Poortena et al. (2017) investigated the robustness of five removal models which endeavored to incorporate fish behavior. They mentioned that in the past, it was assumed that variation around the mean declines in the catch per unit effort (CPUE, discussed in the next chapter) was caused by random variation. However, subsequent work has demonstrated that there can often be transitory or persistent changes in capture probability, leading to substantial bias ranging between 30 % and 50% in abundance estimates. A variety of fish behavior patterns can possibly lead to changes in the capture probability. For example, if the most catchable fish are removed first, the average probability of capture will decline causing a negative bias in the population estimate. Alternatively, if catchability is linked to abundance,

namely, the probability of catching a fish increases with less fish as with schooling populations, then the average capture probability will increase causing a positive bias in the population estimate.

The authors included marked fish in the removal process, but as already mentioned above, they did not recommend including capture-recapture data because of the difficulties in assessing the underlying assumptions. Extensive mathematical details are briefly summarized in their paper.

10.2 Bayesian Methods

Random catchability can commonly occur with removal methods, an example being size selectivity, which suggests the use of prior probabilities. Carle and Strub (1978) introduced a beta prior distribution for p , the probability of capture, and then obtained a marginal probability distribution for N . Wyatt (2002) introduced a Bayesian model, which assumed constant p and used a two-stage scheme in which a number of sites are randomly selected from the sections of river followed by a sample of fish from each site selected. Mäntyniemi et al. (2005) extended this model to allow for randomly distributed individual catchabilities in the population, which we now consider below.

10.2.1 Random Catchability

Following Mäntyniemi et al. (2005), let N_0 be the initial size of the population. Let p_h be the probability of capture for an individual h ($h = 1, \dots, N_0$) with prior density function $f(p_h | \theta)$, where θ is assumed to be $(\mu, \sigma)^t$ and μ and σ are the respective mean and standard deviation. Now, conditionally on the catchability p_h , let z_h be a Bernoulli distributed indicator variable that takes the value 1 if individual h is captured in the first removal and value 0 otherwise. Then the probability of obtaining z_h given p_h is $p_h^{z_h}(1 - p_h)^{1-z_h}$, and we have the distribution

$$\begin{aligned} f(z_h | \mu, \sigma) &= \int_0^1 p_h^{z_h} (1 - p_h)^{1-z_h} f(p_h | \mu, \sigma^2) dp_h \\ &= \mu^{z_h} (1 - \mu)^{1-z_h}. \end{aligned} \tag{10.3}$$

It is assumed that the z_h are conditionally independent so that n_1 caught in the first sample is binomial, namely,

$$n_1 = \sum_{h=1}^N z_h \mid N, \mu \sim \text{Binomial}(N, \mu).$$

After the first removal, $N_0 - n_1 = N_1$ individuals still remain in the population. The distribution of catchabilities in the remaining population has also changed from the initial distribution $f(p_h | \mu, \sigma)$ because, typically, individuals with the highest catchabilities have been removed. Then, from (10.3), and using Bayes' formula,

$$\begin{aligned} f(p_h | z_h = 0, \mu, \sigma) &= \frac{\Pr(z_h = 0 | p_h) f(p_h | \mu, \sigma)}{\Pr(z_h = 0 | \mu, \sigma)} \\ &= \frac{p_h^0 (1 - p_h)^1 f(p_h | \mu, \sigma)}{(1 - \mu)} \\ &\propto f(p_h | \mu, \sigma) (1 - p_h). \end{aligned}$$

Using a similar argument, the distribution of catchability before the j th removal is proportional to $f(p_h | \mu, \sigma)(1 - p_h)^{j-1}$. The authors assumed a beta prior distribution $\text{Beta}(\alpha, \beta)$, namely,

$$f(p_h | \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} p_h^{\alpha-1} (1 - p_h)^{\beta-1},$$

with mean $\alpha/(\alpha + \beta)$. Multiplying the above distribution by $(1 - p_h)^{j-1}$ yields a $\text{Beta}(\alpha, \beta + j - 1)$ distribution, which has a mean catchability of $q_j = \alpha/(\alpha + \beta + j - 1)$. Because of the awkwardness in parameterizing a beta distribution in terms of its mean μ and standard deviation σ , the authors used a different parameterization, namely, $\eta = (\mu(1 - \mu) - \sigma^2)/\sigma^2 = \alpha + \beta$, which leads to $\alpha = \mu\eta$ and $\beta = (1 - \mu)\eta$. Thus, by assuming that the catchabilities are a sample from a beta distribution, and knowing that only the mean catchability q_j in the remaining population after $j - 1$ removals is needed to model the successive catches, we can write the model in the form

$$\begin{aligned} n_j | N_j, q_j &\sim \text{Binomial}(N_j, q_j) \tag{10.4} \\ N_j &= N_{j-1} - n_{j-1}, \quad n_0 = 0, \\ q_j &= \mu\eta/(\eta + j - 1), \quad j = 1, \dots, s. \end{aligned}$$

From the above equation, there are three parameters $\boldsymbol{\phi} = (N_0, \mu, \eta)'$, and the joint posterior probability density of these, given the data $\{n_i\}$, is

$$f(\boldsymbol{\phi} | \{n_i\}) = \frac{f(\{n_i\} | \boldsymbol{\phi}) f(\boldsymbol{\phi})}{\int f(\{n_i\} | \boldsymbol{\phi}) f(\boldsymbol{\phi}) d\boldsymbol{\phi}} \propto f(\{n_i\} | \boldsymbol{\phi}) f(\boldsymbol{\phi}) \tag{10.5}$$

The authors used vague priors for the posterior distributions of the three parameters. Because of the need for numerical integration for the denominator of the above expression, they used Markov chain Monte Carlo simulation to draw samples from

the posterior distributions of the model parameters using **WinBUGS** software. An example showed that if there is only vague prior information about the variation of catchability among individuals, a very high number of successive removals may be needed to correctly estimate the population size. They applied the method to a known population of juvenile brown trout. A special case of the above model with variable catchability and covariates was given by Bord et al. (2014) for *Ixodes ricinus* ticks using a hierarchical Bayesian model with a random effect.

Diffuse Priors

Bord et al. (2018) sounded a cautionary note about using diffuse priors with a Bayesian removal model as Bayes estimates based on default improper priors lead to improper posteriors or infinite estimates. Similarly, weakly informative priors give unstable estimators that are sensitive to the choice of hyperparameters. They provided some guidelines for the choice of the prior distribution and showed that the prior distribution must penalize large population sizes and/or small sampling rates. In particular, default priors when no information is available, such as reference priors, have to be banished since they lead to improper posteriors and infinite estimates whatever the data are. In the same way, Bayes estimates of the abundance from diffuse proper priors may be highly dependent on the choice of the hyperparameters and lead to unstable estimates. They provided four case studies.

Different Techniques

Bohrmann and Christman (2012) used a similar model to Bohrmann and Christman (2013) but with the two variables, namely, the number of removals and the technique used. An addition was the superscript t referring to the technique used for the removal/depletion (e.g., vessel, gear type, personnel, etc.). Their focus was on using Bayesian sample size methodology along with hierarchical modeling and presenting a method for estimating the efficiency of previously employed depletion techniques.

10.2.2 Use of Effort Information

A removal model is very appropriate in dealing with invasive species as we ultimately want to deplete the population. St. Clair et al. (2013) considered several models using removal effort to estimate animal abundance. Let p_{ij} be the removal probability for region i and period j , which is the same for each animal in region i , and let y_{ij} be the number of animals removed then. Let $y_i = \sum_{j=1}^J y_{ij}$ be the total number of animals removed in region i , and let N_i animals be in region i prior to

the start of the first removal period. Then, $\mathbf{y}_i = (y_{i1}, \dots, y_{iJ})'$ has a multinomial distribution

$$f(\mathbf{y}_i | N_i, \boldsymbol{\pi}_i) = \frac{N_i!}{(N_i - y_{i.})! \prod_{j=1}^J y_{ij}!} \prod_{j=1}^J \pi_{ij}^{y_{ij}} \left(1 - \sum_{j=1}^J \pi_{ij}\right)^{N_i - y_{i.}},$$

where

$$\pi_{ij} = p_{ij} \prod_{k=1}^{j-1} (1 - p_{ik})$$

is the probability that an animal is removed during, but not before, period j . It is also assumed that the \mathbf{y}_i 's are independent vectors, given the local abundance parameter N_i and the removal vector parameter $\boldsymbol{\pi}_i$. The goal is to estimate both the local abundance N_i and local removal parameters p_{ij} .

Let f_{ij} measure the removal or hunting effort expended in region i during period j . The authors considered two methods of modeling the removal probability p_{ij} as a function of effort f_{ij} . For method 1, let θ_i be the probability that an animal is removed with one unit of effort in region i . If we assume that effort is additive (i.e., each unit of effort acts independently and has the same probability of removing an animal), and θ_i is constant across all j periods, then

$$p_{ij} = 1 - (1 - \theta_i)^{f_{ij}}.$$

One way of modeling θ_i is to use $\text{logit}(\theta_i) = \eta_i$, where the η_i are independently distributed as $N(a_i, b_i)$. Alternatively, we could use a hierarchical specification with $a = \mu_\eta$ and $b = \sigma_\eta^2$, where the latter are given prior distributions. The above model was compared with a second model that uses

$$\text{logit}(p_{ij}) = \alpha_{0i} + \alpha_{1i} f_{ij},$$

where the $\boldsymbol{\alpha}_i = (\alpha_{0i}, \alpha_{1i})'$ have independent bivariate normal distributions.

To estimate local abundance N_i it was assumed that $N_i \sim \text{Poisson}(\lambda_i)$, where $\lambda_i = A_i \gamma_i$ is the mean animal abundance in region i with area A_i and γ_i is the average density of animals per unit area in region i . Let $\phi_i = \log(\gamma_i)$, where ϕ_i is $N(\mu_\phi, \sigma_\phi^2)$ and μ_ϕ and σ_ϕ^2 have prior distributions. Then the joint posterior for any of the models will have the form (using vectors for the parameter lists over i , e.g., $\mathbf{y} = \{\mathbf{y}_i; i = 1, \dots, I\}$)

$$f(\mathbf{N}, \boldsymbol{\Psi}, \boldsymbol{\phi}, \psi, \mu_\phi, \sigma_\phi^2 | \mathbf{y}, \mathbf{f})$$

$$\propto \left(\prod_{i=1}^I f(\mathbf{y}_i | N_i, \boldsymbol{\Psi}_i, \mathbf{f}_i) f(\boldsymbol{\Psi}_i | \psi) f(N_i | \phi_i) f(\phi_i | \mu_\phi, \sigma_\phi^2) \right) f(\psi) f(\mu_\phi) f(\sigma_\phi^2),$$

where Ψ_i represents each model's removal probability parameters (θ_i or α_i) and, if needed, with ψ representing the model's meta-level parameters. The multinomial probabilities π given above for the multinomial distribution are simply functions of these Ψ_i . Using posterior distributions, the method was applied to estimating turkey abundance. A marginalized version of the model was fitted using data augmentation (Appendix A.5). Their use of catch per unit effort is considered further in the next chapter.

Davis et al. (2016) considered such a method slightly more general than St. Clair et al. (2013) above and applied it to feral swine (*Sus scrofa*). Let $y_{hij} = 1$ with probability p_{ij} if the h th individual is removed from site i at period j and 0 otherwise, given the individual is present at site i , i.e., when $z_{hi} = 1$ with probability ψ_i (and 0 if not present). Here, ψ_i was given a uniform prior distribution. It was then assumed as before that

$$p_{ij} = 1 - (1 - \theta_i)^{f_{it}}.$$

The θ_i were then modeled using a vague prior from the beta distribution. The size of the population N_i at site i is $N_i = \sum_{t=1}^T z_{it}$. Data augmentation was used with w_i individuals for site i .

10.2.3 Aerial Sightings

Davis et al. (2016) developed a Bayesian hierarchical model to estimate abundance from removal data accounting for varying levels of effort and used simulations to assess the conditions under which reliable population estimates are obtained. They applied the model to estimate site-specific abundance of an invasive species, feral swine (*Sus scrofa*), using removal data from aerial sightings. Simulations showed that abundance estimates were generally accurate when effective removal rates (removal probability accounting for total effort) were above 0.40. We now briefly outline their model (using their notation).

Let $y_{ijk} = 1$ with probability p_{ij} that individual k ($k = 1, \dots, m_i$) on site i ($i = 1, \dots, n$) with pass (removal) j ($j = 1, \dots, J$) was removed and 0 otherwise. Also, let $z_{ik} = 1$ with probability ψ_j , with a uniform prior, if individual i is present on site j and 0 otherwise. Define θ_j to be the probability that an individual will be removed from site j with one unit of effort, and let $N_j = \sum_i z_{ij}$ be the number of individuals on site j . If f_{jk} is the amount of effort (here helicopter hours) for site j and pass k , then

$$p_{ij} = (1 - (1 - \theta_j))^{f_{ij}}.$$

They used a vague prior from the beta distribution to model the site-level removal probability θ_j . The various distributions are listed below (from the supplementary material from their paper).

$$y_{ijk} = \begin{cases} 0 & , z_{ik} = 0, \\ \text{Bernoulli}(p_{ij}) & , \sum_{l < j} y_{ilk} = 0 , z_{ik} = 1 \\ 0 & , \sum_{l < j} y_{ilk} > 0 , z_{ik} = 1 \end{cases}$$

$$z_{ik} \sim \text{Bernoulli}(\psi_i)$$

$$\psi_i \sim \text{Beta}(\alpha_\psi, \beta_\psi)$$

$$\theta_i \sim \text{Beta}(\alpha_\theta, \beta_\theta)$$

Using the square bracket notation $[a \mid b]$ for the density function for a given b , we have the final distribution using vectors for the parameter lists, namely,

$$[\mathbf{z}, \boldsymbol{\psi}, \boldsymbol{\theta} \mid \mathbf{y}, \mathbf{f}] \propto \prod_{i=1}^n \prod_{j=1}^J \left[\prod_{k=1}^{m_i} ([y_{ijk} \mid p_{ij}]^{z_{ik}} 1^{(1-z_{ik})} [z_{ik} \mid \psi_i]) [[\psi_i \mid \alpha_\psi, \beta_\psi] [\theta_j \mid \alpha_\theta, \beta_\theta]] \right].$$

The authors also gave the conditional posterior distributions of the parameters in their supplement.

10.2.4 Some Electrofishing Applications

Prevost et al. (2008) applied hierarchical Bayesian modeling to the estimation of the population of Atlantic salmon (*Salmo salar*) juveniles in the Oir River (France). The data set consisted of ten sampling sites sampled by one or two removals (passes) over a period of 20 years (1986–2005). The sites were selected from the river stretch of interest, and then the fish were sampled from each site by the successive removal method. Each site was a section of the river associated with a unique habitat type, i.e., rapid-riffle or run, and each sampling unit was identified by three indices i (year), h (habitat type), and k which stands for repetition per stratum (ih). The two removals are, respectively, denoted by $n_{1(ihk)}$ and $n_{2(ihk)}$ for year i , habitat h , and repetition k from an initial population N_{ihk} size and the probability of capture p_{ihk} . We then have

$$n_{1(ihk)} \mid N_{ihk}, p_{ihk} \sim \text{Binomial}(N_{ihk}, p_{ihk})$$

$$n_{2(ihk)} \mid n_{1(ihk)}, N_{ihk}, p_{ihk} \sim \text{Binomial}(N_{ihk} - n_{1(ihk)}, p_{ihk}).$$

The population size N_{ihk} depends on the expected fish density d_{ihk} (fish per square meter) and the surface area S_{ihk} of a sampling site. It was assumed that $N_{ihk} \mid \lambda_{ihk} \sim \text{Poisson}(\lambda_{ihk})$, where $\lambda_{ihk} = d_{ihk} S_{ihk}$. Habitat and time covariates were introduced giving us

$$\text{E}[\log(d_{hi})] = \mu_d + \alpha_{di} + \beta_{dh},$$

and we have the following independent variables:

$$\log(d_{ihk}) \mid \mu_d, \alpha_{di}, \beta_{dh}, \sigma_d \sim N(\text{E}[\log(d_{ih})], \sigma_d^2),$$

with independent random effects

$$\alpha_{di} \mid \sigma_{\alpha_d} \sim N(0, \sigma_{\alpha_d}^2).$$

Also, we have the independent and identically distributed

$$\text{logit}(p_{ihk} \mid \mu_p, \sigma_p \sim N(\mu_p, \sigma_p^2)).$$

Further models and prior distributions were considered by the authors, along with model fitting and model checking. Some goodness-of-fit tests were provided, and assumptions were carefully discussed.

Laplanche (2010) developed a similar type of model that has multiple locations for estimating fish abundance in alpine streams and uses several removals per site but without an index for the habitat type, i.e., i for site and j for removal number. Again, abundance was assumed to be Poisson, and gamma and beta priors were used for other parameters. Eleven model reductions were considered. Ruiz and Laplanche (2010) focused on a hierarchical model that accounted for (i) variability of the abundance with fish length (as a distribution mixture), (ii) spatial variability of the abundance, (iii) variability of the catchability with fish length (as a logit regression model), (iv) spatial variability of the catchability, and (v) residual variability of the catchability with fish.

Harris et al. (2016) carried out a rare removal experiment where the population of larval lampreys in a wadeable stream was of known size in experimental enclosures using a Bayesian approach. With five removals, they found that a hierarchical removal model where the probability of capture p has a distribution or modeled using covariates was an effective method for estimating abundance.

10.3 Two and Three Removals

Since removal methods are often limited by the number of removals s that can be carried out at one site, we now consider the special cases of three and two removals in some detail. One removal is discussed briefly in Sect. 10.4.

10.3.1 Three Removals

As mentioned previously, in addition to small mammals, the removal method has also been used for small fish populations, say in brooks and rivers, using electrofishing. When $s > 1$, the term used is “multipass electrofishing” and is discussed further in Sect. 10.4. Under certain conditions, the variance of \hat{N} decreases only slightly as s increases beyond 3 so that the case of three removals is of some interest as it gives explicit solutions to the above equations for \hat{N} and \hat{p} (cf. Seber, 1982, 315), namely,

$$\hat{N} = \frac{6X^2 - 3XY - Y^2 + Y(Y^2 + 6XY - 3X^2)^{1/2}}{18(X - Y)}$$

and

$$\hat{p} = \frac{3X - Y - (Y^2 + 6XY - 3X^2)^{1/2}}{2X},$$

where $X = 2n_1 + n_2$ and $Y = r = n_1 + n_2 + n_3$; we note that the experiment “fails” if $n_2 > n_1$, for then \hat{N} is negative. One of the problems with the removal method is the possible variation in catchability. Seber and Whale (1970) (see also Seber, 2008: 316–7) carried out an analysis for this situation. It is assumed that u_j ($j = 1, 2, \dots, N$) is the probability that the j th member of the population is caught in a sample, given that it is in the population. This probability is assumed to remain constant from sample to sample, and individuals are independent. The joint distribution of the n_i is then found to depend on the k th moments a_k of the distribution of the u_j , namely,

$$f(n_1, n_2, n_3) = \frac{N!}{\left(\prod_{i=1}^3 n_i!\right)(N - Y)!} a_1^{n_1} (a_1 - a_2)^{n_2} (a_1 - 2a_2 + a_3)^{n_3} \\ \times (1 - 3a_1 + 3a_2 - a_3)^{N-Y},$$

so that the effect of catchability can be modeled using a variety of density functions $f(u)$. For example, if $f(u)$ is uniform on $[c, d]$ ($0 \leq c \leq d \leq 1$) and $w = c/d$, then $a_1 = d(1+w)/2$, $a_2 = d^2(1+w+w^2)/3$, and $a_3 = d^3(1+w)(1+w^2)/4$. Using the above joint density function for the n_i , the asymptotic mean of \hat{N} , obtained by replacing X and Y by their expected values, is given by BN , where

$$B = \frac{3(1+w)^2}{4(1+w+w^2)}.$$

Seber has a table of B for different values of w and d . It transpires that when $w = 1$, $B = 1$. When $w < 1$, $B < 1$, and B is sensitive to the value of w . When $w \geq 0.5$, the

asymptotic bias is negligible, and when $c > 0.01$, $B > 0.8$. If $f(u)$ has a unimodal (e.g., beta type) distribution defined on $[c, d]$, then the corresponding values of B will be larger indicating, asymptotically, that \hat{N} is fairly insensitive to variation in catchability. However, if there is considerable variation in catchability, the more catchable ones will be captured first, so that the average probability of capture will decrease from one trapping to the next, and \hat{N} will underestimate N .

10.3.2 Two Removals

Seber (1982, 318–323) pointed out that in some situations, it may be unwise or impracticable to take more than just two samples. For example, the weather may have a big effect on p so that a short intensive survey may be needed if p is to remain constant. If p is large, three samples may severely deplete the population unless tagging can be used to represent “removal.” On the other hand, if one wishes to reduce the population as much as possible, then for large p , it may be a waste of time or uneconomic to take a third sample. For example, in sorting animals out of samples of mud, soil, or grain, p the probability of being found will be large so that a third sorting may be a waste of effort. Seber (1970) and Seber and Le Cren (1967) considered this case $s = 2$ in detail. When $n_1 > n_2$, the maximum likelihood (and moment) estimates are

$$\hat{N} = n_1^2 / (n_1 - n_2)$$

and

$$\hat{q} = n_2 / n_1,$$

with large sample means and variances

$$\begin{aligned}\text{E}[\hat{N}] &\approx N + \frac{q(1+q)}{p^3}, \\ \text{V}[\hat{N}] &\approx \frac{Nq^2(1+q)}{p^3}, \\ \text{E}[\hat{p}] &\approx p - \frac{q}{Np},\end{aligned}$$

and

$$\text{V}[\hat{p}] \approx \frac{q(1+q)}{Np}.$$

From the relative bias terms included above, the formula for \hat{p} is roughly of the order of $(Np)^{-1}$, while that for \hat{N} is roughly to the order of $(Np)^{-3}$ so that the asymptotic bias and variance of \hat{N} may not be valid for quite reasonable values of N and p . By considering a coefficient of variation of no more than 25%, Seber and Le Cren (1967) suggested as a rough rule that the approximations are usually satisfactory if $Np^3 > 16q^2(1 + q)$. Seber and Whale (1970) gave approximately unbiased estimates of p and $(N + 1)^{-1}$, namely,

$$p^* = 1 - [n_2/(n_1 + 1)]$$

and

$$(n_1 - n - 2)/[(n_1 + 1)(n_1 + 2)],$$

respectively, for using in the construction of confidence intervals. A catchability investigation along the line of the three-sample case considered above revealed that the validity of the estimates can be sensitive to the catchability distribution. Harding et al. (1984) took a closer look at the above method and showed that the asymptotic approximations used for confidence intervals can fail under conditions commonly arising in practice. The same thing happens if the lower limit of the interval falls below $n_1 + n_2$. As is the case in other areas of statistics, likelihood intervals are preferred when n_1 , n_2 , and N are sufficiently small.

The two-removal model has been used for counting whales where two observation places (perches) are used (George et al., 2004). Once a whale is seen at the first perch, that whale is then regarded as “removed.” The authors added the additional feature of fitting a linear model with the log-log link function for the k th data point ($k = 1, \dots, K$), as there were several data points from the season, namely,

$$\log(-\log \hat{p}_k) = \mathbf{x}'_k \boldsymbol{\beta},$$

where covariates such as visibility, number of observers, and offshore distance can be used and $\boldsymbol{\beta}$ can be estimated. More recently, acoustic and visual location data were combined to estimate the number of whales passing Point Barrow, Alaska. If an estimate \tilde{p} of p is available from a two-sample removal or other data, it may be possible to apply this estimate to a different one-sample situation giving an estimate of N as $\tilde{N} = n_1/\tilde{p}$. For further discussion on the two-sample method, see Seber and Le Cren (1967) and Seber (1982).

10.4 Electrofishing

With rivers and small streams, a common removal method is to use electrofishing. Here, a portion of the waterway is blocked off, and the technique uses direct current electricity at high voltage flowing between a submerged cathode and anode to

temporarily stun fish so they can be caught. The equipment is portable and can be carried as a backpack, which is convenient for awkward terrain. Electrofishing can also be carried out by boat where the boat itself is the cathode and the anode(s) is generally mounted off the bow. The stunned fish swim toward the anode, where they are caught alive using a dip net.

The probability p by electrofishing will depend on a number of factors such as fish size and habitat characteristics such as the cross-sectional area of the stream, the amount of undercut banks' size, and the amount of cobble sub-substrata. The electrofishing crew can also have an effect when estimating abundance (Benejam et al., 2011). The probability p can be modeled using logistic regression, and typical covariates such as sampling gear, fish size, and stream habitat can be used (Dauwalter & Fisher, 2007). The closure assumption can be achieved by using block nets, although violation of this assumption in small streams is often minor. As electrofishing is a stressful capture technique, fish that evade capture on the first removal (depletion) seek cover and become less vulnerable to capture on subsequent removals. It is important to allow fish sufficient time (e.g., several days) to recover between removals.

Peterson et al. (2004) evaluated how effective were multipass (multisample) removal estimates of fish abundance by comparing the probability of capture (capture efficiency) for two species of trout with the efficiencies measured by the recapture of known numbers of marked individuals. As with all recaptures, there is the possible effect of handling and tagging. Capture efficiency measured by the recapture of marked fish was low for the first electrofishing pass (mean, 28%) and decreased considerably with successive passes. On average, the removal methods overestimated three-pass capture efficiency by 39% and underestimated fish abundance by 88%, across both species and all size classes. The study was carried out using backpack electrofishing in 43 first-through third-order streams located primarily in national forests and Bureau of Land Management lands in central Idaho and southwest Montana.

Two methods were used, the constant probability model described above and the generalized removal model M_{bh} (Sect. 12.5.7). Linear regression was used to determine the effect of various regressors including class size, habitat characteristics, and the number of removal passes. There was also an evaluation of fish escape from blocked off area as the population is assumed to be closed during sampling. Further details with regard to the model used including confidence intervals are given by the authors. The general conclusion was that most removal-type estimators from electrofishing were likely to be biased and that these biases are related to stream characteristics, fish species, and size.

A similar study was carried out by Carrier et al. (2009) using a dual-gear approach, which uses different sampling techniques for marking and recapturing fish. This ensures that individuals biased (either positively or negatively) toward the first capture gear are represented in the recapture sample independent of any initial bias. The authors used minnow trapping, marking and releasing juvenile salmonids in a stop-netted reach, followed by electrofishing the reach 1 hour after the release of marked fish. Several different experiments were carried out to test the assumption of

equal catchability of marked and unmarked fish and to compare population estimates and capture efficiencies from mark-recapture and depletion.

Meyer and High (2011) investigated the accuracy of removal electrofishing estimates of trout abundance in Rocky Mountain streams using four removals. Extensive practical details are given, and the MicroFish software package (Van Deventer & Platts, 1989) was used for estimation for the case of constant probability of capture. Capture-recapture was also used to determine the extent of bias in the estimates. They concluded that “multiple-pass removal sampling in typical Rocky Mountain streams can produce population estimates that are consistently but not drastically biased and are, therefore, probably adequate for most basic fish population monitoring even without correction, especially if electrofishing settings and crew training balance the need to minimize injury with effective fish sampling.”

Rosenberger and Dunham (2005) carried out an extensive study of the assumptions underlying both the removal and Petersen capture-recapture methods. They found that both methods gave misleading estimates if their assumptions were violated. This was particularly the case with the removal method due to the decreasing values of p , the probability of capture (called sampling efficiency by the authors). The term “mark-recapture sampling efficiency” for the estimated sampling efficiency was based on mark-recapture estimates divided by the total number of captured fish. They obtained separate removal estimates of fish abundance based on two, three, or all four passes (removals). For four-pass removal estimates only, they used the generalized removal model M_{bh} that can account for heterogeneity in sampling efficiency among passes detected by use of a goodness-of-fit test. Covariates were also incorporated using some logistic models.

Baumgartner (2004) compared the Petersen estimate with a removal jackknife estimate (Pollock & Otto, 1983, see Sect. 12.6.2) at 3-month intervals to determine the composition and sizes of migratory fish populations. Temporal changes in the population sizes of seven migratory species were assessed over a 2-year period prior to the construction of a fishway on the Murrumbidgee River, Australia. No difference in population size estimates was detected between the two methods. However, the jackknife removal estimator used to cope with heterogeneity of capture generally provided less variable estimates for a greater number of species. The number of removals is important, and the author used five.

Single Removal

Foley et al. (2015) investigated the effectiveness of the one-sample (one-pass) approach, as it can lead to a wider coverage of a stream using, say, an electrofishing backpack. Although a greater number of removals are more efficient with regard to estimation methods, they are costly and time-intensive, limiting the spatial extent of investigation. For those aiming to investigate whole-watershed patterns of fish distribution and abundance, this requires trading high-effort and site-specific sampling with low-effort, extensive sampling of entire stream lengths. Frequently, the method is conducted in a continuous manner, sampling every meter of stream between predetermined points. Although labor-intensive, it is ideal for sampling fish in headwater streams. Here, fish presence may be limited to a few kilometers

of stream and to where representative reach approaches fail to capture variability in fish distribution or the importance of habitat features present throughout entire riverscapes. Meador et al. (2011) assessed the efficacy of single-pass backpack electrofishing using two passes to estimate the number of species (species richness) and found that “the removal model failed 48% of the time because the number of individuals of a species was greater in the second pass than in the first pass.”

10.5 Design of a Removal Experiment

Bohrmann and Christman (2013) considered the problem of designing a removal (depletion) survey with imperfect capture or detection rates, which has had little attention. Several sites are used. Sites at which multiple removal passes are conducted (called multiple-pass sites) may be used to primarily provide information about detectability, but these constitute only a portion of a larger designed survey. Beginning with I multipass sites, each with r removals (the authors used p), it is assumed that a fixed amount of effort is used for each removal sample giving a total sampling effort of $e = Ir$. The probability of capture in a sample is assumed to be constant and equal to p (the authors used q).

Let $\mathbf{y}_i = (y_{i1}, \dots, y_{ir})'$ denote the removal numbers for the i th site ($i = 1, \dots, I$). Let N_i be the initial size of the population at site i and N_{ij} be the number available to be caught for the j th removal at the site. Assuming individuals are caught independent of other individuals, we have

$$y_{ij} | N_{ij}, p \sim \text{Binomial}(N_{ij}, p). \quad (10.6)$$

Also, $N_{i1} = N_i$ and $N_{ij} = N_i - \sum_{k=1}^{j-1} y_{ik}$ for $j = 2, \dots, r$. The elements of \mathbf{y}_i together with $y_{i,r+1}$, the number not caught (and unobservable), have a multinomial distribution, conditional on parameter N_i , with index N_i and cell probabilities $\mathbf{f}(p) = (f_1(p), \dots, f_r(p), f_{r+1}(p))'$ with $f_k(p) = p(1-p)^{k-1}$ for $k = 1, \dots, r$ and $f_{r+1}(p) = (1-p)^r$. Because of the unknown N_i and p , a more useful modification is the multinomial model conditional on $T_i = \sum_{k=1}^r y_{ik}$ with parameter T_i and cell probabilities given by $\mathbf{f}^*(p)$, which has elements

$$f_k(p)^* = \frac{f_k(p)}{\sum_{m=1}^r f_m(p)} = \frac{p(1-p)^{k-1}}{\sum_{m=1}^r p(1-p)^{m-1}} = \frac{p(1-p)^{k-1}}{1 - (1-p)^r}$$

for $k = 1, \dots, r$. Here, for $i = 1, \dots, r$,

$$T_i | N_i, p \sim \text{Binomial}(N_i, (1 - (1-p)^r)).$$

Bohrmann and Christman (2013) then showed in an extensive appendix that the Fisher information for p is

$$\begin{aligned} \mathcal{I}(p) = & \left\{ \frac{(1 - (1 - p)^r)}{rp^2} \right. \\ & + \left(\frac{p(1 - p) + 2p(1 - p)^2 + \dots + (r - 1)p(1 - p)^{r-1}}{r(1 - p)^2} \right) \\ & \left. - \left(\frac{r(r - 1)(1 - p)^{r-2}(1 - (1 - p)^r) + r^2(1 - p)^{2r-2}}{r(1 - (1 - p)^r)} \right) \right\} (e\bar{N}), \end{aligned}$$

where $\bar{N} = \sum_{i=1}^I N_i/I$ is the average number of animals at the sample sites. Excluding the term $e\bar{N}$, the above expression for p and r for a given level of effort e can be maximized over r . That expression can be calculated over a fine grid of p values and for integer values of r , and maxima for each value of p obtained. That is, one obtains the number of removals (depletion passes) r maximizing information gain about the catchability parameter p , for given values of p and e fixed. Once this was done, simulation of 60 depletion sites allowed the authors to explore allocations of 120 total removals from 2 removals at 30 sites to 6 removals at 20 sites, for each of the 5 values of p . Under model (10.1), and using a uniform prior [0,1] for p along with a uniform prior on N_i of $[T_i, T_i + 200]$, they catalogued the Bayesian point estimates of p for each simulation, along with the estimated standard deviations of these estimates, which they used for comparison among the various effort allocations.

10.6 Removal Methods for Subpopulations

Dorazio et al. (2005) described a method of estimating the abundance of spatially distinct subpopulations surveyed using removal sampling. It was assumed that exchanges were not permitted during the survey, though the restriction was not permanent. Once sampling was completed, animals could move freely among different subpopulations without consequences to abundance estimators. Using their model, they found that their abundance estimates have similar or better precision than those computed using the conventional approach of analyzing the removal counts of each subpopulation separately. Hierarchical models were developed using the Poisson and negative binomial distributions to model variation in abundance among subpopulation and using the beta distribution to model variation in capture probabilities. We now describe their models, which are motivated by the need to monitor changes in abundance of an endangered fish species, the Okaloosa darter.

Suppose removal samples are taken from each of I spatially distinct and closed subpopulations that constitute a representative sample of the entire population and we treat the subpopulations separately. Let x_{ij} denote the number of animals observed in the j th removal ($j = 1, \dots, J_i$) taken from the i th subpopulation ($i = 1, \dots, I$). For a given i , it is assumed that $x_{i1}, x_{i2}, \dots, x_{iJ_i}$, denoted by \mathbf{x}_i , have a multinomial distribution with index N_i and overall cell probabilities $\{p_{ij}\}$

or \mathbf{p} . Let θ_i denote the probability of capture during a single removal from the i th subpopulation. Then $p_{ij} = \theta_i(1 - \theta_i)^{j-1}$ is the probability of capture of animals during the j th removal. The probability distribution for \mathbf{x}_i is

$$f(\mathbf{x}_i | N_i, \theta_i) = \frac{N_i!}{(\prod_{j=1}^{J_i} x_{ij}!)(N_i - x_{i\cdot})!} (1 - \theta_i)^{J_i(N_i - x_{i\cdot})} \prod_{j=1}^{J_i} [(\theta_i)(1 - \theta_i)^{j-1}]^{x_{ij}}, \quad (10.7)$$

where $x_{i\cdot} = \sum_{j=1}^{J_i} x_{ij}$ is the total number of animals removed from subpopulation i . Differentiating the likelihood with respect to θ_i gives us

$$\theta_i = \frac{x_{i\cdot}}{J_i(N_i - x_{i\cdot}) + \sum_{j=1}^{J_i} j x_{ij}}, \quad (10.8)$$

and substituting this θ_i back into (10.7) yields a profile likelihood function for N_i that may be evaluated for admissible values of N_i ($= x_{i\cdot}, x_{i\cdot} + 1, \dots$) to determine the maximum likelihood estimate \hat{N}_i . Substituting this estimate of N_i into (10.8) gives the maximum likelihood estimate $\hat{\theta}_i$ of θ_i .

To combine the I subpopulations, we begin by assuming that the N_i are random variables from a discrete probability function $g(N_i | \boldsymbol{\phi})$, where g could be the Poisson or negative binomial probability distribution and $\boldsymbol{\phi}$ are one or more distributional parameters. Using (10.7), we have

$$f(\mathbf{x}_i | \boldsymbol{\phi}, \theta_i) = \sum_{N_i=x_{i\cdot}}^{\infty} f(\mathbf{x}_i | N_i, \theta_i) g(N_i | \boldsymbol{\phi}). \quad (10.9)$$

Assuming observations from different subpopulations are mutually independent, we have the likelihood

$$L(\boldsymbol{\phi}, \{\theta_i\} | \mathbf{X}) = \prod_{i=1}^I f(\mathbf{x}_i | \boldsymbol{\phi}, \theta_i), \quad (10.10)$$

where $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_I)'$ denotes the matrix of observed numbers of animals removed during the survey. The above likelihood can be used to obtain maximum likelihood estimates $\hat{\boldsymbol{\phi}}$ and $\hat{\theta}_i$. Here, g to be specified can be regarded as a prior mixing distribution for N_i , our main parameter of interest. Using Bayes, we can then obtain the following posterior density for any admissible value of N_i , conditional on $\hat{\boldsymbol{\phi}}$ and $\hat{\theta}_i$, namely,

$$\pi(N_i | \mathbf{X}, \hat{\boldsymbol{\phi}}, \hat{\theta}_i) = \frac{f(\mathbf{x}_i | N_i, \hat{\theta}_i) g(N_i | \hat{\boldsymbol{\phi}})}{f(\mathbf{x}_i | \hat{\boldsymbol{\phi}}, \hat{\theta}_i)}.$$

The authors recommended using the above equation to compute a bootstrap approximation to the marginal posterior density $\pi(N_i \mid \mathbf{X})$, which integrates over uncertainty in the model parameters.

The model (10.9) can be extended to include variation in capture probabilities as follows:

$$f(\mathbf{x}_i \mid \boldsymbol{\phi}, \eta) = \sum_{N_i=x_i}^{\infty} \left[\int_0^1 f(\mathbf{x}_i \mid N_i, \theta_i) h(\theta_i \mid \boldsymbol{\eta}) d\theta_i \right] g(N_i \mid \boldsymbol{\phi}),$$

where h needs to be specified, leading to

$$L(\boldsymbol{\phi}, \boldsymbol{\eta} \mid \mathbf{X}) = \prod_{i=1}^I f(\mathbf{x}_i \mid \boldsymbol{\phi}, \boldsymbol{\eta}).$$

This equation can be used to obtain maximum likelihood estimates $\hat{\boldsymbol{\phi}}$ and $\hat{\boldsymbol{\eta}}$. Empirical Bayes estimates for N_i can be computed using bootstrapping from

$$\pi(N_i \mid \mathbf{X}, \hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\eta}}) = \frac{g(N_i \mid \hat{\boldsymbol{\phi}}) \int_0^1 f(\mathbf{x}_i \mid N_i, \theta_i) h(\theta_i \mid \hat{\boldsymbol{\eta}}) d\theta_i}{f(\mathbf{x}_i \mid \hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\eta}})}.$$

We now consider possible default choices for $g(N_i \mid \boldsymbol{\phi})$ and $h(\theta_i \mid \boldsymbol{\eta})$, unspecified above. The first is a multinomial-Poisson mixture using $N_i \sim \text{Poisson}(\lambda)$. Using (10.9), we get the closed-form Poisson distribution

$$f(\mathbf{x}_i \mid \lambda, \theta_i) = \prod_{j=1}^{J_i} \frac{\exp[-\lambda \theta_i (1 - \theta_i)^{j-1}]}{x_{ij}!} [\lambda \theta_i (1 - \theta_i)^{j-1}]^{x_{ij}}, \quad (10.11)$$

with likelihood function $\prod_{i=1}^I f(\mathbf{x}_i \mid \lambda, \theta_i)$. Typical transformations are $\log(\lambda) = \alpha$ and $\log(\theta_i(1 - \theta_i)) = \beta_i$ to keep the parameters within their admissible ranges, or with covariates, $\log(\lambda_i) = \mathbf{w}'_i \boldsymbol{\alpha}$ and $\log(\theta_{ij}/(1 - \theta_{ij})) = \mathbf{v}'_{ij} \boldsymbol{\beta}$. The validity of the model can be tested using deviance statistics, as described by the authors.

Spatial randomness does not usually hold for most animal populations so that the Poisson distribution may not be appropriate. An alternative model is a multinomial-negative binomial mixture, which uses the negative binomial

$$g(N_i \mid \lambda, \varepsilon) = \frac{\Gamma(N_i + \varepsilon^{-1})}{N_i! \Gamma(\varepsilon^{-1})} \left[\frac{\varepsilon \lambda}{1 + \varepsilon \lambda} \right]^{N_i} \left[\frac{1}{1 + \varepsilon \lambda} \right]^{\varepsilon^{-1}},$$

for which $E[N_i] = \lambda$ and $\text{var}[N_i] = \lambda + \varepsilon \lambda^2$. Here, the overdispersion parameter ε is positive or tending to 0 for the Poisson distribution. Again, deviance statistics can be used.

We next consider a model for the density function h by extending the multinomial-Poisson mixture to the multinomial-Poisson-beta mixture. Here, we model the spatial distribution of capture probabilities using the beta distribution

$$h(\theta_i \mid \mu, \tau) = \frac{\Gamma(r)}{\Gamma(\tau\mu)\Gamma(\tau(1-\mu))} \theta_i^{\tau\mu-1} (1-\theta_i)^{\tau(1-\mu)-1},$$

so that $E[\theta_i \mid \mu, \tau] = \mu$, and $\text{var}[\theta_i \mid \mu, \tau] = \mu(1-\mu)/(\tau+1)$. The above model then leads to

$$f(\mathbf{x}_i \mid \lambda, \mu, \tau) = \int_0^1 f(\mathbf{x}_i \mid \lambda, \theta_i) h(\theta_i \mid \mu, \tau) d\theta_i,$$

which requires numerical integration. Transformations $\log(\mu)/(1-\mu) = \beta$ and $\log(\tau) = \phi$ can be used, with the possibility of including covariates.

Alternatively, we can use the negative binomial instead of the Poisson distribution in the previous model but with the same mixing distribution h to get

$$\begin{aligned} q(\mathbf{x}_i \mid N_i, \mu, \tau) \\ = \int_0^1 f(\mathbf{x}_i \mid N_i, \theta_i) h(\theta_i \mid \mu, \tau) d\theta_i \\ = \frac{N_i!}{\prod_{j=1}^{J_i} x_{ij}!(N-x_{i\cdot})!} \frac{B\left(\tau\mu + x_{i\cdot}, \tau(1-\mu) + J_i(N_i - x_{i\cdot}) - x_{i\cdot} + \sum_{j=1}^{J_i} jx_{ij}\right)}{B(\tau\mu, \tau(1-\mu))} \end{aligned}$$

where $B(\cdot, \cdot)$ denotes the beta function. Then

$$f(\mathbf{x}_i \mid \lambda, \varepsilon, \mu, \tau) = \sum_{N_i=x_{i\cdot}}^{\infty} q(\mathbf{x}_i \mid N_i, \mu, \tau) g(N_i \mid \lambda, \varepsilon).$$

Maximizing the likelihood function obtained by multiplying the product of the above over i is computationally intensive but straightforward. Again, the adequacy of the model can be tested using deviance statistics. In applying the method to a real-life example and simulated examples, it was found that by combining the subpopulations, efficiencies were gained as each site's estimate of N_i through borrowed strength from the data observed from the other sites.

In conclusion, we note that the above will apply to multiple sites that are individually closed, which is now considered.

10.6.1 Multiple Sites

Haines (2020) extended the work on the multinomial N-mixture models described in the previous section and, in particular, extended the use of the negative binomial, but with a change of notation. Let y_{ij} be the count at the i th site (containing N_i individuals, $i = 1, \dots, s$) and j th removal, and define $y_{i\cdot} = \sum_j y_{ij}$. Let p_{ij} be the probability of capturing animal i in sample j , and let π_{ij} be the probability of animal on site i which is removed in sample j . Then $\pi_{ij} = \prod_{k=1}^{j-1} (1 - p_{ik}) p_{ij}$, and denote $\pi_{i\cdot} = \sum_{j=1}^J \pi_{ij}$. Now μ_i , the mean of N_i , can be expressed in terms of covariates, namely, $\log(\mu_i) = \mathbf{x}'_i \boldsymbol{\beta}$. Also, we can use $\text{logit}(p_{ij}) = \mathbf{z}'_{ij} \boldsymbol{\gamma}$. Using a negative binomial distribution for the counts with shape parameter r , Haines developed a general likelihood function $L(\boldsymbol{\beta}, \boldsymbol{\gamma}, r)$ consisting of the product of two likelihood functions. The first one is conditional on the site totals, namely,

$$L_C(\boldsymbol{\gamma}) = \prod_{i=1}^s \left(\frac{y_{i\cdot}!}{\prod_{j=1}^J y_{ij}!} \right) \prod_{j=1}^J \left(\frac{\pi_{ij}}{\pi_{i\cdot}} \right)^{y_{ij}},$$

and the second is the marginal likelihood

$$L_M(\boldsymbol{\beta}, \boldsymbol{\gamma}, r) = \prod_{i=1}^s \frac{\Gamma(r + y_{i\cdot})}{\Gamma(r)y_{i\cdot}!} \left(\frac{\mu_i \pi_{i\cdot}}{r + \mu_i \pi_{i\cdot}} \right)^{y_{i\cdot}} \left(\frac{r}{r + \mu_i \pi_{i\cdot}} \right)^r.$$

The overall likelihood is the product of the two, namely,

$$L(\boldsymbol{\beta}, \boldsymbol{\gamma}, r) = \prod_{i=1}^s \left[\frac{\Gamma(r + y_{i\cdot})}{\Gamma(r) \prod_{j=1}^J y_{ij}!} \prod_{j=1}^J \left(\frac{\mu_i \pi_{ij}}{r + \mu_i \pi_{i\cdot}} \right)^{y_{ij}} \left(\frac{r}{r + \mu_i \pi_{i\cdot}} \right)^r \right].$$

For general p_{ij} and no common parameters, there is parameter redundancy with L_C , so some restrictions are needed. The one mentioned above is $p_{ij} = p_i (= \theta_i)$. The reader is referred to Haines (2020) with regard to maximum likelihood estimation and computational details, along with information matrices for the parameters. Adjustments that can be made when there are some zero-site totals are mentioned. The Poisson mixing distribution is also discussed using the above framework, and tests of the Poisson versus the negative binomial distribution are considered briefly.

10.7 Removal Methods for Point-Count Surveys

A removal method can be used to directly estimate the detection probability during point-count surveys of birds (Sect. 7.3). All the birds heard or seen during a fixed amount of time at widely spaced count locations were recorded. As previously

discussed, the probability of a bird being recorded depends on (1) the probability the bird sings during the count and (2) the probability the bird is detected, given that it sings. Two methods for dealing with (2) were considered, firstly the distance method, where the probability of detection falls off with increasing distance, and secondly the double observation method where we model the probability that a bird is recorded given that it could be detected by at least one observer. However, neither of these approaches deals with the first component of the detection probability in (1).

In Sect. 7.9, we developed methods of estimating the product of the two probabilities, and here, we consider another method where distances are not measured but counts are separated into time intervals, as proposed by Farnsworth et al. (2002). They began with the simplest case of just two equal time intervals $(0, t)$ and $(t, 2t)$ and point j ($j = 1, \dots, s$, say) within a detection circle of radius W . An observer records all the birds seen or heard in the first time interval and then continues the point count, recording any additional birds detected in the second interval. Let n_1 be the first count, n_2 the additional count, and p_i the probability of detection for the i th count; then the first count can be regarded as a “removal.”

Let N be the size of the initial population. The authors recommended using more than two (possibly unequal) intervals so that the assumption of equal detectability used for a two-sample removal model can be relaxed. They proposed intervals of the first 3 min, the subsequent 2 min, and the last 5 min with n_i ($i = 1, 2, 3$) counted in the i th interval. Since we have too many parameters, the three p_i and N , but only three counts, we need to add a constraint. The authors divided the local population of N of birds into two groups. Group 1 is composed of the birds that are easily detected and are assumed to be all detected in the first interval, and group 2 includes those more difficult to detect. Let c be the probability that a randomly selected bird is a member of group 2 (and hence the expected proportion of the population in group 2), and let q be the probability of failing to detect a member of group 2 within 1 minute. Then, for the birds detected within the first time interval of 3 minutes,

$$\mathbb{E}[n_1] = N(1 - c) + Nc(1 - q^3) = N(1 - cq^3) = Np_1, \text{ say.}$$

This follows from the fact that all of the members of group 1 plus some from group 2 will be detected in that interval. The probability that a bird in group 2 will be missed during the first 3 min is q^3 , and the complement of this $(1 - q^3)$ is thus the probability of being detected at least once in the first interval. Similarly, the expected value of the number of birds initially detected within the next time interval (2 min) is

$$\mathbb{E}[n_2] = Ncq^3(1 - q^2) = Np_2, \text{ say.}$$

Here, the birds first recorded during the middle interval must be missed in the first 3 min and not missed in the subsequent 2 min. Thus, these are all members of group 2. Finally, the expected number of birds counted in the last 5 min of the count is

$$\mathbb{E}[n_3] = N_j cq^5(1 - q^5) = Np_3, \text{ say.}$$

This follows from the fact that for a bird to be first counted in the final interval, it must be a member of group 2, and it must be missed during the first 5 min but not missed during the last 5 min. If $n. = \sum_i n_i$, the expected total number of birds counted by the end of the full 10 min is then

$$E[n.] = N[c(1 - q^{10}) + (1 - c)] = N(1 - cq^{10}) = Np., \text{ say.}$$

The joint distribution of the n_i is the multinomial distribution

$$f(n_1, n_2, n_3 | N) = \frac{N!}{(\prod_i n_i!)(N - n.)!} \left(\prod_{i=1}^3 p_i^{n_i} \right) (1 - p.)^{N-n.}, \text{ say.}$$

Since N is unknown, we can condition on the sum $n.$ to get the likelihood

$$L(c, q | n.) \propto \prod_{i=1}^3 \pi_i^{n_i},$$

where $\pi_i = p_i/p.$ for all $i.$ Maximum likelihood estimates of $c, q,$ and ultimately $p. = 1 - cq^{10}$ can be found using the available software. Here, $p.$ is the combined probability of detecting a bird on the sampled area. Also, N can be estimated by $\hat{N} = n./\hat{p}.$ and population density by

$$\hat{D} = \hat{N}/(s\pi W^2).$$

Farnsworth et al. (2002) discussed the underlying assumptions in detail and suggested further extensions such as using more than one observer. The above approach is basically a generalization of the behavioral model M_b considered in Chap. 12.

10.8 Time-to-Detection Model

We now consider the time-to-detection method as it is essentially a removal method as later detections are ignored. Martin-Schwarze et al. (2017) considered a class of N-mixture models, with a focus on birds, that allows for detection heterogeneity over time through a flexibly defined time-to-detection distribution (TTDD), and allowed for fixed and random effects for both abundance and detection. Given S surveys, let N_s individuals be present for survey s ($s = 1, \dots, S$), of which $n_s^{(\text{obs})}$ are observed. Suppose that an observer could remain at the survey location until every individual is detected, recording the time-to-detection T_{sb} for each individual b (bird) ($b = 1, \dots, N_s$). It is assumed that detection times for all individuals at a survey are independent, identically distributed (i.i.d.) according to a common TTDD

for T_{sb} , the time to first detection, with cumulative distribution function (cdf) $F_T(t)$ and density function (pdf) $f_T(t)$. In practice, times to first detection are truncated due to a finite survey length of C , meaning each individual has detection probability $p^{(\text{det})} = F_T(C)$. The conditional distribution of observed detection times then has pdf

$$f_C(t \mid \text{det}) = f_T(t)/F_T(C), \quad 0 < t < C,$$

where $F_C(t \mid \text{det}) = \int_0^t f_C(x \mid \text{det})dx$, and instantaneous detection rate or hazard function, $h(t) = f_T(t)/[1 - F_T(t)]$. Various possible TTDDs for T are the following, where α is a shape parameter:

Exponential: $f(x; \varphi) = \varphi e^{-\varphi x}, \varphi > 0, E[x] = 1/\varphi.$

Gamma: $f(x; \alpha, \varphi) = \frac{\varphi^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\varphi x}. E[x] = \alpha/\varphi.$

Weibull: $f(x; \alpha, \varphi) = \alpha \varphi (\varphi x)^{\alpha-1} e^{-\varphi x}. E[x] = \Gamma(1 + 1/\alpha)/\varphi.$

Lognormal: Here, $\log x$ has a normal distribution so that

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\frac{(\log x - \mu)^2}{2\sigma^2}\right].$$

If $\sigma = \alpha$ and $\varphi = e^{-\mu}$, $E[x] = \exp[\alpha^2/2]/\varphi$.

One common strategy to deal with data that do not fit a constant detection assumption is to model the TTDD as a mixture of two distributions—a continuous-time distribution and a point mass for increased detection probability in the initial observation period. We now consider an N-mixture model and assume that the following variables are independently distributed with density functions:

$$n_s^{(\text{obs})} \sim \text{Binomial}(N_s, p^{(\text{det})}) \quad \text{and} \quad N_s \sim \text{Poisson}(\lambda).$$

Then N_s is split into observed and unobserved portions so that

$$n_s^{(\text{obs})} \sim \text{Poisson}(\lambda p^{(\text{det})}) \text{ and, independently, } n_s^{(\text{unobs})} \sim \text{Poisson}(\lambda[1 - p^{(\text{det})}]).$$

Covariates were introduced using

$$\log(\lambda_s) = \mathbf{x}'_{sA} \boldsymbol{\beta}_A + \mathbf{z}'_{sA} \boldsymbol{\xi}^A,$$

where \mathbf{x}_{sA} are covariates, $\boldsymbol{\beta}_A$ is a vector of fixed effects, \mathbf{z}_{sA} specifies random effect levels, and the random effects $\boldsymbol{\xi}_{jA}$ are i.i.d. $N(0, \sigma_{A[j]}^2)$, where $A[j]$ assigns the appropriate variance for the j th abundance random effect. We also have

$$\log(\varphi_s) = \mathbf{x}'_{sD} \boldsymbol{\beta}_D + \mathbf{z}'_{sD} \boldsymbol{\xi}_D,$$

where \mathbf{x}_{sD} are covariates, $\boldsymbol{\beta}_D$ is a vector of fixed effects, \mathbf{z}_{sD} specifies random effect levels, and the random effects ξ_{jD} are i.i.d. $N(0, \sigma_{D[j]}^2)$, where $D[j]$ assigns the appropriate variance for the j th detection random effect.

The next step is to incorporate the fact that the detection data are typically not recorded exactly, but are instead censored into I intervals. Letting C_i for $i = 1, \dots, I$ indicate the right endpoint of the i th interval, then C_I is the total survey duration, and, letting $C_0 = 0$, the i th interval is $(C_{i-1}, C_i]$. Let n_{si} be the number of individuals counted during interval i on survey s , so that $n_s^{(\text{obs})} = \sum_{i=1}^I n_{si}$, and define $\mathbf{n}_s = (n_{s1}, \dots, n_{sI})'$. Assuming independence among individuals and sites, the \mathbf{n}_s elements are independently distributed as the multinomial distribution with index $n_s^{(\text{obs})}$ and cell probabilities \mathbf{p}_s , where $\mathbf{p}_s = (p_{s1}, \dots, p_{sI})'$. Then

$$p_{si} = F_T(C_i \mid \text{det}) - F_T(C_{i-1} \mid \text{det}) = \int_{C_{i-1}}^{C_i} f_T(t \mid \text{det}) dt.$$

Since it is common for avian point counts to observe increased detections in the first interval, a TTDD can be defined with a mixture component to increase the probability of observing individuals in the first interval. A mixing parameter $\gamma \in [0, 1]$ is introduced giving a point mass during the first observation interval and a continuous-time detection distribution $F_{T(C)}(t)$ leading to

$$F_T(t) = (1 - \gamma) + \gamma F_{T(C)}(t), \quad t > 0.$$

Using Bayesian methods with prior distributions, Martin-Schwarze et al. (2017) conducted three simulation studies to explore the behavior of models with nonconstant TTDDs. The first study compared mixture vs non-mixture models. The second study compared the TTDD families. In the first two studies, they utilized intercept-only models to focus the attention on the robustness of the TTDD choice in the most simple of scenarios. For the third study, they included fixed and random effects for both abundance and detection and again compared the distribution families. There is an extensive comparison of the performance of the various models, and their paper should be consulted for details.

10.9 Change in Ratio: Two Classes

Change-in-ratio (CIR) methods for population size estimation are closely related to removal methods as they are based on removals taking place from the population. Here, the population is subdivided into distinct population classes, and the removals are taken from each class. The method can also be combined with capture-recapture data. Other information can sometimes be added to obtain estimates of other parameters (e.g., recruitment, harvest, and finite annual population growth from data on moose observations and harvest, Solberg et al., 2005). We first consider the case where we have one removal and two samples. Some of the earlier material is referred to by Seber (1982, 354–380), which we follow.

10.9.1 Theory

Suppose we have a closed population consisting of two types (classes) of animals, x -type and y -type, for example, males and females or sport fish and coarse fish, and known removals take place between times t_1 and t_2 . Suppose there is a differential change in the ratio of x -type to y -type animals between times t_1 and t_2 . Let

X_i = number of x -type animals in the population at time t_i

Y_i = number of y -type animals in the population at time t_i

$N_i = X_i + Y_i$ (total population size at time t_i)

$P_i = X_i/N_i (= 1 - Q_i)$

$R_x = X_1 - X_2$

$R_y = Y_1 - Y_2$

$R = R_x + R_y = N_1 - N_2$

$f = R_x/R$

Now,

$$P_2 = \frac{X_1 - R_x}{N_1 - R} = \frac{P_1 N_1 - R_x}{N_1 - R},$$

and

$$N_1 = \frac{R_x - RP_2}{P_1 - P_2}. \quad (10.12)$$

Given estimates \hat{P}_i ($i = 1, 2$), we have the estimates

$$\begin{aligned} \hat{N}_1 &= \frac{R_x - R\hat{P}_2}{\hat{P}_1 - \hat{P}_2} \\ \hat{X}_1 &= \hat{P}_1 \hat{N}_1 \\ \hat{N}_2 &= \hat{N}_1 - R. \end{aligned} \quad (10.13)$$

Using the delta method (Appendix), we have for independent estimates of the P_i ,

$$\text{var}[\hat{N}_1] \approx (P_1 - P_2)^{-2} \left\{ N_1^2 \text{var}[\hat{P}_1] + N_2^2 \text{var}[\hat{P}_2] \right\} \quad (10.14)$$

$$\text{var}[\hat{X}_1] \approx (P_1 - P_2)^{-2} \left\{ N_1^2 P_2^2 \text{var}[\hat{P}_1] + N_2^2 P_1^2 \text{var}[\hat{P}_2] \right\}. \quad (10.15)$$

We now consider some methods of obtaining the estimates \widehat{P}_i . Suppose a random sample of fixed size n_i animals is taken with replacement, at time t_i , and x_i animals are found to be of x -type. Then $\widehat{P}_i = x_i/n_i$ is an unbiased estimate of P_i with variance $\text{var}[\widehat{P}_i] = P_i Q_i/n_i$, and (10.14) reduces to

$$\text{var}[\widehat{N}_1] \approx (P_1 - P_2)^{-2} \left\{ \frac{X_1 Y_1}{n_1} + \frac{X_2 Y_2}{n_2} \right\}.$$

Note that the above theory still holds if the removals R_x and R_y are negative, that is, additions. Many of the later papers assume additions with removals that are negative, and this needs to be kept in mind. For example, Paulik and Robson (1969), in their excellent review of earlier material, define $R_x = X_2 - X_1$, etc., though they use the above notation when discussing the planning of CIR experiments. We shall use the term “removal” throughout, as in this chapter we are considering removal methods.

10.9.2 Binomial Model

For the case of binomial sampling, \widehat{N}_1 and \widehat{X}_1 were first obtained by Chapman (1954) who showed that they are maximum likelihood estimates for the product of binomial models (with $y_i = n_i - x_i$)

$$f(x_1, x_2 | \{X_i, Y_i, n_i\}) = \prod_{i=1}^2 \binom{n_i}{x_i} \left(\frac{X_i}{N_i} \right)^{x_i} \left(\frac{Y_i}{N_i} \right)^{y_i}.$$

The binomial model is applicable when, for example, animals are just observed, so that a particular animal may be seen more than once. The assumptions underlying this model are:

- (i) The population is closed.
- (ii) All animals have the same probability of being caught in the i th sample ($i = 1, 2$).
- (iii) The removals R_x and R_y are known exactly.

When sampling is without replacement, we have a product of hypergeometric functions

$$f(x_1, x_2 | \{X_i, Y_i, n_i\}) = \prod_{i=1}^2 \binom{X_i}{x_i} \binom{N_i - X_i}{n_i - x_i} / \binom{N_i}{n_i}.$$

The maximum likelihood estimates of N_1 and X_i are still the same, and asymptotic variances are given by (10.14) and (10.15) with

$$\text{var}[\widehat{P}_i] = \frac{P_i Q_i}{n_i} \cdot \frac{N_i - n_i}{N_i - 1}.$$

This can be estimated by

$$\widehat{\text{var}}[\widehat{P}_i] = \frac{\widehat{P}_i \widehat{Q}_i}{n_i - 1} \left(1 - \frac{n_i}{\widehat{N}_i}\right).$$

When the sampling fraction is small, we can use the binomial model as a good approximation.

Assumption (i)

The binomial model can still hold when there is mortality both between time t_1 and the removal time, and between the time of removal and time t_2 , provided that in each period, the survival rates are the same for both types of animal (i.e., the P_i remain constant). However, N_1 now refers to the population just prior to the removal.

Assumption (ii)

This assumption implies that (a) the probability of capture is the same for all animals of a given type and (b) the two types of animal have the same probability of capture in the sampling procedure. If care is exercised in the sampling procedure, (a) may be realistic, though (b) may not be true in some situations. For example, cock pheasants are more easily seen than hen pheasants. If the two types of animal refer to different species, they may be sampled at different rates or captured differently. Suppose that λ is the ratio of the probability of sampling y -type animals to the probability of sampling x -type animals, and let $d = 1 - \lambda$; then x_i is Binomial $(n_i, P_{i\lambda})$, where

$$P_{i\lambda} = \frac{X_i}{X_i + \lambda Y_i} = \frac{X_i}{N - d Y_i}.$$

The estimates are now biased and expressions are available for their asymptotic means (Seber, 1982, 358). If $R_y = 0$, \widehat{X}_1 is not only asymptotically unbiased but, by the delta method, the usual variance estimate (cf. 10.15 of Sect. 10.9.1 with parameters replaced by estimators) is an asymptotically unbiased estimate of its asymptotic variance

$$(P_{1\lambda} - P_{2\lambda})^{-2} \left\{ (X_1 + \lambda Y_1)^2 P_{2\lambda}^2 \text{var}[\widehat{P}_1] + (X_2 + \lambda Y_2)^2 P_{1\lambda}^2 \text{var}[\widehat{P}_2] \right\}.$$

Assumption (iii)

This assumption may be false as R_x and R_y could be underevaluated because of unknown natural mortality, unreported kills, and possible “crippling losses.” If the

total kills are determined by questionnaire, there can be all sorts of biases such as misreporting, nonresponse, and memory loss, to name a few (cf. (Seber, 1982, 360–361)). Suppose that the actual removals are $R_x + \Delta_x$ and $R_y + \Delta_y$, then from (10.12) in Sect. 10.9.1,

$$\begin{aligned} N_1 &= \frac{R_x + \Delta_x - P_2(R + \Delta_x + \Delta_y)}{P_1 - P_2} \\ &= \frac{R_x - P_2 R}{P_1 - P_2}(1 + b_1), \end{aligned}$$

where

$$b_1 = \frac{\Delta_x - P_2(\Delta_x + \Delta_y)}{R_x - P_2 R},$$

and asymptotically (large n_i)

$$\mathbb{E}[\widehat{N}_1 = N_1/(1 + b_1)].$$

Also,

$$\mathbb{E}[\widehat{X}_1] = P_1 \mathbb{E}[\widehat{N}_1] = X_1/(1 + b_1),$$

and asymptotically

$$\mathbb{E}[\widehat{N}_2] = N_2/(1 + b_2),$$

where

$$b_2 = \frac{\Delta_x - P_1(\Delta_x + \Delta_y)}{R_x - P_1 R}.$$

If it is assumed that the proportion k of unreported animals is the same for both classes, then $b_1 = b_2 = k$ and the estimates will be slightly underestimated (Chapman, 1955).

10.9.3 Survival Rates

Let x_i and y_i be the respective numbers of x -type and y -type animals taken in the sample at time t_i . Suppose that the removals are due to exploitation and possibly natural mortality. Let ϕ_x and ϕ_y be the fractions of x -type and y -type, respectively, surviving from time t_1 to time t_2 . Then, from Paulik and Robson (1969),

$$X_2 = X_1 \phi_x, \quad Y_2 = Y_1 \phi_y,$$

and hence,

$$\theta = \frac{\phi_x}{\phi_y} = \left(\frac{Y_1}{X_1} \right) / \left(\frac{Y_2}{X_2} \right) = \left(\frac{1 - P_1}{P_1} \right) / \left(\frac{1 - P_2}{P_2} \right).$$

Then, if \hat{P}_i is an estimate of P_i , θ is estimated by

$$\begin{aligned}\hat{\theta} &= \left(\frac{1 - \hat{P}_1}{\hat{P}_1} \right) \left(\frac{\hat{P}_2}{1 - \hat{P}_2} \right) \\ &= (y_1 x_2) / (y_2 x_1).\end{aligned}$$

Using the delta method,

$$\text{var}[\hat{\theta}] = [P_1(1 - P_2)]^{-4} \{(1 - P_2)^2 P_2^2 \text{var}[\hat{P}_1] + (1 - P_1)^2 P_1^2 \text{var}[\hat{P}_2]\}.$$

Paulik and Robson (1969) gave various variations of the above method when additional information is available. To separate exploitation and natural mortality, Chapman and Murphy (1965) suggested considering two cases: “instantaneous” removal, when the removal takes place in a short period of time, and “continuous” uniform removal, when the removal takes place uniformly over an extended period of time. For details of the two cases, see Seber (1982, 384–387).

10.9.4 Removals Estimated

Following on from the above discussion, consider the situation where the removals have to be estimated by independent methods. If the removals R_x and R_y have respective independent unbiased estimates \hat{R}_x and \hat{R}_y , then N_1 can be estimated by

$$\hat{N}_1 = \frac{\hat{R}_x - \hat{P}_2 \hat{R}}{\hat{P}_1 - \hat{P}_2},$$

where $\hat{R} = \hat{R}_x + \hat{R}_y$. Using the delta method,

$$\begin{aligned}\text{var}[\hat{N}_1] &\approx (P_1 - P_2)^{-2} \left\{ N_1^2 \text{var}[\hat{P}_1] + N_2^2 \text{var}[\hat{P}_2] \right. \\ &\quad \left. + (1 - P_2)^2 \text{var}[\hat{R}_x] + P_2^2 \text{var}[\hat{R}_y] \right\}.\end{aligned}$$

Alternatively, if we have estimates of R and $f_x = R_x/R$, then

$$\hat{N}_1 = \frac{\hat{R}(\hat{f}_x - \hat{P}_2)}{\hat{P}_1 - \hat{P}_2},$$

and

$$\begin{aligned}\text{var}[\widehat{N}_1] \approx & (P_1 - P_2)^{-2} \left\{ N_1^2 \text{var}[\widehat{P}_1] + N_2^2 \text{var}[\widehat{P}_2] \right. \\ & \left. + R^2 \text{var}[\widehat{f}_x] + (f_x - P_2)^2 \text{var}[\widehat{R}] + 2R(f_x - P_2) \text{cov}[\widehat{f}_x, \widehat{R}] \right\}.\end{aligned}$$

A similar expression is obtained for X_1 , namely,

$$\begin{aligned}\text{var}[\widehat{X}_1] \approx & (P_1 - P_2)^{-2} \left\{ N_1^2 P_2^2 \text{var}[\widehat{P}_1] + N_2^2 P_1^2 \text{var}[\widehat{P}_2] \right. \\ & \left. + P_1^2 R^2 \text{var}[\widehat{f}_x] + P_1^2 (f_x - P_2)^2 \text{var}[\widehat{R}] + 2RP_1^2 (f_x - P_2) \text{cov}[\widehat{f}_x, \widehat{R}] \right\}.\end{aligned}$$

The covariance terms will depend on the model used. If the estimates are independent, the above covariance terms are zero.

A possible scenario is when R is known but f has to be estimated by means of a sample from the R removals. In this case, $\text{var}[\widehat{R}] = 0$, and \widehat{f} will usually be a binomial proportion.

10.9.5 Optimum Allocation

If N_1 is to be estimated with minimum variance, subject to $n_1 + n_2$ fixed, then Chapman (1955) showed that the optimal sample allocation is

$$\frac{n_2}{n_1} = \left(\frac{X_2 Y_2}{X_1 Y_1} \right)^{1/2} = \left(1 - \frac{R_x}{X_1} \right)^{1/2} \left(1 - \frac{R_y}{Y_1} \right)^{1/2}$$

and n_2 should be slightly smaller than n_1 . The above also holds for $R_y = 0$. If X_1 is to be estimated, we now have the optimum sample allocation

$$\frac{n_2}{n_1} = \frac{N_2}{N_1} \left(\frac{X_1 Y_2}{X_2 Y_1} \right)^{1/2}.$$

For the case $R_y = 0$, $Y_1 = Y_2$, and we have (Chapman, 1955)

$$\frac{n_2}{n_1} = \left(1 - \frac{R_x}{N_1} \right)^{1/2} \left(1 - \frac{R_x}{X_1} \right)^{-1/2}.$$

Here, n_2 should be slightly larger than n_1 , so that if \widehat{N}_1 and \widehat{X}_1 are both required, a good compromise is to have $n_1 = n_2$. Tables for choosing n_2/n_1 from Chapman (1955) are given by Seber (1982, 363–364), who also describes several confidence intervals for N_1 from Paulik and Robson (1969).

10.9.6 Exploitation Rate

If $f = R_x/R$, Paulik and Robson (1969) showed that the rate of exploitation is

$$u = \frac{R}{N_1} = \frac{P_1 - P_2}{f - P_2}$$

and the rates for x -type and y -type are, respectively,

$$u_x = \frac{fR}{P_1 N_1} = \frac{fu}{P_1}$$

and

$$u_y = \frac{(1-f)u}{(1-P_1)}.$$

If f is known or an estimate \hat{f} is available, then estimates are

$$\hat{u} = \frac{\hat{P}_1 - \hat{P}_2}{\hat{f} - \hat{P}_2}, \quad \hat{u}_x = \frac{\hat{f}\hat{u}}{\hat{P}_1}$$

and using the delta method,

$$\text{var}[\hat{u}] \approx (f - P_2)^{-4} \left\{ (f - P_2)^2 \text{var}[\hat{P}_1] + (f - P_1)^2 \text{var}[\hat{P}_2] + (P_1 - P_2)^2 \text{var}[\hat{f}] \right\}.$$

If $R_y = 0$, then

$$u_x = \frac{X_1}{N_1} = \frac{P_1 - P_2}{P_1(1 - P_2)}.$$

10.9.7 Subsampling

Sometimes, data is obtained from sampling subareas. Suppose the total population area is divided up into K subareas and k are selected at random. Let x_{ij} be the number of x types in the i th sample of n_{ij} individuals at time t_i from the j th subarea ($i = 1, 2, j = 1, \dots, k$). Let P_{ij} be the proportion of x -types in the j th subarea at time t_i . Then two estimates of P_i are available, namely, the pooled estimate

$$\hat{P}_i = \left(\sum_{j=1}^k x_{ij} \right) \left(\sum_{j=1}^k n_{ij} \right) = x_i/n_i$$

and the average

$$\bar{P}_{i \cdot} = \frac{1}{k} \sum_{j=1}^k \left(\frac{x_{ij}}{n_{ij}} \right) = \frac{1}{k} \sum_{j=1}^k \hat{P}_{ij}, \text{ say.}$$

We can test $P_{ij} = P_i$ (i.e., the data in all the subareas can be pooled) using the chi-squared test of homogeneity

$$T = \sum_{J=1}^k \left\{ \frac{(x_{ij} - n_{ij} \hat{P}_i)^2}{n_{ij} \hat{P}_i (1 - \hat{P}_i)} \right\},$$

with $k - 1$ degrees of freedom. If the hypothesis is rejected, then $\bar{P}_{i \cdot}$ should be used, and if k/K is small, an approximately unbiased estimate of $\text{var}[\bar{P}_{i \cdot}]$ is

$$\widehat{\text{var}}[\bar{P}_{i \cdot}] = \frac{1}{k(k-1)} \sum_{j=1}^k (\hat{P}_{ij} - \bar{P}_{i \cdot})^2.$$

When $n_{ij} = n$ for all i, j , the two estimators of P_i are identical.

10.9.8 Continuous Case

The case of continuous removals was considered by (Seber, 1982, 386–392) using instantaneous fishing (exploitation) and natural mortality rates for both x - and y -type fish. Because there are too many parameters, some assumptions needed to be made about the equality of some of the parameters. Other assumptions can be made to reduce the number of parameters. For example, Claytor and Allard (2003) proposed an extension of the CIR method to estimate the exploitation rate for lobsters based on continuous sampling concurrent with fishing. They assumed that (1) the population was closed, (2) the ratio of catchability between classes is constant throughout the season for all traps, (3) the ratio of catchability by the monitoring traps and by the commercial traps is constant over the season for all classes, and (4) the ratio of the fleet effort to the monitoring trap effort either is constant over the season or can be estimated up to a constant factor.

The two classes used were the exploited length class $i = e$ and reference length class $i = r$ using the above and below minimum length classes, respectively, where the latter referred to all lobster smaller than the minimum carapace length that had to be returned to the water. Let $j = 1, \dots, J$ refer to the J individual sequential fishing events that occurred during the fishing season, E_j denote the fleet effort at fishing event j , and e_j denote the corresponding monitoring effort. Let c_{ij} denote

the monitoring trap catch of length class i at event j , and let $p_j = c_{ej}/(c_{ej} + c_{rj})$. Exploitation rates were then estimated from the p_j .

10.10 Multiple Removals and Samples

Chapman (1955) extended the CIR theory to the case where there are two classes but several selective removals, each followed by a random sample. For $i = 1, \dots, s$, let R_{xi} and R_{yi} be the total (cumulative) removals from the x and y types, respectively, prior to the i th sample. Let

$$R_i = R_{xi} + R_{yi}, \quad (R_1 = 0)$$

$$X_i = X_1 - R_{xi}$$

$$N_i = N_1 - R_i$$

$$P_i = X_i/N_i$$

If sampling can be assumed to be binomial, the maximum likelihood estimators \hat{X}_1 and \hat{N}_1 are solutions of

$$\sum_{i=1}^s (x_i/X_i) = \sum_{i=1}^s (n_i/N_i) \text{ and}$$

$$\sum_{i=1}^s (y_i/Y_i) = \sum_{i=1}^s (n_i/N_i),$$

which can be obtained iteratively. The asymptotic variances are

$$\text{var}[\hat{X}_i] = \sum_{i=1}^s \left(\frac{n_i P_i^2}{X_i Y_i d} \right)$$

$$\text{var}[\hat{N}_i] = \sum_{i=1}^s \left(\frac{n_i}{X_i Y_i d} \right),$$

where

$$d = \left(\sum_{i=1}^s \frac{n_i}{X_i Y_i} \right) \left(\sum_{i=1}^s \frac{n_i P_i^2}{X_i Y_i} \right) - \left(\sum_{i=1}^s \frac{n_i}{N_i Y_i} \right)^2$$

$$= \sum_j \sum_{i < j} \left(\frac{n_i n_j (P_i - P_j)^2}{X_i Y_i X_j Y_j} \right).$$

A pair of first approximations for the iterative process is given, for instance, by the first and last samples, i.e.,

$$\tilde{N}_1 = \frac{R_s - R_{xs} \hat{P}_s}{\hat{P}_1 - \hat{P}_s}, \quad \tilde{X}_1 = \hat{P}_1 \tilde{N}_1.$$

Otis (1980) extended the above theory to three types (classes) of animals and two samples under the key assumptions that all members of the population have the same probability of being sampled in each of the pre-removal and post-removal samples. Let the three classes have sizes X_i , Y_i , and Z_i for samples n_i ($i = 1, 2$), with corresponding removals R_u ($u = x, y, z$). Let $N_i = X_i + Y_i + Z_i$, $R = R_x + R_y + R_z$, $X_2 = X_1 + R_x$, etc. Let $n_i = x_i + y_i + z_i$ ($i = 1, 2$) be the size of the i th sample with observations in the three classes. If sampling is with replacement, we now have trinomial distributions instead of binomial distributions, namely,

$$f(x_1, x_2, y_1, y_2) = \prod_{i=1}^2 \frac{n_i!}{x_i! y_i! z_i!} \left(\frac{X_i}{N_i}\right)^{x_i} \left(\frac{Y_i}{N_i}\right)^{y_i} \left(\frac{Z_i}{N_i}\right)^{z_i}.$$

Using a different notation, Otis gave expressions for finding the maximum likelihood estimators for X_1 , Y_1 , N_1 , and $Z_1 = N_1 - X_1 - Y_1$. Asymptotic variances were obtained in the usual manner from the information matrix. Optimization can be done numerically perhaps using some moment estimates as starting values.

Assuming sampling with replacement, the extension to more than $r = 3$ classes and $s = 2$ samples is straightforward by replacing the trivariate distributions by general multinomial distributions. Also, the numbers of samples and removals can be extended in a sequential manner. Skalski and Millspaugh (2006) and Skalski et al. (2005) gave details of this theory, under the assumption that all animals have the same probability of detection within a sample, and calculated the maximum likelihood estimates using a computer program called USER.

Udevitz and Pollock (1991) generalized the above by relaxing the assumption of equal detection probabilities and allowing different probabilities, provided that the ratios of encounter probabilities remained constant over time: the constant probability ratio assumption. Here, λ_i is the ratio of the encounter probability for individuals in subclass i to that for individuals in subclass 1, irrespective of the sample number ($\lambda_1 = 1$). Under the Otis (1980) model, these ratios are assumed to have the value of 1. If λ_2 or λ_3 is given a value (constraint), then explicit moment estimates, which are also maximum likelihood, can be found for $r = 3$ subclasses and $s = 2$ sampling periods. The authors extended this to general $r \geq 2$ subclasses and $S \geq 2$ samples. If $s = 2$ and $r \geq 3$, only one constraint is needed.

Paulik and Robson (1969) recommended that the CIR method should not be used if subclass proportions changed by less than 0.05 and only with caution if they changed by less than 0.1. They noted that when there are more than two subclasses, more than two samples, or unequal encounter probabilities, the situation becomes more complex and it does not appear that such simple rules of thumb can

be developed. However, when there are substantial changes in at least some of the proportions, the model selection procedures and the estimators under the correct model can perform quite well, whereas they can perform quite poorly when all changes in proportions are small (Udevitz & Pollock, 1991).

10.11 Index-Removal Method

The index-removal (IR) method makes use of the decline in a measure of relative abundance due to a known removal, where the relative abundance is measured in surveys before and after the removal. The method requires the assumptions that (1) the population is closed except for the removals, which are known exactly or can be estimated, and (2) all animals have the same probability of being captured in a survey and the capture probability is constant over time. The second assumption is relaxed later.

From Eberhardt (1982), we have a simplified model which assumes that the pre-removal index is a count, denoted by n_1 , and that it has some constant relationship (p) to the initial population size (N_1) as does the post-removal index (n_2) to the current population size. Then $E[n_1] = pN_1$ and $E[n_2] = p(N_1 - R)$, where R is the total removal. Taking the ratio of the two expectations gives us

$$\widehat{N}_1 = n_1 R / (n_1 - n_2),$$

and substituting this estimate in $E[n_1]$ gives us

$$\widehat{p} = (n_1 - n_2) / R.$$

If the n_i 's are assumed to follow independent Poisson distributions, then the above estimates are also maximum likelihood estimates. If $p^* = R/N_1$ is the proportion removed, then $\widehat{p}^* = (n_1 - n_2)/n_1$. In the next section, we introduce sampling effort into the mix!

10.12 Known Sampling Effort

Let $E[x_i] = a_i X_i$ and $E[y_i] = b_i Y_i$; then, with binomial sampling, we effectively assume that $a_i = b_i = n_i/N_i$. However, if the sample sizes are determined by sampling effort so that the n_i are now random variables, we can make the alternative assumption (Chapman and Murphy, 1965) that $a_i = b_i = K f_i$, where K is the so-called binomial catchability (Sect. 11.1) and the sampling efforts f_1 and f_2 are

known. Also, $E[n_i] = Kf_iN_i$, so that replacing $E[n_i]$ by n_i and setting $c_i = n_i/f_i$, the catch per unit effort, we have the equations (Seber, 1982, 375–376)

$$\frac{c_1}{N_1} = \frac{c_2}{N_1 - R} = K = \frac{c_1 - c_2}{R},$$

leading to the estimators

$$\widehat{N}_1 = \frac{Rc_1}{c_1 - c_2} \text{ and } \widehat{N}_2 = \frac{Rc_2}{c_1 - c_2}.$$

The estimator \widehat{N}_1 was first given by Petrides (1949) and is particularly useful when it is not possible to distinguish between the two classes, e.g., deer tracks in the snow (Davis, 1963). We also have

$$\widehat{N}_1 = R/(1 - c_2/c_1),$$

so that animal signs can be used to measure the number of animals, provided the index signs per animal is the same for both samples. The exploitation rate $u = R/N_1$ can be estimated by $\widehat{u} = (c_1 - c_2)/c_1$.

Using the delta method (Appendix A.4), we have the approximate formula

$$\text{var}[\widehat{N}_1] = (E[c_1 - c_2])^{-2}\{N_1^2\text{var}[c_2] + N_2^2\text{var}[c_1]\}.$$

If n_1 and n_2 are assumed to be independent Poisson random variables, then $E[c_i]$ and $\text{var}[c_i]$ can be estimated by c_i and c_i/f_i , respectively. When the two classes are distinguishable, X_1 is estimated by $\widehat{X}_1 = (x_1/n_1)\widehat{N}_1$.

10.12.1 Several Subclasses, Removals, and Samples

Udevitz and Pollock (1995) made use of effort information to extend the above models to having several subclasses, removals, and samples. Let $r > 1$ be the number of subclasses and $s > 1$ the number of sampling periods. Between each pair of sampling periods, a known number of individuals are removed from one or more subclasses. Let x_{ij} be the number removed in sample j from subclass i . The j th sample then consists of $n_j = \sum_{i=1}^r x_{ij}$ encounters with individuals from subclass i ($i = 1, \dots, r$, $j = 1, \dots, s$). Sampling is with replacement, or the sampling fractions are negligible without replacement. Encounters with individuals in the samples are independent. If sampling efforts are known, and f_j is the effort used to obtain sample j , we assume that

$$x_{ij} \mid f_j \sim \text{Poisson}(\gamma_{ij}f_j X_{ij}),$$

where X_{ij} is the number of subclass i individuals in the population during the j th sampling period, and these numbers are independent. Here, γ_{ij} is the Poisson intensity function and can be interpreted as the per-unit-effort encounter probability for any given type i individual at time j . The likelihood function is

$$\prod_{i=1}^r \prod_{j=1}^s \frac{1}{x_{ij}!} (\gamma_{ij} f_j X_{ij})^{x_{ij}} \exp[-(\gamma_{ij} f_j X_{ij})]. \quad (10.16)$$

If R_{ij} is the number of subclass i individuals removed from the population between sampling periods j and $j + 1$, then

$$X_{ij} = X_{i1} - \sum_{k=1}^{j-1} R_{ik}, \quad i = 1, \dots, r, \quad j = 1, \dots, s$$

so that the initial subclasses, X_{i1} ($i = 1, \dots, r$), are the unknown parameters of interest. With the unknown γ_{ij} , there are a total of $rs + r$ unknown parameters, and the parameters are not identifiable without imposing at least r constraints on the γ_{ij} . There are several possibilities, and one method is to assume that the ratios of encounter probabilities for different subclasses remain constant over time as described above, namely, $(r - 1)(s - 1)$ constraints

$$\frac{\gamma_{i1}}{\gamma_{11}} = \frac{\gamma_{ij}}{\gamma_{11}} \equiv \lambda_i, \quad i = 2, \dots, r; \quad j = 2, \dots, s,$$

where λ_i represents the unknown but constant ratio of the encounter probability for subclass i to that for subclass 1. A consequence of this assumption is that the ratios of encounter probabilities at different times must be the same for all subclasses. This can be seen by rearranging the above equation to get an equivalent set of $(r - 1)(s - 1)$ constraints

$$\frac{\gamma_{1j}}{\gamma_{11}} = \frac{\gamma_{ij}}{\gamma_{11}} \equiv \beta_j, \quad i = 2, \dots, r; \quad j = 2, \dots, s.$$

If the λ_i ($i = 1, \dots, r$) and the β_j ($j = 2, \dots, s$) are unspecified, then the two sets of constraints are entirely equivalent, and either set can be viewed as a consequence of the other.

Reparameterizing according to the two sets of constraints, the likelihood (10.16) becomes

$$\prod_{i=1}^r \prod_{j=1}^s \frac{1}{x_{ij}!} (\gamma_{11} \lambda_i \beta_j f_j X_{ij})^{x_{ij}} \exp[-(\gamma_{11} \lambda_i \beta_j f_j X_{ij})], \quad (10.17)$$

where $\lambda_1 = \beta_1$, by definition. If $s \geq 3$, the unknown parameters can be estimated directly from this likelihood with or without additional constraints. If $s = 2$, the parameters are not identifiable without at least one additional constraint.

Instead of (10.16), we can use the conditional distribution

$$f(x_{1j}, \dots, x_{rj} | n_j) = \left[\frac{n_j!}{x_{1j}! \cdots x_{rj}!} \prod_{i=1}^r \left(\frac{\lambda_i X_{ij}}{\sum_{i=1}^r \lambda_i X_{ij}} \right)^{x_{ij}} \right], \quad (10.18)$$

and the likelihood is the product of the above over j . If no assumptions are made about the λ_i ($i = 2, \dots, r$), maximum likelihood estimates can be obtained from (10.18). The likelihoods presented above by Chapman (1955) for $r = 2$ and by Otis (1980) for $r = 3, s = 2$ are obtained as special cases by setting $\lambda_i = 1$ ($i = 1, \dots, r$). Udevitz and Pollock (1995) considered various special cases. They emphasized that CIR methods perform well only when there are substantial changes in subclass proportions.

As noted by Dawid and Lauritzen (1993), the above method requires the assumption that all animals have the same probability of being captured by one unit of effort. This was not the case with their application to crabs, as the sampling is selective for x -type crab, in this case the legal-sized crab. However, recalling equation

$$\widehat{N}_1 = \frac{R c_1}{c_1 - c_2}$$

from Sect. 10.12, we have a similar equation

$$\widehat{X}_1 = \frac{R_x c_{x1}}{c_{x1} - c_{x2}},$$

where c_{xi} is the catch per unit effort of x -type individuals in sample i . If an estimate \widehat{P}_1 of $P_1 = X_1/N_1$ is available, then we have the estimate

$$\widehat{N}_1 = \widehat{X}_1 / \widehat{P}_1.$$

Expressions for the asymptotic variances were given by the authors.

Ihde et al. (2008b) gave a multiple-year IR model, called the 2qIR model, which allowed catchability to vary by season, thus extending two previous models, the “one-year” model of Eberhardt (1982) above and the multiyear model 1qIR model of Ihde et al. (2008a). Ihde et al. (2008b) compared the three models using simulation under various changes of parameter values. The new model can be used when (1) the pre- and post-harvest survey indices of abundance have been obtained in at least 2 years, (2) the exploitation pre- and post-harvest survey indices of abundance have been obtained in at least 2 years, (3) the exploitation rate varies among years, and (4) the seasonal catchability coefficients remain constant across years.

Let N_{ij} be the population size at the time of survey j in year i , let the catch be x_{ij} for that survey, and let f_{ij} be the effort expended. Let q_j be the catchability coefficient in season j . For the case $i = j = 2$, we assume independent Poisson models for the x_{ij} , namely,

$$L = \prod_{i=1}^2 \prod_{j=1}^2 \frac{(q_j N_{ij} f_{ij})^{x_{ij}} \exp[-q_j N_{ij} f_{ij}]}{x_{ij}!}.$$

If N_i is the population size at the beginning of year i and R_i the removal in year i between seasons, then $N_{11} = N_1$, $N_{12} = N_1 - R_1$, $N_2 = N_{21}$, and $N_{22} = N_2 - R_2$. We have four equations and four unknown parameters so we can obtain four moment estimates by equating each random variable to its expected value, where

$$\begin{aligned}\mathbb{E}[x_{11}] &= q_1 f_{11} N_1 \\ \mathbb{E}[x_{12}] &= q_2 f_{12} N_1 - R_1 \\ \mathbb{E}[x_{21}] &= q_1 f_{21} N_2 \\ \mathbb{E}[x_{22}] &= q_2 f_{22} (N_2 - R_2)\end{aligned}$$

These give the moment estimators

$$\begin{aligned}\widehat{N}_1 &= \frac{x_{11}(X_{12}R_2 - x_{22}R_1)}{X_{12}x_{21} - x_{11}x_{22}} \\ \widehat{N}_2 &= \frac{x_{21}(X_{12}R_2 - x_{22}R_1)}{X_{12}x_{21} - x_{11}x_{22}} \\ \widehat{q}_1 &= \frac{x_{12}x_{21} - x_{11}x_{22}}{x_{12}R_2 - x_{22}R_1} \\ \widehat{q}_2 &= \frac{x_{12}x_{21} - x_{11}x_{22}}{x_{11}R_2 - x_{21}R_1}.\end{aligned}$$

The above model can be extended to n surveys per year by modeling

$$x_{ij} \sim \text{Poisson}(q_j f_{ij} N_{ij}), \quad i = 1, \dots, n; j = 1, 2,$$

and $N_{i2} = N_{i1} - R_i$. When more than 2 years of data are available, nonlinear maximization software is required to obtain parameter estimates. The authors noted some problems such as when the pre- and post-harvest catch rates are similar in magnitude, extremely large abundance estimates can result for all IR models. They found that the “2qIR model estimates were always more accurate and precise than those of the other models examined and other model scenarios in which there was moderate contrast in exploitation among years, regardless of the seasonal difference

between survey catchability coefficients.” We consider a combined approach in the next section.

10.13 Omnibus Removal Methods

In the previous theory, we considered three methods that can be used to estimate population size when survey data are collected just before and just after two or more known harvests: change-in-ratio (CIR), index-removal (IR), and catch-effort methods.

10.13.1 Combining CIR and IR

Dawe et al. (1993) combined the CIR and IR methods and applied the model to snow crabs using the above CIR theory and the previous theory for the index-removal method. The removals R and R_x were estimated, and methods for doing this were described, so that the theory of Sect. 10.9.4 applies. As the focus was on X_1 , the initial number of legal-sized lobsters, let \widehat{X}_{1a} be the CIR estimate and \widehat{X}_{1b} be the index-removal estimate with variances σ_a^2 and σ_b^2 , respectively, and covariance σ_{ab} . If the estimates give similar results, a combined estimate is given by the weighted average $\widehat{X}_{1com} = \alpha\widehat{X}_{1a} + (1 - \alpha)\widehat{X}_{1b}$, with variance

$$\text{var}[\widehat{X}_{1com}] = \alpha^2\sigma_a^2 + (1 - \alpha)^2\sigma_b^2 + 2\alpha(1 - \alpha)\sigma_{12},$$

where α is chosen to minimize the above variance, namely,

$$\alpha = \frac{\sigma_b^2 - \sigma_{12}}{\sigma_a^2 + \sigma_b^2 - 2\sigma_{12}}.$$

The exploitation rate u_x for x -type crabs is estimated by $\widehat{u}_x = \widehat{R}_x/\widehat{X}_1$, and extensive variance and covariance estimators are given by the authors in the appendix of their paper.

Frusher et al. (2007) considered ways of evaluating the catchability assumptions for the two methods as applied to lobster. These two methods can be used to estimate the exploitation rate when a survey is conducted before and after a harvest. When the magnitude of the harvest is also known, it is possible to estimate the abundance and the survey catchability coefficient. For the CIR method, it is not necessary to assume that the two defined groups (generally, legal- and sublegal-sized animals) have equal catchability. If sublegal-sized animals are unexploited, the estimated exploitation rate for the legal-sized group is shown below to be unbiased. The assumption for the CIR method that the ratio of survey catchabilities of the two size groups does

not vary over time can be evaluated by comparing the size composition of sublegal-sized animals over time in the surveys.

For the IR method, we have $u = R/N_1$, which is then estimated by $\hat{u} = (c_1 - c_2)/c_1$ on the assumptions that all animals are equally catchable and catchability does not change from one survey to the next. It may be informative to plot the catch rates from each survey against the size class to look for systematic differences between the surveys. This assumption can be evaluated by comparing the catch rates of sublegal-sized animals before and after the fishing season. Thus, if the catch rate of animals just below the legal size limit has not changed from one survey to the next, then it is reasonable to suppose the catchability of legal-sized animals has not changed. If a formal test is desired of whether the catch rates have changed from one survey to the next, it can be obtained using a t -test for the size class immediately below the legal size limit or using Hotelling's T^2 test (see Seber, 1984) to test the equality of catch rates over time for several size classes simultaneously.

The authors referred to other procedures for checking the assumptions. They also compared the two methods when there are natural mortality and recruitment between surveys. The most likely problem using the IR method is a change in catchability from one survey to the next. In contrast, the CIR method is unaffected by change in catchability between surveys, provided the ratio of catchability remains constant over time. Thus, the weaker assumption of the CIR method (constant ratio of catchability) is more likely to be met in practice than the assumption of the IR method (constant catchability).

Referring to the theory in Sect. 10.9.2 assumption (ii), if λ is the ratio of the probability of sampling y-type animals to the probability of sampling x-type animals, then x_i is Binomial($n_i, P_{i\lambda}$), where

$$P_{i\lambda} = \frac{X_i}{X_i + \lambda Y_i}.$$

Assuming only x -type individuals are removed so that $Y_1 = Y_2$, the exploitation rate for x -type individuals is

$$u_{\lambda x} = \frac{P_{1\lambda} - P_{2\lambda}}{P_{1\lambda}(1 - P_{2\lambda})},$$

and the authors showed that this is equal to $u_x = X_1/N_1$, as before. To test the effect of a change in relative catchability of the two components between surveys, we replace the catchability parameter λ with a survey-specific value, λ_i , for $i = 1, 2$. Frusher et al. (2007) in their appendix proved that

$$u_{\lambda_1\lambda_2x} = 1 - \frac{\lambda_1}{\lambda_2} + \frac{\lambda_1 R_x}{\lambda_2 X_1} = 1 - \frac{\lambda_1}{\lambda_2} + \frac{\lambda_1}{\lambda_2} u_x,$$

which reduces to u_x when $\lambda_1 = \lambda_2$.

10.13.2 Combining Three Methods

Chen et al. (1998) gave a model combining all three methods. They modeled the survey and removal processes as a Poisson point process and a linear death process, respectively, and then combined the two processes. The complete data likelihood can be factored into three parts: the general likelihood function of the index-removal method, the general likelihood function of the change-in-ratio method, and the general likelihood function of the catch-effort method. They developed the model in three stages. We distinguish between efforts for sampling (f) and efforts for removals from harvesting (f^*).

We begin by using a Poisson distribution and the notation of the previous model by Udevitz and Pollock (1995) to obtain the likelihood (10.17) with $\gamma_{11} = \mu$. In the first survey, there are f_1 units of sampling effort and a total of n_1 animals with x_{11}, \dots, x_{r1} animals being seen in each subclass, respectively. The first survey is followed by a harvest with $R_{11}, R_{21}, \dots, R_{r1}$ removals in each subclass with a total of f_1^* units of sampling effort. The survey-removal procedures are repeated until all s surveys and $s - 1$ removals are completed, giving (10.16). Using the same parameter constraints for identifiability, we have $\gamma_{ij} = \mu\lambda_i\beta_j$. This gives us a model for the survey process.

We now consider a model for the index-removal method. The total number of encounters ($n_j = \sum_{i=1}^r x_{ij}$) is the sum of the independent Poisson random variables x_{ij} ($i = 1, \dots, r$). This implies that n_j is also Poisson so that we have the independent distributions

$$(n_j \mid f_j) \sim \text{Poisson} \left(\mu f_j \beta_j \sum_{i=1}^r \lambda_i X_{ij} \right), \quad j = 1, \dots, s,$$

where we recall that $X_{ij} = X_{i1} - \sum_{k=1}^{j-1} R_{ik}$. The likelihood for the total sample counts is now

$$L_1 = \prod_{j=1}^s \left[\frac{1}{n_j!} \left(\mu \beta_j f_j \sum_{i=1}^r \lambda_i X_{ij} \right)^{n_j} \exp \left[-(\mu \beta_j f_j \sum_{i=1}^r \lambda_i X_{ij}) \right] \right]. \quad (10.19)$$

If we assume all animals have the same probability of being caught by one unit of sampling effort regardless of class and survey (i.e., $\gamma_{ij} = \mu$ for all i, j), then L_1 ends up as the likelihood of the index-removal method. If the capture probabilities for the subclasses are not all the same, the initial total population ($\sum_i X_{i1}$) is not estimable in L_1 , but the problem of heterogeneity of capture probability can be solved by making separate estimates for each subclass provided that the sampling probabilities within each subclass are homogeneous.

The probability model for the change-in-ratio method is considered next. Using (10.18), this is given by

$$\begin{aligned} L_2 &= \prod_{j=1}^s f(x_{1j}, \dots, x_{rj} | n_j) \\ &= \prod_{j=1}^s \left[\frac{n_j!}{x_{1j}! \cdots x_{rj}!} \prod_{i=1}^r \left(\frac{\lambda_i X_{ij}}{\sum_{i=1}^r \lambda_i X_{ij}} \right)^{x_{ij}} \right]. \end{aligned} \quad (10.20)$$

We now consider the removal model, but assume the removal rates are linearly proportional to the population size. The removal system can then be modeled as a pure linear death process where it is assumed that η_{ij} , the individual death rate, is the same for all individuals of a given subclass during a given removal period. In fisheries work, η_{ij} is referred to as the catchability coefficient. Under these assumptions, we have

$$(R_{ij} | f_j^*) \sim \text{Binomial}(X_{ij}, 1 - \exp[-\eta_{ij} f_j^*]).$$

If the constant probability ratio assumption is imposed on the death rate we have $\eta_{ij} = v\rho_i \tau_j$, giving the full likelihood function for the removal process

$$L_3 = \prod_{j=1}^{s-1} \prod_{i=1}^r \binom{X_{ij}}{R_{ij}} (1 - \exp[-v\rho_i \tau_j f_j^*]^{R_{ij}}) (\exp[-v\rho_i \tau_j f_j^*]^{X_{ij} - R_{ij}}).$$

Finally, we now put the removal and survey sampling together by assuming the sample counts (x_{ij}) and the removals (R_{ij}) are random variables and factorizing the likelihood into three parts, namely,

$$\begin{aligned} L &= \Pr(x_{1j}, \dots, x_{rj} | j = 1, \dots, s; R_{1j}, \dots, R_{rj} | j = 1, \dots, s-1) \\ &\quad f_j | j = 1, \dots, s; f_j^* | j = 1, \dots, s-1 \\ &= \prod_{j=1}^s \Pr(n_j | R_{1j}, \dots, R_{rj}; f_j, X_{i1}) \prod_{j=1}^s \Pr(x_{1j}, \dots, x_{r-1,j} | n_j, \{R_{ij}\}, X_{i1}) \\ &\quad \times \prod_{j=1}^{s-1} \Pr(R_{1j}, \dots, R_{rj} | f_j^*; X_{i1}) \\ &= L_1 L_2 L_3, \end{aligned}$$

where L_1 is the general likelihood of the index-removal model, L_2 is the general change-in-ratio likelihood, and L_3 is the general likelihood function of catch-effort or removal models. This approach should lead to large gains in efficiency because

of more complete use of the data collected. Extensive simulations were carried out by the authors.

10.14 Summary

This chapter is about the so-called removal method for estimating the size of a closed population. It consists of removing a sequence of samples from the population, which can be useful for dealing with an invasive species where depletion is desirable. It has also been used for capture-recapture experiments when recaptures are ignored so that an individual recaptured is effectively “removed” from the study. This can avoid such problems as the possible effect of a declining population on the probability of capture p or “catchability.” The alternative is to temporarily remove individuals using, for example, electrofishing.

If the probability of capture is allowed to vary, then we have too many parameters, together with the initial population size, to estimate. Several solutions are presented for the probability parameters: (1) use logistic regression for the probabilities; (2) assume p is constant, and also use regression models; (3) combine with a marked release; and (4) use knowledge of sampling effort.

Variable catchability can be directly modeled in various ways using the so-called generalized removal model and Bayesian methods. Because of their usefulness, two and three removals are given special attention. Subpopulations are also considered, including multiple sites. Removal methods for point-count methods (Chap. 7) and times to first capture (detection), the so-called TTDD model, are described.

Indirectly related to the removal method for estimating population sizes is the so-called change-in-ratio (CIR) method where removals are carried out from two or more classes (e.g., males and females) changing the class ratios and for two or more removals. Various associated parameters such as exploitation rate are estimated, and effort information can be incorporated. Omnibus removal methods which combine various methods such the removal, CIR, and the index-removal method are described in detail.

Chapter 11

Catch-Effort Models



11.1 Types of Model

Catch-per-unit-effort methods for closed populations are an extension of the previous chapter on removal sampling. There, use is based on the general assumption that the size of a sample caught from a population is proportional to the effort put into taking the sample, though sometimes proportionality may not exist (e.g., Ye & Darren, 2009). A brief history of the topic is given by Seber (1982, Chapter 7).

Let P_i be the probability an individual is caught in the i th sample ($i = 1, \dots, s$) with effort f_i and removed from a closed population. If sampling is a Poisson process with regard to effort, with intensity k , then we find that $Q_i = 1 - P_i = \exp[-kf_i]$. Here k has usually been called the (Poisson) catchability coefficient (or q in fishery research). When kf_i is small, then $P_i \approx kf_i$, which can be made exact if we redefine k as K giving the so-called the binomial catchability coefficient, the average probability of being captured with one unit, where the units of effort are independent and additive. Then $P_i = Kf_i$, and this model is used later for regression methods.

Catch effort can be incorporated with count data through covariates. For example, Link and Sauer (1999) proposed a model of the form

$$\log(\mu_{ij}) = A_i^* + B^* \psi_{ij}^p + \phi_j,$$

where μ_{ij} is the expected value of a count for the i th circular plot and the j th year of the survey, $A_i^* = A_i - B/p$, $B^* = B/p$, and ϕ_j is a year effect. They considered the procedure for different values of p and tested whether $B = 0$, i.e., whether an effort adjustment was necessary. Their method of selecting an optimal value for p required substantial data and considerable computation. An important aspect of catch-effort methods is standardizing commercial catch and effort data for complex fisheries (Bishop, 2006).

The use of effort data can be built more directly into a model as follows. Let p_{hi} be the probability of catching the individual h ($h = 1, \dots, N$) in the i th removal ($i = 1, \dots, s$). It is assumed that individuals act independently and the units of effort are independent, e.g., traps do not compete with each other. The data consist of a sequence of s samples (removals) and the corresponding efforts $\{(n_1, f_1), \dots, (n_s, f_s)\}$. Here, for sample i , f_i denotes the effort, and n_i denotes the removal (which we can also think of as “unmarked” when capture equates to “removal” from the population). The ratio n_i/f_i is referred to as the catch-per-unit-effort (CPUE). The cumulative removals and efforts are defined by

$$x_i = \sum_{r=1}^{i-1} n_r, \quad \text{and} \quad F_i = \sum_{r=1}^{i-1} f_r, \quad i = 2, \dots, s+1, \quad x_1 = F_1 = 0.$$

Chao and Chang (1999) gave the following suite of models (with some changes in notation), as well as sources for their estimators.

| | |
|-----------------------------------|------------------------------------------------|
| Homogeneous Constant Effort (1) | $p_{hi} = \lambda$. |
| Heterogeneous Constant Effort (2) | $p_{hi} = \lambda_h$. |
| Homogeneous Variable Effort (3) | Multiplicative. $p_{hi} = \lambda f_i$. |
| Homogeneous Variable Effort (4) | Poisson. $p_{hi} = 1 - \exp[-kf_i]$. |
| Homogeneous Variable Effort (5) | Logit(p_{hi}) = $\lambda + \beta f_i$. |
| Heterogeneous Variable Effort (6) | Multiplicative. $p_{hi} = \lambda_h f_i$. |
| Heterogeneous Variable Effort (7) | Poisson. $p_{hi} = 1 - \exp[-\lambda_h f_i]$. |
| Heterogeneous Variable Effort (8) | Logit $p_{hi} = \lambda_h + \beta f_i$. |

These models are discussed below, except for model (8) for which estimation is not available, except for special cases given later.

11.2 Homogeneous Models

We first consider models which are homogeneous with regard to subscript h . Model (1) is the standard removal method with constant p described in Sect. 10.1.1 and does not involve effort. Model (4) assumes a Poisson process, where the probability of an individual not being caught when subjected to δf units of effort is $q_i = 1 - p_i = \exp[-kf_i]$, and k (q in fishery research) is the Poisson catchability coefficient. Let $r = \sum_{i=1}^s n_i$, $Q = q_1 q_2 \cdots q_s$, and $P = 1 - Q$. We recall from Chap. 10 that the joint distribution of the $\{n_i\}$ for the removal model is the multinomial distribution

$$f(\{n_i\}) = \frac{N!}{(\prod_{i=1}^s n_i!) (N-r)!} p_1^{n_1} (q_1 p_2)^{n_2} \cdots (q_1 q_2 \cdots q_{s-1} p_s)^{n_s} Q^{N-r-s+1}. \quad (11.1)$$

From (11.1) and model (4), the maximum likelihood estimates of k and N are the solutions of

$$kF_s = -\log(1 - r/N), \text{ and } NF_s = \sum_{i=1}^s f_i(x_i + n_i p_i^{-1}),$$

where $x_i = \sum_{j=1}^{i-1} n_j$, ($x_1 = 0$), and $F_s = \sum_{i=1}^s f_i$. These equations can be solved iteratively using, for example, regression estimates described below as a first approximation. Estimates can also be obtained using the conditional approach of (11.6). Since kf_i is usually small, their solution is facilitated by using the approximation $p_i \approx kf_i(1 - \frac{1}{2}kf_i)$. Seber (1982, 297), using Stirling's approximation for large factorials, showed that the asymptotic variance-covariance matrix of the maximum likelihood estimates \hat{N} and \hat{k} is \mathbf{V}^{-1} , where

$$\mathbf{V} = \begin{pmatrix} \mathbb{E}\left[\frac{x_{s+1}}{N(N-x_{s+1})} + \frac{1}{2N^2} - \frac{1}{2(N-x_{s+1})^2}\right], & F_{s+1} \\ F_{s+1}, & \mathbb{E}\left[\sum_{i=1}^s n_i f_i^2 q_i / p_i^2\right] \end{pmatrix}.$$

The $(1, 1)$ element of \mathbf{V}^{-1} is the asymptotic variance of \hat{N} . \mathbf{V} can be estimated by replacing the expectations by the random variables and the unknown parameters by their estimates. A statistic for testing the adequacy of the above multinomial model is

$$\sum_{i=1}^s \left\{ \frac{n_i - \hat{N}\hat{q}_1 \cdots \hat{q}_{i-1}\hat{p}_i)^2}{\hat{N}\hat{q}_1 \cdots \hat{q}_{i-1}\hat{p}_i} \right\},$$

which is asymptotically distributed as chi-squared with $s - 2$ degrees of freedom.

If we use the binomial catchability coefficient K and assume units of effort are independent and additive, then $p_i = Kf_i$, and we have model (3) in the above list models with $\lambda = K$, which can be analyzed using the above approach. We now consider using a different method for model (3).

11.2.1 Method of Estimating Functions

From Chao and Chang (1999) we have, with homogeneity, $p_{hi} = p_i = \lambda f_i$, say, where n_i is the i th removal, and there are s removal samples. Let $\theta = (N, \lambda)'$, let $x_{hi} = 1$ if the h th animal ($h = 1, \dots, N$) has been removed in samples 1 to i , and 0 otherwise, and let $\mathbf{x}_i = (x_{1i}, \dots, x_{Ni})'$. Just after the $(k-1)$ th sample, there are

$$D_{k-1} = \sum_{i=1}^{k-1} n_i = \sum_{h=1}^N x_{h,k-1},$$

total removals, and there are $N - D_{k-1} = \sum_{h=1}^N (1 - x_{h,k-1})$ animals left in the population ($k = 1, 2, \dots, s$, $D_0 \equiv 0$) with $D_s = r$, the total number of removals. We therefore have

$$\mathbb{E}[n_k \mid \mathbf{x}_{k-1}] = (N - D_{k-1})p_k.$$

For each sample k , we can construct an unbiased estimating function for N and the p_k , namely,

$$g_k = n_k - (N - D_{k-1})p_k, \quad k = 1, \dots, s.$$

To combine these s estimating functions, we consider a weight associated with g_k to be a function of \mathbf{x}_{k-1} . The optimal unbiased estimating equation with corresponding root that has minimum asymptotic variance is given by (e.g., see Liang & Zeger, 1995)

$$\mathbf{g} = \sum_{k=1}^s \Phi_k v_k^{-1} g_k,$$

where $v_k = \text{var}[g_k \mid \mathbf{x}_{k-1}]$ and

$$\Phi_k = (\mathbb{E}[\partial g_k / \partial N \mid \mathbf{x}_{k-1}], \mathbb{E}[\partial g_k / \partial \lambda \mid \mathbf{x}_{k-1}])'.$$

Substituting the partial derivatives and the variance $v_k = (N - D_{k-1})p_k(1 - p_k)$ into \mathbf{g} , we obtain the following estimating equations:

$$\begin{aligned} \sum_{k=1}^s [(N - D_{k-1})(1 - p_k)]^{-1} [n_k - (N - D_{k-1})p_k] &= 0 \\ \sum_{k=1}^s \left(\frac{\partial p_k}{\partial \lambda} \right) [p_k(1 - p_k)]^{-1} [n_k - (N - D_{k-1})p_k] &= 0. \end{aligned}$$

If we set $p_k = \lambda f_k$, we get model (3), and setting $p_k = 1 - \exp[-\lambda f_k]$, we get model (4). For model (5), the logistic model with additional parameter β now has $\boldsymbol{\theta} = (N, \lambda, \beta)'$, and Φ_k is three-dimensional, giving an additional estimating equation

$$\sum_{k=1}^s \left(\frac{\partial p_k}{\partial \beta} \right) [p_k(1 - p_k)]^{-1} [n_k - (N - D_{k-1})p_k] = 0.$$

The above three equations are usually referred to as quasi-likelihood equations, which have to be solved iteratively, and their solutions give so-called maximum quasi-likelihood estimators (MQLEs). If the efforts are scaled from f_k to bf_k

($b > 0$), then the solution for λ is multiplied by b^{-1} , but the solution for N remains unchanged.

When $\theta = (N\lambda)'$, Chao and Chang (1999) give an expression for the asymptotic variance V of the estimator \widehat{M}_{MQLE} of N . They also showed that the maximum likelihood estimates of N (say \widehat{N}) and λ satisfy

$$1 - \frac{D_s}{N} = \prod_{k=1}^s (1 - p_k)$$

$$\sum_{k=1}^s \left(\frac{\partial p_k}{\partial \lambda} \right) [n_k - (N - D_{k-1})p_k] = 0.$$

The asymptotic variance of \widehat{N} is the same as V , which only holds for large samples. Therefore, the authors suggested the use of a bootstrap procedure by generating a sequence of values of the underlying multinomial distribution, giving \widehat{N}_B , as follows. If $n_0 = N - r$, we begin by generating the first batch $\{n_1^*, \dots, n_s^*, n_0^*\}$ from a multinomial distribution with parameter \widehat{N}_B and cell probabilities $\{n_1/\widehat{N}_B, \dots, n_s/\widehat{N}_B, 1 - r/\widehat{N}_B\}$. Then based on the batch data, an estimate N^* is calculated. After k replications, the average of the k values of the N^* is calculated, and its variance is estimated by the sample variance of the values.

11.3 Heterogeneous Models

Referring to the models in Sect. 11.1, model (2) is known as the generalized removal method and does not involve effort data. Model (5) involves a logistic regression

$$\text{logit}(p_{hi}) = k + \beta f_i,$$

which is generally straightforward for maximum likelihood estimation. We now focus on the remaining heterogeneous models. For the multiplicative model, (6) we have

$$p_{hi} = \lambda_h f_i,$$

where λ_h is the unknown individual h 's removal (or capture) probability for a unit of effort and f_i is the known effort in the i th sample. Here, λ_h and f_i are defined only up to a multiplicative constant. Since $\lambda_h f_i$ may be greater than 1, a suitable modification is $p_{hi} = \min(\lambda_h f_i, 1)$. Model (6) with $p_{hi} = 1 - \exp[-\lambda_h f_i]$ is derived under the assumption that sampling of the i th individual is a Poisson process with respect to effort with parameter λ_h , a Poisson catchability coefficient. We discuss this model below.

11.3.1 Using Coverage and Estimating Functions

These are given below by Chao and Chang (1999) for model (6) with algebraic details in their appendix. Treating the λ_h as fixed parameters, define $\bar{\lambda} = \sum_{h=1}^N \lambda_h / N$ and the coefficient of variation (CV)

$$\gamma = \sum_{h=1}^N [(\lambda_h - \bar{\lambda})^2 / N]^{1/2} / \bar{\lambda},$$

where CV is a measure of the degree of heterogeneity. The sample coverage of the first k samples, C_k , is defined as the probability-weighted fraction of the population that was removed in the first k samples, i.e.,

$$C_k = \frac{\sum_{h=1}^N \lambda_h x_{hk}}{\sum_{h=1}^N \lambda_h},$$

where $x_{hk} = 1$ if the h th animal has been removed in samples 1 to k , and 0 otherwise. We recall that

$$D_{k-1} = \sum_{i=1}^{k-1} n_i = \sum_{h=1}^N x_{h,k-1}.$$

If the λ_h are all equal, then $C_k = D_k / N$, so that a natural estimator of N without heterogeneity, i.e., homogeneity, is D_k / \widehat{C}_k , where (from the authors' appendix)

$$\widehat{C}_k = 1 - \frac{n_{k+1}/f_{k+1}}{n_1/f_1}.$$

The expected value of \widehat{C}_k is theoretically positive for all $k = 1, \dots, s-1$, but may be zero or negative for some k . Therefore, let τ be the largest k such that $\widehat{C}_k > 0$, i.e., such that $n_{k+1}/e_{k+1} < u_1/e_1$. Therefore, an estimate of N under the homogeneity model is $\widehat{N}_{hom} = D_\tau / \widehat{C}_\tau$, where in most cases $\tau = s-1$. We now develop the optimal estimating equation for the heterogeneous case.

Given the state vector $\mathbf{x}_{k-1} = (x_{1,k-1}, \dots, x_{N,k-1})'$, we have

$$n_k = \sum_{h=1}^N (x_{hk} - x_{h,k-1}) = \sum_{h=1}^N x_{hk} - (1 - x_{h,k-1})$$

because x_{hk} and $x_{h,k-1}$ take values 0 or 1 only. Then

$$\begin{aligned} \mathbb{E}[n_k \mid \mathbf{x}_{k-1}] &= \sum_{h=1}^N \lambda_h f_k (1 - x_{h,k-1}) \\ &= (1 - C_{k-1}) \left(\sum_{i=1}^N \lambda_i \right) f_k \\ &= (N - NC_{k-1}) \bar{\lambda} f_k, \quad C_0 \equiv 0. \end{aligned}$$

In the heterogeneous case $NC_{k-1} \neq D_{k-1}$. For notational simplicity let $NC_{k-1} \equiv D_{k-1}^*$. The idea here is to find an estimator for the expected discrepancy between NC_{k-1} and D_{k-1} . Using some approximations and estimating equations, then after some algebra, the authors obtained an estimate of N as

$$\hat{N}_{het} = \frac{\sum_{k=2}^s (f_k^2/n_k) [D_{k-1} + A_{k-1} n_{k-1} \hat{\gamma}^2 I(\hat{C}_{k-1} > 0)]}{\sum_{k=2}^s (f_k^2/n_k) \hat{C}_{k-1} I(\hat{C}_{k-1} > 0)},$$

where I is the usual indicator function, $A_{k-1} = (\sum_{i=1}^{k-1} f_i)/f_{k-1}$, and

$$\hat{\gamma}^2 = \max \left\{ \hat{N}_{hom} [n_1 - n_2(f_1/f_2)]/n_1^2 - 1, 0 \right\}.$$

A previous estimator given by Lee and Chao (1994) is

$$\hat{N}_{sc} = \frac{D_\tau + A_\tau n_\tau \hat{\gamma}^2}{\hat{C}_\tau}$$

Again a bootstrap variance estimator and confidence intervals can be constructed as in the homogeneity models. The authors compared the various estimators, including the regression estimators given below. They found that, except for sparse data, the bootstrap standard error estimates are generally satisfactory and concluded that for highly heterogeneous populations, the sample coverage approach is recommended. Finally, we note that model (7) can be dealt with in the same way as model (6) above.

11.4 Regression Models

From Eq. (10.1), a version of (11.1), we obtain the regression model

$$\mathbb{E}[n_i \mid x_i] = p_i(N - x_i) = K f_i (N - x_i),$$

and if we define $Y_i = n_i/f_i$, the catch per unit effort, we have

$$\mathbb{E}[Y_i | x_i] = K(N - x_i), \quad (i = 1, \dots, s), \quad (11.2)$$

a regression model first given by Leslie and Davis (1939) and De Lury (1947, 1951). Although the variances differ, other factors can affect the scatterplot, so it is suggested that an unweighted regression model be fitted (Seber, 1982, 298–299). A simple visual estimate of N can be obtained from the plot by setting $Y = 0$, giving $N = x$. In the following, we recall that K and k are the respective binomial and Poisson catchability coefficients.

A second model of De Lury (1947) uses (11.1) to give

$$\begin{aligned} \mathbb{E}[n_i] &= Nq_1q_2 \cdots q_{i-1}p_i \\ &= Np_i \exp[-kF_i]. \end{aligned} \quad (11.3)$$

If $p_i \approx kf_i$, then taking logarithms and setting $z_i = \log Y_i = \log(n_i/f_i)$ leads to

$$\mathbb{E}[z_i] = \log(kN) - kF_i. \quad (11.4)$$

Another model, given by Ricker (1958, 149), which treats the whole experiment as F_{s+1} samples, each with one unit of effort, gives

$$\mathbb{E}[N - x_i] = N(1 - K)^{F_i}.$$

Combining this with (11.2) and setting $z_i = \log Y_i = \log(n_i/f_i)$ leads to

$$\mathbb{E}[z_i] = \log(KN) + [\log(1 - K)]F_i. \quad (11.5)$$

When K is small, $\log(1 - K) \approx K$, $K = k$, and (11.5) leads to (11.4).

A different approximation is used by Paloheimo (1961) and Chapman (1961). From (11.3)

$$\begin{aligned} \mathbb{E}[Y_i] &= Nf_i^{-1}p_i \exp[-kF_i] \\ &= kN \exp[-kF_i](1 - \exp[kf_i])/(kf_i), \end{aligned}$$

and using the approximation

$$\log[(1 - e^{-w})/w] \approx -w/2$$

for small w , we have

$$\mathbb{E}[Y_i] = kN \exp[-k(F_i + \frac{1}{2}f_i)].$$

Then taking logarithms

$$\text{E}[z_i] \approx \log(kN) - k(F_i + \frac{1}{2}f_i).$$

Braaten (1969) showed that this model is more robust to changes in k than (11.4).

Ricker recommended the use of (11.4) when k and K are small, say less than 0.02, and (11.5) when K is larger. The reason for this is that when K is of moderate size so that one unit of effort catches a reasonable proportion of the population, then the effects of sampling without replacement during each sample can no longer be ignored. Although (11.2) depends on less assumptions, it is helpful to plot several different regression models, as departures from underlying assumptions are liable to be expressed in different ways. It is important that there are sufficient removals so that there is a significant decline in the catch per unit effort, though this clashes with the need for having kf_i small for the approximation $p_i \approx kf_i$ to hold. Further details relating to the above regression methods are given by Seber (1982, 297–304).

Gould and Pollock (1997) pointed out some shortcomings of the above regression methods, for example, equating the expected value of the log to the log of the expectation, and not allowing for a variable variance structure as the CPUE decreases as the population is depleted. Their results indicated that “maximum likelihood estimation consistently provides less biased and more precise estimates than the regression methods and allows for greater model flexibility necessary in many circumstances.” They recommended the use of maximum likelihood estimation in catch-effort studies and used the model (11.1) expressed as $L(\theta | \{n_i\})$, but in the factorized form using two likelihoods, namely, (with $r = \sum_i n_i$)

$$L(\theta | \{n_i\}) = L_1(\theta | r)L_2(r | \{n_i\}), \quad (11.6)$$

where, setting $Q = q_1q_2 \cdots q_s$,

$$L_1(\theta | r) = \frac{N!}{r!(N-r)!}(1-Q)^r Q^{N-r}$$

and

$$L_2(r | \{n_i\}) = \frac{r!}{\prod_{i=1}^s n_i!} \left(\frac{p_1}{1-Q} \right)^{n_1} \left(\frac{q_1 p_2}{1-Q} \right)^{n_2} \cdots \left(\frac{q_1 q_2 \cdots q_{s-1} p_s}{1-Q} \right)^{n_s}.$$

Using this conditional maximum likelihood (ML) approach is asymptotically equivalent to using the unconditional ML (Sanathanan, 1972a) and has several advantages. For instance, L_2 can be maximized over k (or K) separately from N , and then we can use this estimate of k to maximize L_1 over N by solving the equality (Lindsay & Roeder, 1987)

$$\frac{L_1(N, \{p_i\} | r)}{L_1(N-1, \{p_i\} | r)} = 1.$$

This results in the estimator given in Chapman (1951)

$$\hat{N} = r/(1 - \hat{Q},)$$

where \hat{Q} is Q with k replaced by its estimator.

Computations can be carried out using a standard computer package and other computational methods, as described by the authors. For example, using the computer package **SURVIV**, they fitted a logistic model to p_i , namely,

$$\text{logit}(p_i) = \beta_0 + \beta_1 f_i + \beta_2 t_i,$$

where t_i is the temperature.

Chao and Chang (1999) applied the regression methods to a population of known size and found the estimates to be too low. However, in spite of the shortcomings of the regression methods, a plot is always a useful visual indication as to how well a model fits the data, as the models will be affected differently by departures from the underlying assumptions.

11.5 Underlying Assumptions

In any study, it is important to investigate the validity of any underlying assumptions, which can be listed as follows (Seber, 1982, 305–308).

Population closed. The whole experiment needs to be in one area for as short a period of time as possible to avoid mortality, migration, etc. This will, for example, have an effect on the size of s , the number of removals.

Constant catchability coefficient. Catchability can vary through a number of factors such as fish size, etc. so that K , for example, is effectively the average coefficient over all sizes in the catch. This average will be constant only if the distribution of size in the population remains constant; otherwise, stratifying the size classes and estimating each size class separately might be needed. However, environmental conditions can produce changes in the catchability coefficient. Perhaps K_i for the i th removal can be modelled using a logistic model with suitable covariates or expressed as a random effects model.

Constant trap efficiency. In fisheries, many kinds of gear decrease in efficiency the longer they are left before being lifted and reset, so that the catch per unit time begins to decrease. For example, as more fish are caught on a set line, the less vacant hooks there are, and eventually, a point of saturation may be reached. On the other hand, with net fishing, the presence of some fish already in the net may tend to scare others away, so that saturation may be reached long before the net is full. Gear saturation may need to be corrected for (e.g., Murphy, 1960). If different fishing gears are used for the same population, these need to

be combined in some way. Some suggestions for doing this are given by Seber (1982, 306–308).

Independent units of effort. Dependence can occur when traps are so close together that “physical” competition exists between them, which is independent of population size. For example, too many anglers at a pool may frighten the fish, or setting a new gill net near one already in operation may scare the fish away from the latter.

Random sampling. Sampling may not be random because of a number of factors such as not all the population being subject to the catching process because of inaccessibility of some part of the fishing grounds.

Complete record of effort. In commercially exploited populations information on effort may be missing or unavailable for some parts of the catch. Using the rest of the catch, its corresponding effort can be scaled up for the total catch. If effort records are complete and catch records incomplete, the previous procedure can be used in reverse.

We now consider some departures from constant catchability by incorporating effort data as in the previous chapter.

11.5.1 Some Variable Catchability Models

Vajas et al. (2021) considered a Bayesian catch-effort model using the logs of hunters as a management tool for studying exploited populations using wild boar as a case study. A range of factors such as habitat and hunting pressure can affect catchability. Consider a given management unit u belonging to the catchability group g , a given hunting season s , and a given hunting day i belonging to the month m . Let h_{usi} be the known hunting effort applied by the hunters of the whole u during the i th hunting day of the season s (here measured by the total number of hunters participating to a drive in this unit this day). Let P_{usi} be the hunting pressure (i.e., the proportion of animals in the units that are culled by the drive hunt), assumed to have a prior beta distribution $\text{Beta}(\alpha_{usi}, \beta_{usi})$. Consider the model

$$p_{usi} = E[P_{usi}] = 1 - \exp[-\gamma_{gm} h_{usi}],$$

where γ_{gm} is the unknown catchability of the wild boar in the catchability group g during the month m . Here

$$\alpha_{usi} = p_{usi}\phi, \text{ and } \beta_{usi}(1 - p_{usi})\phi,$$

where ϕ is a dispersion parameter and $\text{var}[p_{usi}] = p_{usi}(1 - p_{usi})/(\phi + 1)$. Let $N_{us} = D_{us}A_{us}$ be the initial population size of the unit u in season s , with D_{us} and A_{us} the corresponding density and area, respectively. Let c_{us1} be the number culled

on the first hunting day, and assume

$$c_{us1} \sim \text{Binomial}(N_{us}, p_{us1}),$$

leaving $N_{us} - c_{us1}$ in the population at the end of the day. We now have the usual removal model. For the next day, we have

$$c_{us2} \sim \text{Binomial}(N_{us} - c_{us1}, p_{us2}),$$

and, more generally, on day i , we have

$$c_{usi} \sim \text{Binomial}(N_{us} - \sum_{j=1}^{i-1} c_{usj}, p_{usi}).$$

This model supposes that the only source of mortality during the hunting season is caused by hunting, which is not a stringent assumption for the species considered. The parameters to be estimated were γ_{gm} , ϕ , and the initial population densities D_{usi} . Prior distributions were used for these parameters with models fitted using Markov Chain Monte Carlo and checked for goodness of fit.

An important question relates to sustainable of subsistence hunting in an indigenous reserve. Shaffer et al. (2017) used encounter rates from transect surveys and hunter catch-per-unit effort from hunter surveys to validate model predictions for several species in Southern Guyana.

An eradication program for feral goats was carried out on Kangaroo Island, Australia, using a range of methods including helicopter surveys, Judas goats, walking transects, camera trapping, scat count transects, and opportunistic community observations (Southgate et al., 2022). This raises an important question, which we now consider.

11.5.2 When Is a Species Extinct?

A major question in dealing with an invasive species is when can one be sure that a species is extinct especially when monitoring efforts have not wielded any presence of the animal for a while? A decision needs to be made when to terminate an eradication program, as it is difficult to evaluate the costs if total eradication is wrongly assumed. More recently, there has been a shift in the research to try and deal adequately with uncertain observations (e.g., blurry photographs) that in the past were either ignored or included as certain, rather than using a probabilistic approach.

Thompson et al. (2013) noted that estimating the probability that a species is extinct/extant based on historical sighting records is important when deciding how much effort and money to invest in conservation policies. Their aim was to infer

the state of a species by calculating the probability that a species is extant given the sighting data using Bayesian methods. The problem has several random variables, including the probabilities of an observation, given the species is extinct/extant especially when there may be false observations. A distinction is made between confirmed observations and uncertain observations, and possible false sightings need to be allowed for.

The *true* state σ_t of a species in year t over T years is $\sigma = (\sigma_1, \dots, \sigma_T)'$, where σ_t can only take two values, either χ for extant or e for extinct. The corresponding *observed* sighting state s_t over T years is $s = (s_1, \dots, s_T)'$, which could include one or a combination of specimens, photographs, sounds, observations, reports, surveys, etc. If the species is actually extant in year t ($\sigma_t = \chi$), it must be extant in all previous years. Similarly, if the species is extinct in year t , it must be extinct in all subsequent years. Let T_N denote the year of the last confirmed (true) positive sighting. The only possible states of nature over the T year period are then either $\sigma_t = \chi$ for $t = 1, \dots, T$, or $\sigma_{j-1} = \chi$ and $\sigma_j = e$. This gives us either $\chi = (\chi, \chi, \dots, \chi)$ or $e_j = (\chi, \chi, \dots, \chi, e, \dots, e)$, where the first e entry is in the j th position. The state of χ is equivalent to the species being extant at time T , which we denote by the event X_T , and e_j is equivalent to the species becoming extinct (as opposed to being extinct) in year j , which we denote by the event E_j .

We require the conditional probability

$$\Pr(X_T \mid s) = \frac{\Pr(X_T \text{ and } s)}{\Pr(s)} = \frac{\Pr(s \mid X_T)\Pr(X_T)}{\Pr(s)}.$$

Similarly, we can write

$$\Pr(E_j \mid s) = \frac{\Pr(s \mid E_j)\Pr(E_j)}{\Pr(s)},$$

for $j = T_N + 1, \dots, T_N$. The events X_T and E_j , $j = T_N + 1, \dots, T$, are mutually exclusive and exhaustive. By the law of total probability,

$$\Pr(s) = \Pr(s \mid X_T)\Pr(X_T) + \sum_{j=T_N+1}^T \Pr(s \mid E_j)\Pr(E_j).$$

Assuming independent sighting, an expression was obtained for $\Pr(X_T \mid s)$ and also when sightings are definite or uncertain. Lee (2014) showed how one of the above methods can be implemented in an Excel spreadsheet and showed how to include survey effort to the model as an additional discrete variable. He also included an extension to allow for uncertainty in the prior belief that the species is extant.

Ramsey et al. (2009) considered the same question with regard to the eradication of feral pigs from Santa Cruz Island, California. Following completion of the hunting phase in a zone, either eradication (E) was successful ($E = 1$) or pigs still persisted ($E = 0$), and probabilistic support for one of these was sought as there was

imperfect detection. A Bayesian framework was used, and it involved information on the number of pigs removed per unit time (1 week) and distance traveled per unit time for aerial and ground hunting. This was used to estimate the probability of detection (and removal) p and the population size N prior to the commencement of each hunting method with a catch-effort model. This model expressed $\text{logit}(p)$ as a linear regression in model hunting effort.

Rout et al. (2014) also considered the same question of determining extinction in the context of managing an invasive red fox eradication program on Phillip Island, Australia. If extinction is not done properly, there can be a rebound in population growth. Five methods had been used to find and remove the foxes: leg-hold trapping, baiting, hunting, spotlighting, and den searches. Let p_{it} be the removal probability for method i in year t ; then using the Poisson model with effort f_{it} , we have

$$p_{it} = 1 - \exp[-a_i f_{it}],$$

where a_i is the estimated “effectiveness” and f_{it} is the amount of effort in person-hours. For hunting and spotlighting, some foxes were only sighted and not removed. In this case, the probability of sighting a fox was

$$p_{it}^* = 1 - \exp[-b_i f_{it}],$$

where b_j was the method’s estimated effectiveness in sighting foxes. For each management method i , the number of removals in each year t was modelled as a binomial process:

$$n_{it} \sim \text{Binomial}(p_{it}, \lambda N_{t-1}),$$

where λ is the annual growth rate of the population and N_{t-1} is the number of foxes in the population at the start of time t . A similar model was used for the number of sightings (without removal). A Bayesian approach was used with priors on the parameters, and Markov chain Monte Carlo sampling in **OpenBUGS** was carried out. The authors gave a number of references to previous eradication programs.

11.5.3 Fisheries

A major application of catch-effort methods is in fisheries, where estimating the catch-per-unit effort can be more difficult as it can, for example, involve individual anglers or individual boat catches and fleets and landing catches. Thorson and Berkson (2010) considered the following model for varying catchability changing by time and population density. If k_t is the Poisson coefficient of catchability, the authors considered the density dependent model

$$k_t = k_0 \exp[t\beta] \bar{N}_t^{-\alpha},$$

where \bar{N}_t is average abundance in year t and k_0 is the base catchability coefficient. Here $k_t = k_0$ when $\alpha = 0$ and $\beta = 0$. The authors investigated various submodels using simulations and concluded “that imputation of catchability parameters from similar stocks can be useful to develop estimates of annual catchability that are more accurate than the estimates implied by assuming constant catchability.” It is important that catchability is controlled as it can increase with improved fishing efficiency, which can lead to overfishing. A similar “power” model was used by Hillary, Ward et al. (2013) to investigate hyperstability, where catch per unit effort remains high as fish density declines. Many factors affect the variability of catch-effort data (Ward, 2008). The question of whether fish learn to avoid anglers was considered by Wegener et al. (2018).

Comesaña and Guerra (2019) used the Gómez-Muñoz model to estimate catch and effort in the artisanal fishery for the squid *Loligo vulgaris* in North-West Spain. It was based on interviews of professional and sport fishermen, fish sellers, members of fishermen’s associations, and owners of bars who have fished or purchased the fish during 2011–2015.

Hartill et al. (2020) noted that there is a paucity of information used to inform the management of the highly variable recreational fishing. Many developed nations require commercial fishers to submit catch/effort return data that are used to inform stock assessment models and the setting of catch limits. However, this is not generally required by recreational fishers so that most of the available information on the scale and nature of recreational catch and effort has come from intermittent and labor-intensive surveys that provide only fragmented insights. Alternatively, digital cameras offer a more cost-effective means of continuously monitoring trends in recreational fishing effort. The authors reviewed early case studies using time-lapse camera systems and discussed in detail practical aspects of the methodology. They suggested the use of more traditional methods as well as using cameras.

Much of the literature is about commercial fishing where the population is open, and there is natural mortality as well as mortality from fishing are operating. As our concern is with closed populations, we refer to the earlier work on open populations in Seber (1982, chapter 8). A major change since has been the development of hierarchical Bayes’ models, for example, Zhou et al. (2008, prawns)

11.6 Summary

This chapter is a continuation of the previous chapter relating to removal models but more focused on the use of effort data. Here, the emphasis is on models involving catch-per-unit effort. A suite of eight models with four homogeneous and four heterogeneous models are introduced and considered individually. The use of estimating functions and including coverage is considered for some models.

A variety of linear regression models using both Poisson and Binomial catchability coefficients are described that are useful visually. Underlying model assumptions

are discussed, and a few variable catchability models are given as examples. The question of when a species is extinct is considered. Finally, fishery examples are referred to briefly as the main application is generally to open populations, which is not considered (cf. Seber, 1982, chapter 8).

Chapter 12

Capture-Recapture: Frequentist Methods



12.1 Introduction

Capture-recapture methods have a long history and have been extensively used for studying animal populations and more recently for human populations. They were originally used for closed populations and then extended to open populations where there is birth, death, and migration (cf. Seber & Schofield, 2019). There are a large number of research papers on closed populations, and we endeavor to consider them in groups of models, although there is some overlap between them. Because of the amount of material, we have split the methods for closed populations into two chapters, frequency, and Bayesian capture-recapture methods for both ease of exposition and chapter size. However, the split is somewhat arbitrary as Bayesian models are generally well integrated into population studies.

A wide variety of marks and tags are used to identify an individual uniquely, which are discussed in detail in Seber and Schofield (2019) along with methods of handling tag loss. Also natural marking such as spots or stripes can be used as unique tags along with photographic “captures,” for example, felids such as tigers, leopards, jaguars, and ocelots. Such surveys are typically conducted over short period of 39–60 days to ensure closure and involve photo-capturing both flanks of the cat in order to assign a unique identity to the individual. Also auxiliary data such as date, time, etc. is usually collected.

In this book, there is some overlap of this chapter with Chaps. 7 and 8, where distance methods sometimes make use of capture-recapture methodology, also mentioned below. A connection is also made with removal methods in Chap. 11, where captures can be regarded as “removals” and capture-recapture can be combined with spatial models as in Chap. 14, particularly with camera trapping, where resighting methods can also be used.

Closure can be achieved a number of ways. For example, if the population is changing slowly, sampling can be carried out over just a comparatively short period

of time. Birth is still possible, though sampling could take place at a nonbreeding time, or juveniles can be recognized and excluded from the sampling.

Sometimes, a population is totally enclosed, for example, on an island or in a fenced sanctuary, but birth and mortality can possibly take place. Mortality will depend on a species life span. However, geographic closure can be difficult to maintain. For example, grid trapping using live traps is a common method to study populations of small rodents. One problem frequently encountered is that traps in the outer sector of the grid tend to catch more animals than those near the center. This so-called edge effect is discussed in Sect. 2.4. Some of the effects of departures from the closure assumption were discussed by Kendall (1999). Some packages such as **R** and **MARK** have a test for closure. Covariate methods have become increasingly popular, though there are sometimes missing values and can be incompletely observed. Measurement errors for covariates can bias estimates (Hwang & Huang, 2003).

The general capture-recapture method for closed populations is readily described using just two samples to begin with, referred to as the Petersen method, which we now consider.

12.2 Two-Sample Capture-Recapture

Suppose we have a closed population of N animals and a random sample of M animals is captured without replacement. The captured animals are then given an individual identifying mark or tag and then released back into the population. After a short period of time, a second sample of size n is taken without replacement, and m , the number marked, is noted. Given M and n , m has a hypergeometric distribution and $E[m] = Mn/N$. A moment estimate of N can then be obtained by equating the sample proportion of marked with the population proportion of marked leading to

$$\hat{N} = Mn/m, \quad (12.1)$$

the so-called Petersen or Lincoln-Petersen estimate. This estimate and the procedure known as the Petersen method has been extensively studied by Seber (1982, chapter 3) and developed further in Seber and Schofield (2019, section 1.2), where Bayesian methods are also introduced. We also note in passing that if N is known but M is unknown, we can still use the same moment equation to get an estimate of M , namely, $\hat{M} = mN/n$. An example of this is in quality control where we have batch of light bulbs of known number and we want to estimate the number of defective bulbs by taking a random sample of bulbs from the batch and determining the number that is defective. Another example occurs with estimating occupancy, say of a plant. Here N sites are sampled, and M are the unknown number of occupied sites.

As the method is extended to several samples below, it is helpful to have a change of notation to allow for more generality. Let $M = n_1$ be the size of the first sample

and $n = n_2$ the size of the second sample. Looking at just the tagged animals, let m_{10} be the number of tagged caught in just the first sample only, m_{02} the number of tagged caught in just the second sample only, and m_{12} the number tagged caught in both samples. Then $n_1 = m_{12} + m_{10}$, $n_2 = m_{12} + m_{02}$, and $\widehat{N} = n_1 n_2 / m_{12}$. If $r = m_{12} + m_{10} + m_{02}$ is the number of different individuals captured, the estimator of the unmarked population is sometimes quoted and is given by

$$\widehat{N} - r = \frac{m_{10}m_{02}}{m_{12}}.$$

Also an almost unbiased modification, sometimes referred to as Chapman estimate of \widehat{N} , is

$$N^* = \frac{(n_1 + 1)(n_2 + 1)}{m_{12}} - 1.$$

An approximately unbiased estimator of its variance is Seber (1970)

$$v = \frac{(n_1 + 1)(n_2 + 1)(n_1 - m_{12})(n_2 - m_{12})}{(m_{12} + 1)^2(m_{12} + 2)}.$$

12.2.1 Underlying Assumptions

As we will be applying the Petersen method to human populations also, we need to revisit the assumptions underlying the method as applied to animal populations and highlight some of the material from Seber (1982, chapter 3), with human populations also in mind. One source for examples from human populations is Böhning et al. (2018). The assumptions are:

- (1) The population is closed.
- (2) All animals have the same probability of being caught in the first sample.
- (3) Marking does not affect the catchability of an animal.
- (4) The second sample is a simple random sample, that is, each of the possible samples has the same probability of being chosen.
- (5) Animals do not lose their marks or tags between samples.
- (6) All marks are reported on recovery in the second sample.

The Petersen method can be modelled several ways (Seber & Schofield, 2019, chapter 1). In particular, another approach is to use the first sample just to obtain a tagged proportion of the population, and it then does not matter how it was obtained provided the second sample is a simple random sample. The resulting statistical development can then be carried out conditional on fixed n_1 . Depending on how the second sample is taken, we might need to assume that there is random mixing of tagged and untagged or endeavor to achieve this, for example, in some fishery studies.

A variety of distributional forms for the joint distribution of the m_{ij} are available, depending on whether n_1 , n_2 , or both are random. For example, similar to Salasar et al. (2015), Chatterjee and Mukherjee (2016a) used an integrated likelihood model to provide several estimates of N . Prior distributions were also used. They added a behavioral effect, giving dependent samples (Chatterjee & Mukherjee, 2016b, 2018) and considered an adjusted profile-likelihood for this situation (Chatterjee & Mukherjee, 2021).

We shall now consider the effects of departures from the above assumptions. As these are discussed in some detail in Seber (1982, Chapter 3), we shall highlight just those pertaining to a closed population. We cannot test for assumption 1 about closure unless we take more samples (discussed later). Assumption 5 about mark loss will depend on the type of mark or tag used and can be incorporated into the model using tag loss methods such as double tagging (cf., Seber & Schofield, 2019, Chapter 3). We revisit this problem when considering two applications below. We note that assumptions 2 and 3 can affect assumption 4. This leaves us with dealing with variable catchability.

12.2.2 Variable Catchability

Departures from assumption (2) that all animals have the same probability of being caught in the first sample and from (3) that marking does not affect the catchability of an animal, under the title of variable catchability, often both occur at the same time, and the effects can be difficult to separate. We now consider these together.

For the j th member of the population ($j = 1, \dots, N$), let x_j be the probability that it is caught in the first sample; let y_j be the conditional probability that it is caught in the second sample, given it is caught in the first; and let z_j be the conditional probability that it is caught in the second sample, given it is not caught in the first. Then, assuming that the population consists of a random sample of N triples (x_j, y_j, z_j) from a trivariate distribution $f(x, y, z)$ (a key assumption), Seber (1970) showed that for the approximately unbiased estimate N^* given above

$$E[N^*] \approx k(N - n_1) + n_1,$$

where

$$k = (p_3 - p_{13})p_1 / [p_{12}(1 - p_1)],$$

$p_1 = E[x]$, $p_3 = E[z]$, $p_{12} = E[xy]$, and $p_{13} = E[xz]$. We see then that N^* is approximately unbiased when $k = 1$. Seber also discussed what happens to variances. In studying the values of k for departures from assumptions (2) and (3), it is useful to make the transformation

$$k = (B - p_1) / (1 - p_1), \quad \text{where } B = p_1(p_3 - p_{13} + p_{12}) / p_{12}.$$

Then $k = 1$ if $B = 1$.

Case 1: Catching Does Not Affect Future Catchability

Here assumption (3) holds but not assumption (2). We now have $y_j = z_j$ ($j = 1, \dots, N$) so that $p_{13} = p_{12}$, $p_3 = p_2$ ($= E[y]$), and

$$1 - B = 1 - (p_1 p_2 / p_{12}) = \text{cov}[x, y] / E[xy].$$

The bias can be severe as Seber found that if $f(x)$ is the uniform $[0, 1]$ distribution, for example, then $p_1 = \frac{1}{2}$, $B = \frac{3}{4}$, and $k = \frac{1}{2}$. The above equation was first studied by Junge (1963), who noted that $B = 1$ (and hence $k = 1$) when x and y are uncorrelated. Since we would generally expect the correlation to be positive, N would then be underestimated as $k < B < 1$, a common observation in studies. Jung also noted that x and y are uncorrelated if either x or y is constant so that $B = 1$ if at least one of the two samples is random, a point made with the epidemiological example described in Sect. 12.2.4, where the second sample was a random survey. This means that if the first sample is random, $B = 1$. Also variation in catchability due to, say, trapping selectivity could exist for both samples without introducing bias if the sources of selectivity in the two samples were independent. This lends support to the statement by several authors that the bias due to catchability can be reduced by using a different sampling method for each sample.

If catching for an individual does vary from sample to sample, this is essentially the heterogeneity model M_h described below. Alternatively, for the case of just two samples, one can use a logistic model (Alho, 1990).

Case 2: Zero Correlation Between Samples

If there is no correlation between the two samples as far as catchability is concerned (irrespective of the effects of trapping and handling), then x_j and (y_j, z_j) are uncorrelated, and $p_{13} = p_1 p_3$, $p_{12} = p_1 p_2$, and $k = p_3 / p_2$. Hence, $k = 1$ if and only if the average catchability of the tagged in n_2 is the same as the average catchability of the untagged in n_2 . Of course x_j and (y_j, z_j) are uncorrelated when x_i is constant. If the first sample is random, $k = 1$, and the second sample need not be random but can be highly selective, provided the selectivity was independent of mark status. We see then that if $k \neq 1$, we cannot be sure whether this departure is due to tagging affecting either future catchability or to variation in the inherent catchability of individuals, or to both. Methods for separating the two effects are described below.

With human populations, the overall sampling period is generally sufficiently short and, in the case of say a census (discussed below), sufficiently controlled so that population can be treated as closed for assumption 1. Unfortunately, the second assumption is not generally true for both animal and human populations as the “catchability” can vary from individual to individual depending on race, age, sex, etc. for humans. Assumption 3 may also be false as the catching or handling of animals may lead to some animals responding differently to their second capture because of what is referred to as being “trap-happy” (as they like the free trap-food) or “trap-shy” (as they did not like the process of being captured). This type of effect

also happens with humans. Since the first sample is just to mark or tag part of the population, assumption 2 is not critical if assumption 4 holds, which we exploit with the epidemiological example described below. However, variable catchability with assumption 3 not true can affect assumption 4 as simple random sampling may not apply because of the catching method.

As already indicated, the Petersen method has been applied to human populations such as census methods, discussed later, and also to various epidemiological populations to estimate, for example, the size of a particular subpopulation or prevalence of certain noncommunicable diseases. We consider this case in the next section. Before doing so, we mention an alternative approach to the two-sample case developed by Fienberg (1972) as a special case of the s -sample model discussed below. He viewed the capture-recapture model as a 2^s cross-classification contingency table with the absence or presence of an individual in the i th sample defining the categories for the i th dimension. The table is incomplete as there is a missing item, namely, the number in the cell corresponding to those not caught in any sample. It is analyzed by Fienberg (1972) using a log-linear model, which is also described for a closed population in Seber and Schofield (2019, section 10.2). The beauty of this approach is that one can use an analysis of variance method that incorporates main effects and interactions. It should be noted that this approach assumes that the model which describes the observed data also describes the count of the unobserved individuals.

Measure of Dependence

Let $x_{hi} = 1$ if the individual h is caught in sample i ($h = 1, \dots, N, i = 1, 2$) with probability P_{hi} , and 0 otherwise. Let

$$\mu_{11} = N^{-1} \sum_{h=1}^N E[x_{h1}x_{h2}] \text{ and } \mu_i = N^{-1} \sum_{h=1}^N E[x_{hi}].$$

Chao et al. (2008) introduced a measure of dependence

$$\gamma_{12} = \frac{\mu_{11}}{\mu_1\mu_2} - 1,$$

which is 0 if and only if the two samples are independent. The authors used their definition to describe various forms of correlation bias. We recall that the Petersen estimator is $\widehat{N} = n_1n_2/m_{12}$ and

$$N = \frac{E[n_1]E[n_2]}{E[m_{12}]}(1 + \gamma_{12}) \approx E[\widehat{N}](1 + \gamma_{12}),$$

or

$$E[\widehat{N}] \approx N - \gamma_{12}E[\widehat{N}].$$

The correlation bias is approximately $-\gamma_{12}E[\hat{N}]$. It can arise in two situations: (1) the capture probability of an individual in one sample depends on its previous capture history, known as local dependence, and (2) heterogeneity among individuals. Even if the two samples are locally independent for any given individual, they may become globally dependent if capture probabilities are heterogeneous. As already noted above, no correlation bias arises if the equal-catchability assumption holds for the second sample, for example, when the second sample is a random sample.

In the case of just a behavioral response (model M_b described below) where initial capture affects future captures, the authors assumed that in the first sample, the capture probability is p for all animals. In the second sample, the probability of capturing a previously uncaptured individual is p , but capturing a previously captured individual becomes $c = \phi p$. We then find that $\mu_{11} = cp$, $\mu_1 = p$, $\mu_2 = (1 - p + c)$, and $\gamma_{12} = (\phi - 1)(1 - p)/[(1 - p) + \phi p]$. This means that the two samples are positively dependent (“trap happy”) if $\phi > 1$ and negatively dependent (“trap shy”) if $\phi < 1$.

In the case of heterogeneity (model M_h to be precise, described below), p_1 and p_2 are now random variables, and we have that the (p_{h1}, p_{h2}) are a random sample from a bivariate distribution $f(p_1, p_2)$. Then

$$\gamma_{12} = \frac{E[p_1 p_2]}{\mu_{p1}\mu_{p2}} - 1 = \frac{E[(p_1 - \mu_{p1})(p_2 - \mu_{p2})]}{\mu_{p1}\mu_{p2}} = \frac{\text{cov}[p_1, p_2]}{\mu_{p1}\mu_{p2}},$$

where $\mu_{pi} = E[p_i]$ denotes the average capture probability for sample i . If two different sampling schemes are used, for example, trapping and resighting, there will be little covariance and $\gamma_{12} \approx 0$.

The authors considered a natural extension to the above topic of dependency where interest is in estimating a shared population. Suppose there are two populations A and B , where the interest is in estimating the size N_{12} of $A \cap B$ by taking two samples from each population, A_i and B_i , giving overlap samples $A_i \cap B_i$ ($i = 1, 2$). A modified Petersen estimate of N_{12} and a corresponding measure of dependence and correlation bias were derived.

We now consider an alternative method of analysis where the capture probabilities also include covariates.

Nonparametric Method

Chen and Lloyd (2000) considered a population of N individuals each with measurable covariate vector \mathbf{x} , with density function $f(\mathbf{x})$, which includes such variables as age or weight in the mark-recapture experiment context or sighting distance and cluster size in the independent-observer line-transect survey context discussed in Sect. 8.11. We recall that m_{12} is the number caught (detected) in both samples, m_{10} in just the first sample, and m_{02} in just the second sample, so that $r = m_{10} + m_{02} + m_{12}$ is the number of different individuals detected. Let p_0 be the probability that an individual is not detected in either sample so that $p = 1 - p_0$ is detected in at least one sample. Let p_i be the probability of being detected

in sample i . Using detectability functions, the authors showed that the likelihood function $L(p_1, p_2, \alpha, N)$ is proportional to

$$\frac{N!}{(N-r)!} (p_1 - \alpha p_1 p_2)^{m_{10}} (p_2 - \alpha p_1 p_2)^{m_{02}} (\alpha p_1 p_2)^{m_{12}} (1 - p_1 - p_2 + \alpha p_1 p_2)^{N-r},$$

where α is a type of heterogeneity index and N is a function of α . Although we have one too many parameters, α can be estimated separately using a kernel method. The other estimates are essentially moment estimators with α replaced by its estimate. Asymptotic properties of $\hat{N}(\hat{\alpha})$ were obtained.

12.2.3 Two Observers with Time to Detection

We consider the problem of counting animals using two observers over time. Following Stanislav et al. (2010), suppose there are two time intervals of observation. Let p_{dji} be the probability of detecting an individual by observer j time over interval i , ($i = 1, \dots, t$), given it is available (e.g., present or observable), and let p_{ia} be the probability of being available in interval i . Consider a detection history for the two observers over the two intervals such as (11, 01), which denotes an individual detected by both observers in time interval one and detected only by the second observer in time interval two. This has probability

$$p_{11,01} = p_{a1} p_{d11} p_{d21} [p_{a2}(1 - p_{d12}) p_{d22}].$$

Similarly,

$$p_{10,00} = p_{a1} p_{d11} (1 - p_{d21}) [p_{a2}(1 - p_{d12})(1 - p_{d22}) + (1 - p_{a2})].$$

The 00 in the second time interval created the need for two components in the second time interval probability: the first component where we account for the possibility that the bird is available but is missed by both observers and the second where the bird is not available. For the numbers of the observable detection histories $\mathbf{x} = (x_{00,01}, x_{00,11}, \dots, x_{11,11})'$, with probabilities $\mathbf{p} = (p_{00,01}, p_{00,11}, \dots, p_{11,11})'$, we have the multinomial likelihood

$$L(N, \mathbf{p} \mid \mathbf{x}) = \frac{N!}{x_{00,01}! \cdots x_{11,11}!(N-n)!} p_{00,01}^{x_{00,01}} \cdots p_{11,11}^{x_{11,11}} p_{00,00}^{N-n},$$

where $p_{00,00} = Q$, the probability of not being caught during the entire experiment. Here

$$n = \sum_{\omega} x_{\omega}, \quad p_{00,00} = 1 - \sum_{\omega} p_{\omega} = Q,$$

where n is the total number observed and ω is a detection history. Using the usual conditional likelihood approach (Appendix A.3.2), we have

$$L(N, \mathbf{p} | \mathbf{x}) = L_1(N | Q) \times L_2(\{p_\omega\}),$$

where

$$L_1(N | Q) = \frac{N!}{n!(N-n)} (1-Q)^n Q^{N-n},$$

and

$$L_2(\{p_\omega\}) = \frac{n!}{x_{00,01}! \cdots x_{11,11}!} P_{00,01}^{x_{00,01}} \cdots P_{11,11}^{x_{11,11}},$$

with $P_\omega = p_\omega / (1 - Q)$. We first obtain the maximum likelihood estimators for the p_d and p_a using L_2 to obtain \widehat{Q} and then maximize L_1 with this estimate to obtain the maximum estimator of N given by

$$\widehat{N} = n / (1 - \widehat{Q}).$$

Standard errors and confidence intervals for all the populations estimates can be obtained using bootstrapping. The above method can readily be generalized to deal with at least two observers and more than two time intervals. The key assumption of the method is that the observer can keep track of individual birds without error, and when applying the method to birds, the authors recommended the use of at least four time intervals of equal length, perhaps five intervals of two minutes each.

12.2.4 Epidemiological Population: Two Lists

Capture-recapture methods have been used particularly for hard-to-reach human populations such as the homeless, sex workers, drug users, and individuals with subclinical disease. Some examples and methods are described by Chao (2015). It has also been used in ophthalmology to consider the prevalence of a range of conditions such as congenital cataracts and vision impairment (see the review of Ramos et al., 2020). Chaudhary et al. (2020) provided a multi-state occupancy modelling framework for the robust estimation of disease prevalence in multi-tissue disease systems (Sect. 4.18).

Hay and Richardson (2016) gave a review of the prevalence of drug use and noted the more recent use of a Bayesian approach to log-linear methods. They referred to four different lists of drug users, namely, treatment, police, prison, and probation data. By equating “being on list i ” with “being caught in the i th sample,” we can apply the capture-recapture methodology to lists. There is one difference; there is no time ordering of the lists. As with another example such as the population of

diabetics, there is usually information from various incomplete lists obtained from a number of resources such as hospital, doctors, and pharmaceutical records. One problem with the methods is the possibility of some dependence between lists, say through referrals. This has an effect on the interaction terms in log-linear modeling, and Jones et al. (2014) gave several models for handling this situation, as well as Ramos et al. (2020). The key assumptions for applying the traditional closed-population capture-recapture methods to lists are as follows:

- (1) The population is closed, i.e., the population remains unchanged.
- (2) Each member of the population has the same probability of being on a given list.
- (3) The lists are independent. For example, if p_i is the probability of “capture” in list i and p_{12} is the probability of capture in both lists, then the lists are independent if $p_{12} = p_1 p_2$.
- (4) Individuals are matched correctly.

In practice, the first assumption will be satisfactory if the time frame is sufficiently short. Unfortunately, the second and third assumptions will often be false. For example, a list of diabetic patients compiled from doctors’ records will not be independent of a hospital’s admissions list if some doctors always send their patients to that hospital. Furthermore, if a list comes from people who join a particular diabetes support group, there will be those who are joiners and those who are not. The probability of being on such a list will therefore vary from person to person so that there is also heterogeneity.

Assumption 4 will generally be false because there may be patients who give false information deliberately or accidentally or omit some items. There will be transcribing errors such as misspelling a single letter of a complicated name, recording the wrong age, etc. so that the same patient on two lists may be recorded as two different patients. One approach to dealing with departures from assumption 4 is to use probabilistic matching (cf., Howe, 1985; Jaro, 1989, 1995). For every pair of lists, the method involves giving a weight to each unmatched pair of individuals and operates somewhat like sequential sampling. If the weight is large enough, we can decide for a match; if the weight is small enough, we can decide for a nonmatch; otherwise, we are undecided. For example, two very unusual names differing by just one syllable but matching on all the other information would have a large weight. This approach requires a sophisticated statistical package such as AUTOMATCH (Jaro, 1995).

Two lists are sometimes independent, having been constructed using independent methods. For example, one list may come from official sources, while the other may come from an independent sample survey, which satisfies assumption 3, being a simple random sample. In this case, assumption 2 is true for the second list, which is a key assumption. If assumption 4 is not true, then the Petersen estimate cannot be used. Unfortunately, assumption 4 is usually false for at least some of the population as there will be some incorrect matching. The breakdown of the assumption has received some attention in the literature under the title of record-linkage methods such as using probabilistic and deterministic matching, (e.g., Dusetzina et al., 2014, Chapter 4). A matching process is described below.

Seber in Seber et al. (2000) provided a new method of matching based on the idea of tags and tag loss. Each patient on a list has an identifying string of information such as social security number, name, age, sex, address, date of birth, and so on. We now divide this information into two substrings in such a way that if either substring matches or both match with the two lists, the individual is identified uniquely. It is also assumed that having an error in one substring is independent of there being an error in the other substring. The validity of this assumption depends on how we form the substrings. For example, if there are items that a patient is likely to lie about, then they should be in the same substring. If there are items that are more likely to be mistyped because they are complicated, then perhaps they should be together. A further key assumption is that any errors do not lead to accidental matches.

Once we have defined the two substrings (we shall refer to them as tags), each individual has two tags A and B . When a substring is incorrect, the individual “loses” that tag (or perhaps, more appropriately, the tag deteriorates). Assuming the tags are lost independently, we can now use the tag loss methodology of Seber and Felton (1981) that is used for animal populations and apply it here. The computer records are split into the two substrings in such a way that individuals can be matched accurately in the absence of errors using just A alone or B alone. It was also assumed that errors in either A or B will not result in false matches. This should be the case if the record is split in such a way that correct information in each substring uniquely identifies the individual. It was also assumed that errors in either of A or B will not result in false matches, which should be the case if the record is split in such a way that correct information in each substring uniquely identifies the individual.

The two-sample approach, assuming the first three assumptions to be true (the second sample was a simple random sample), was applied by Seber et al. (2000) to finding the size of a diabetic population. The first list came from doctors' records and the second list from a random survey. It was found that the tag loss rates were quite high. We summarize the mathematics from that paper. But first some notation—assuming that all individuals have the same probability $p_i = 1 - q_i$ of being on list i ($i = 1, 2$), let

$$N = \text{diabetic population size},$$

$$\phi_X = \text{probability that tag } X \text{ is retained by an individual } (X = A, B),$$

$$\phi_{AB} = \text{probability that an individual retains both tags},$$

$$\pi_X = \phi_X^2 \quad (X = A, B, AB).$$

We are assuming that ϕ_X is the same for both lists and that ϕ_{AB} and ϕ_X are the same for all members of m_{12} (i.e., no heterogeneity of tag retention). For the individuals in m_{12} (on both lists), let

$$m_{AB,AB} = \text{number retaining both tags on both lists (abbreviated to } m_{AB}),$$

$$m_{AB,AO} = \text{number retaining tag } A \text{ on both lists and } B \text{ on only list 1},$$

$m_{AB,OB}$ = number retaining tag B on both lists and A on only list 1,

$m_{AO,AO}$ = number retaining just A on both lists,

$m_{OB,OB}$ = number retaining just B on both lists,

$$m_A = m_{AB,AO} + m_{AO,AB} + m_{AO,AO} \text{ (number matching only on tag } A\text{),}$$

$$m_B = m_{AB,OB} + m_{OB,AB} + m_{OB,OB} \text{ (number matching only on tag } B\text{),}$$

$$m_T = m_{AB} + m_A + m_B \text{ (members of } m_{12} \text{ observed to match).}$$

We now assume tag independence, so that $\phi_{AB} = \phi_A \phi_B$ and, consequently, $\pi_{AB} = \pi_A \pi_B$. The joint distribution of $m_{AB}, m_{AB,AO}, m_{AO,AB}, m_{AO,AO}, m_{AB,OB}, m_{OB,AB}, m_{OB,OB}$, and $m_{12} - m_T$, conditional on m_{12} , is multinomial with respective cell probabilities $\phi_{AB}^2, \phi_{AB}(\phi_A - \phi_{AB}), \phi_{AB}(\phi_B - \phi_{AB}), (\phi_A - \phi_{AB})^2, \phi_{AB}(\phi_B - \phi_{AB}), \phi_{AB}(\phi_B - \phi_{AB}), (\phi_B - \phi_{AB})^2$, and θ_{OO} . Pooling cells 2–4 and then 5–7, we get

$$\begin{aligned} & \Pr(m_{AB}, m_A, m_B, m_{12} - m_T \mid m_{12}) \\ &= \frac{m_{12}!}{m_{AB}! m_A! m_B! (m_{12} - m_T)!} \theta_{AB}^{m_{AB}} \theta_{AO}^{m_A} \theta_{OB}^{m_B} \theta_{OO}^{m_{12} - m_T}, \end{aligned} \quad (12.2)$$

where $\theta_{AB} = \pi_A \pi_B$, $\theta_{AO} = \pi_A(1 - \pi_B)$, $\theta_{OB} = \pi_B(1 - \pi_A)$, and $\theta_{OO} = (1 - \pi_A)(1 - \pi_B)$. To find the distribution of m_{12} , several capture-recapture models can be used. However, since the sample (list) sizes n_i are random, we can use a multinomial distribution, namely, (Seber & Schofield, 2019, Section 1.2.2)

$$\begin{aligned} & \Pr(m_{12}, m_{10}, m_{02}, N - r \mid N) \\ &= \frac{N!}{m_{12}! m_{10}! m_{02}! (N - r)!} (p_1 p_2)^{m_{12}} \\ & \quad \times (p_1 q_2)^{m_{10}} (q_1 p_2)^{m_{02}} (p_1 p_2)^{m_{12}} (q_1 q_2)^{N-r}, \end{aligned} \quad (12.3)$$

where $r = m_{10} + m_{02} + m_{12}$. It transpires that, in practice, it does not matter whether the n_i are treated as both fixed, both random, or just n_1 fixed. If we set $\pi_A = p_1$, $\pi_B = p_2$, and $m_{12} = N$ in (12.2), we see that models (12.2) and (12.3) are essentially the same. This means that the estimates for these two models have parallel forms.

To obtain estimates of m_{12} , π_A , and π_B , we can maximize the likelihood (12.2) directly. However, we can obtain estimates that are close to the maximum likelihood estimates by setting equal to zero the derivatives of the log likelihood with respect to π_A and π_B and the first difference with respect to m_{12} . These estimates turn out to be the same as the moment estimators and are obtained by equating each random variable to its expected value. These are given by:

$$m_{AB} = m_{12} \pi_A \pi_B,$$

$$\begin{aligned}m_A &= m_{12}\pi_A(1 - \pi_B), \\m_B &= m_{12}(1 - \pi_A)\pi_B,\end{aligned}$$

which have solutions

$$\begin{aligned}\hat{\phi}_A &= (\hat{\pi}_A)^{1/2} = (m_{AB}/m_{\cdot B})^{1/2}, \\ \hat{\phi}_B &= (\hat{\pi}_B)^{1/2} = (m_{AB}/m_{A \cdot})^{1/2}, \\ \hat{m}_{12} &= m_{A \cdot}m_{\cdot B}/m_{AB} = m_T\{1 + (m_A m_B/(m_T m_{AB}))\},\end{aligned}\tag{12.4}$$

where the ϕ_X are the tag retention probabilities, $m_{A \cdot} = m_{AB} + m_A$ and $m_{\cdot B} = m_{AB} + m_B$. Using a similar approach, we get the usual Petersen estimate of N from (12.3), namely, $n_1 n_2 / \hat{m}_{12}$, which, combined with the Eq. (12.4), gives us an estimate of N , namely,

$$\hat{N}_{AB} = n_1 n_2 / \hat{m}_{12}.\tag{12.5}$$

Using the so-called delta method (Appendix), we have, from Seber and Felton (1981, equation (12) with $\theta_A \rightarrow \pi_A$, etc.), the following approximate expression for the variance of \hat{N}_{AB} :

$$\begin{aligned}\text{var}[\hat{N}_{AB}] &\approx E_{n_1}\{\text{var}[\hat{N}_{AB}|n_1]\} \\ &\approx \frac{N}{p_1 p_2} \left\{ \frac{(1 - \pi_A)(1 - \pi_B)}{\pi_A \pi_B} + 1 \right\}.\end{aligned}$$

Estimating p_i by n_i/\hat{N}_{AB} , the above variance can be estimated by

$$\hat{v}_{AB} = \frac{\hat{N}_{AB}^3}{n_1 n_2} \left\{ \frac{(1 - \hat{\pi}_A)(1 - \hat{\pi}_B)}{\hat{\pi}_A \hat{\pi}_B} + 1 \right\}\tag{12.6}$$

If the tags are not independent, Seber et al. (2000) gave the following estimator, which is approximately unbiased if ϕ_A and ϕ_B are large enough, namely,

$$N_{AB}^* = \frac{(n_1 + 1)(n_2 + 1)}{(m_T + 1)} \left\{ 1 - \frac{m_A m_B}{m_T(m_{AB} + 1)} \right\} - 1.$$

Its properties and confidence intervals are discussed in their paper. Also considered is what happens if the tag loss is ignored and if there is heterogeneity. For other medical type examples, see Seber (1982, Section 12.8.3) and IWGDMF (1995a,b).

Lee (2002) considered the same problem of mismatching lists by fitting a non-saturated Poisson log-linear model to the non-missing cells of a contingency table and then using the model to predict the missing cell total. He gave an example where quite small error rates (say about 5% of records being in error) can lead to severe

overestimation of the population size by factors of 100%. He also considered more than two lists.

12.2.5 Epidemiological Population: Several Lists

Although this topic belongs to multiple capture-recapture models considered later, it is easier to mention it briefly here. Lee et al. (2001) extended the above theory to more than two lists that were not necessarily independent. They developed the theory for three and four lists, which indicates how the theory can be generalized to more than four lists, and considered the calculation of standard errors. Sutherland and Schwarz (2005) considered a four list log-linear model using quasi-likelihood estimation and variance estimation and compared various estimators.

Köse et al. (2014) considered the situation where there is an initial list followed by several repeat lists. For example, from the information collected in the surveillance of drug users, list 1 could be a list supplied by the police, whereas source 2 could be lists supplied by hospitals, which also provide the number of repeated contacts with drug users during their treatment. Here source 2 is a set of laboratories that identify a certain patient repeatedly. The information collected is the frequencies of patients listed in the second source that are listed for the two cases, listed and not listed in sample 1. By truncating the data, this leads to a Petersen-type estimator that is calculated using an EM algorithm. It assumes that there are no mismatches and does not use all the data but provides simple population estimators.

King et al. (2014) investigated the prevalence of injecting drug users in England using four lists from (a) probation, (b) drug intervention program (DIP) prison assessments, (c) drug treatment, and (d) DIP community assessments. Individuals were labelled, forename initial, surname initial, and date of birth. All four must agree for inclusion in the study. It was apparently assumed that there were no mismatches as the data sources were reliable. A log-linear model was used with covariates, and a Bayesian analysis was performed. In more general multi-list experiments, there are usually matching errors, which if ignored can have a serious effect on population estimates. Lee and Seber (2001) considered this situation for two, three, and four lists when the lists were not necessarily independent.

12.2.6 Dual Record System

Two problems occur in censusing a human population, the possible presence of undercount when people are missed and the possibility of overcount due to doubling up when counts are divided up into different categories that may overlap. Also many things can go wrong such as lost census forms, bad or missing addresses, and confused or suspicious respondents, all leading to an undercount. Despite follow-up

methods, the problem persists. In the USA, for example, considerable evidence that has accumulated over decades supports the existence of a systematic undercount in the decennial census that differentially affects different population groups, e.g., blacks and nonblacks (Anderson & Fienberg, 1999, chapter 4). One approach to evaluating the undercount applies a demographic analysis using vital data of birth, death, and migration records (Robinson et al., 1993). An earlier bibliography on capture-recapture modelling with application to census undercount adjustment was given by Fienberg (1992).

On the surface, the two-sample Petersen method can be used to estimate population size. This has led to there being a census followed up by a second sample, namely, a so-called post-enumeration survey (PES). In the census literature, the method is referred to as the dual record system or dual system estimate (Hogan, 2003). This two-sample method can be used, provided the underlying assumptions for the method are satisfied, especially the independence of the two samples. However, sample dependence can occur in two main ways.

Firstly, capture in a census may affect capture in the subsequent PES so that the second sample is not a simple random sample and there is a list dependence. Even if stratification is used, which can be helpful if there is heterogeneity, the same problem can occur with each substratum instead of each individual. Secondly, heterogeneity can exist between individuals, so that even if the two lists are independent for an individual, the lists can become dependent through heterogeneity. Some earlier papers that endeavored to relax the independence assumption were given by Chao and Tsay (1998).

Chao and Tsay (1998) mentioned another method that uses a third sample, for example, Zaslavsky and Wolfgang (1993), who used it to assess the level of dependence. For example, in the 2020 US census, the first sample selected 10,050 blocks in 50 states and the District of Columbia plus 400 blocks in Puerto Rico. The PES recapture listing used 541,000 housing units in 50 states and District of Columbia and 24,000 housing units in Puerto Rico. A further subsample was made using 180,000 housing units in 50 states and District of Columbia plus 8,000 housing units in Puerto Rico. Assuming independence within each individual for the three lists, Darroch et al. (1993) used a log-linear model approach with the Rasch model (Rasch, 1961). They used the 2^3 incomplete contingency table method described above under the assumption of no second-order interaction.

Alho et al. (1993) modeled the heterogeneity using a logistic model under the assumption of independence within individuals. Agresti (1994) considered the application of the Rasch model to capture-recapture models, and his results can be readily adapted to census undercount estimation as well. More recent papers were referred to by Chatterjee and Bhuyan (2017), who developed a model for some dependence between two lists. They introduced a paired variable (Y, Z) , where Y_i and Z_i denote, respectively, the list 1 and list 2 inclusion status of the i individual belonging to the population of size N . It is assumed that (Y_i, Z_i) , for $i = 1, \dots, N$, are *iid* bivariate random variables distributed as (X_1, X_2) with probability $1 - \alpha$ and (X_1, X_1) with probability α , where X_1 and X_2 are independently distributed Bernoulli random variables with parameters p_1 and p_2 , respectively. If

$p_{yz} = \Pr(Y = y, Z = z)$ for $y, z \in \{0, 1\}$, then $p_{11} = \alpha p_1 + (1 - \alpha)p_1 p_2$, $p_{10} = (1 - \alpha)p_1(1 - p_2)$, etc. Further models and details are given by the authors.

Chatterjee and Mukherjee (2018) developed a new integrated likelihood for estimating population size in the dependent case. With heterogeneity one can post-stratify the population and model the possible interdependence under suitable assumptions on the dependence structures of the strata (Chatterjee & Bhuyan, 2020a). Stratification is considered further in the following section.

There remains one further problem already alluded to and mentioned in the previous epidemiological example. When we can obtain two independent lists of names and addresses, matching errors can arise when matches are treated as nonmatches. A different approach that uses either the original census plus the PES, or three samples, has been given by Chao and Tsay (1998) using the concept of sample coverage, which we refer to in Sect. 12.8. The choice of model depends on the sampling methods used, and there are variations among countries to allow for count errors. For example, Chipperfield et al. (2017) described the methods used in the USA and UK compared with a method used in Australia, when it comes to estimating the size of population subgroups (e.g., race, sex, age, and geography).

12.2.7 Some Individuals Stratified

Premarathna et al. (2018) discussed the situation in a fishery population where there are two categories such as sex, and some individuals are difficult to sex. If individuals were correctly sexed, one could use the Petersen method to estimate the size of each population separately. However, in this situation, a subsample of the captured fish at each sample occasion is selected, and additional and often more costly measurements are made, such as sex determination through sacrificing the fish. The data then consisted of two types of marked animals: animals whose value of the stratification variable is unknown (here sex), and subsamples at each occasion where the value of the stratification variables were determined. The authors developed new methods for these types of experiments, and given the relative costs of sampling for a simple capture and for processing the subsample, optimal allocation of effort for a given cost could be determined. They applied the methods to estimating the size of the walleye population in Mille Lacs Lake, Minnesota, as follows.

The population is stratified into K categories with the following labeling: O (animal is not captured), U (animal is captured but not stratified), and C (animal is captured and identified as category C), where $C \in \{C_1, \dots, C_K\}$. For example, with males and females, we have $C \in \{M, F\}$. The category label is used for each of the two samples, such as UO (caught but not stratified in the first sample and not captured in the second sample), OC (not captured in the first sample but captured in the second sample and stratified as category C), and OO unobservable, giving seven labels altogether. Since there are K categories for each of CO , OC , and CC , plus the four UO , OU , UU , and OO , there are $3K + 4$ total capture histories.

The probabilities of the capture histories are similarly labeled, e.g., P_{OC} , and P_{OO} , which all sum to 1, with counts n_{UC} , etc., and 7 of each. For $t = 1, 2$ let

p_{tC} = capture probability of animals in category C at sample occasion t ,

λ_C = proportion of category C animals in the population, $\sum_C \lambda_C = 1$,

θ_t = subsample proportion at sample occasion t ,

N = total population size,

N_C = population size of those in category C ; $N_C = N\lambda_C$,

giving us $3K + 2$ parameters to be estimated, with the constraint $\sum_C \lambda_C = 1$. We then have a multinomial distribution with index N , $3K + 4$ cell probabilities P_{XY} , and cell observations n_{XY} , where $n_{OO} = N - n$. We now give the P_{XY} , namely,

$$P_{UO} = \sum_C \lambda_C p_{1C}(1 - \theta_1)(1 - p_{2C}),$$

$$P_{UU} = \sum_C \lambda_C p_{1C}(1 - \theta_1)p_{2C},$$

$$P_{OU} = \sum_C \lambda_C (1 - p_{1C})p_{2C}(1 - \theta_2),$$

$$P_{CO} = \lambda_C p_{1C}\theta_1(1 - p_{2C}),$$

$$P_{CC} = \lambda_C p_{1C}\theta_1p_{2C},$$

$$P_{OC} = \lambda_C (1 - P_{1C})P_{2C}\theta_2,$$

$$P_{OO} = \sum_C \lambda_C (1 - P_{1C})(1 - P_{2C}),$$

where the seven probabilities sum to 1. Logistic regression with covariates were used for the p_{tC} , and maximum likelihood estimates with variance estimates were calculated. The total cost of the experiment was considered as a linear function of sample sizes and an optimal strategy determined using numerical optimization.

12.2.8 Detection Model

Givens et al. (2015) used the two-sample method with detections of two independent observers teams for bowhead whales when there is incomplete detection, which we sketch briefly. Their model, which we sketch briefly, estimated the relationship between detection probabilities and covariates associated with the sightings. They

began with the two-sample basic model with $p_i = p$ ($i = 1, 2$), but conditioned on r , namely,

$$f(m_{10}, m_{02}, m_{12} | r) = \frac{r!}{m_{10}! m_{02}! m_{12}!} (pq)^{m_{10}} (qp)^{m_{02}} (p^2)^{m_{12}} (P)^{-r},$$

where $P = (1 - q^2)$. Now allowing for incomplete detectability and assuming independent individuals, let p_{hx} be the probability that an individual h has capture history x_h with observed catch history y_h . Here, conditioned on r , possible capture histories are (10), (01), and (11) with respective probabilities $p_h q_h / P_h$, $q_h p_h / P_h$, and p_h^2 / P_h , where $P_h = 1 - q_h^2$. The authors assumed that $\text{logit}(p_{hx}) = \boldsymbol{\beta}' \mathbf{z}_h$, where \mathbf{z}_h are appropriate covariates, and the likelihood function for individual h is $L(\boldsymbol{\beta} | y_h, \mathbf{z}_h)$. A weighted likelihood was introduced to allow for some specific problems listed for bowhead whales, namely,

$$L(\boldsymbol{\beta} | \{y_h, \mathbf{z}_h\}) = \prod_{h=1}^r L_h(\boldsymbol{\beta} | y_h, \mathbf{z}_h)^{\lambda_h}.$$

where the λ_h are the weights. The above is an oversimplification of the method used as, for example, the authors discussed in great detail the methods for identifying captures, matching sightings, and dealing with some inconsistencies.

In concluding this section, we realize that two samples do not give much data and some of the previous assumptions may not hold. We therefore need more than two samples, described next along with regression methods (Seber, 1970, Section 5, and Seber, 1982, Section 4.1.3). We begin with three samples.

12.3 Triple-Catch Method

The next step up from two capture-recapture samples is to consider three samples, sometimes known as the triple-catch method, referred to briefly above. It is also referred to as a triple-record system. For example, in the general census coverage evaluation study mentioned in Sect. 12.2.6, a third source of information, known as administrative list supplement (ALS), prepared prior to census operation, is often used for improving post-enumeration survey (PES) coverage. If all the assumptions underlying the two sample method apply to three samples except perhaps for closure, then the extra sample can provide an estimate of the probability of survival between samples one and two.

When the population is closed, Cormack (1966) provided two methods for testing the hypothesis of constant catchability when tagging does not affect future catchability. In using three samples, the first sample provides an identifiable population, or else, one can take two samples from a known population. Both methods are described in detail by Seber (1982, Section 3.2.2(3)), the first using

sampling efforts. Unfortunately, both tests are rather insensitive unless n_1 is large, and a significant proportion is captured in the second and third samples.

In the case of census methods, Darroch et al. (1993) used three samples to assist dealing with some of the problems due to variable catchability and the resulting undercount. Covariates using fixed effects were used. Gold et al. (2015) used three lists from refugee services, public schools, and a local health plan, to estimate refugee populations.

Chatterjee and Bhuyan (2020b) considered the estimation of population size from a dependent triple-record system that is related to the commonly encountered problem of undercoverage/underrepresentation, assuming that the available data are free from any erroneous inclusion. They noted that it is quite natural that some individuals behave independently over the three different capture attempts, while some behavioral dependence may exist for the rest of the population. Let α_1 , α_2 , and α_3 be the proportion of individuals in the population for whom pairwise dependencies between the lists/samples (1 and 2), (2 and 3), and (1 and 3) exist, respectively, and α_4 the proportion for a triple dependency. If $\alpha = \sum_1^4 \alpha_i$, then proportion $1 - \alpha$ behaves independently (no dependence) over the three lists.

Denoting 1 for capture and 0 for noncapture, we have a triplet for an individual such as (101) of which we have, say, y_{101} such individuals denoting capture in samples 1 and 3 and not in 2. The latent capture status x_{ih} takes the value 1 or 0, denoting the presence or absence of the h th individual ($h = 1, \dots, N$) for the i th list ($i = 1, 2, 3$). x_{1h} s, x_{2h} s, and x_{3h} s are independently distributed Bernoulli random variables with parameters p_1 , p_2 , and p_3 respectively, for all $h = 1, \dots, N$. Here p_i refers to the capture probability of a causally independent individual in the i th list.

Let (Y_h, Z_h, W_h) denote a general triplet of binary (Bernoulli) variables representing capture states for individual h ($h = 1, \dots, N$), where N is the population size. Then, for the proportion $1 - \alpha$, individuals who have complete independence, the triplet takes the value (x_{1h}, x_{2h}, x_{3h}) , where the x_{ih} are all different. If there is a dependency between samples 1 and 2 (first order interaction), we have for α_1 individuals ($x_{1h} = x_{2h}$) so that our triple becomes (x_{1h}, x_{1h}, x_{3h}) , and similarly for α_2 and α_3 . For a proportion α_4 with triple dependencies (second-order interaction), we have (x_{1h}, x_{1h}, x_{1h}) . This gives five possible values for (Y_h, Z_h, W_h) with three dependency pairs, one triple dependency, and one with no dependencies. Allowing for N , we have $2^3 - 1 = 7$ observations but eight unknown parameters, namely, three p_i parameters, four α_i parameters, and N . We need one constraint for estimation.

If the samples are ordered in time, then it can be argued that it is unlikely that the status in the third sample will depend on the first sample while dependent on the second sample. So, one may assume α_3 , which implies pairwise dependence between the first and third lists is absent. Here the x_{1h} s, x_{2h} s, and x_{3h} s are independently distributed Bernoulli random variables with parameters p_1 , p_2 , and p_3 , respectively, for all $h = 1, \dots, N$. This is referred to as the trivariate Bernoulli model TBM1.

For $y, z, w = 0$ or 1 , let

$$\Pr(Y = y, Z = z, W = w) = p_{yzw},$$

with marginal sample inclusion probabilities p_Y , p_Z , and p_W given by

$$\begin{aligned} p_Y &= p_{1..} = p_1, \\ p_Z &= p_{.1.} = p_2 + (p_1 - p_2)(\alpha_1 + \alpha_4), \\ P_W &= p_{..1} = p_3(1 - \alpha)\alpha_2 p_1(1 - p_2). \end{aligned}$$

We now have a multinomial distribution for the y_{ijk} with index N and cell probabilities p_{ijk} . If $\theta = (N, \alpha_1, \alpha_2, \alpha_4, p_1, p_2, p_3)'$, the likelihood for model TBM1 is

$$\begin{aligned} L(\theta) \propto & \frac{N!}{(N - y_{000})!} [(1 - \alpha)p_1 p_2 p_3 + \alpha_1 p_1 p_3 + \alpha_2 p_1 p_2 + \alpha_4 p_1]^{y_{111}} \\ & \times [(1 - \alpha)p_1 p_2(1 - p_3) + \alpha_1 p_1(1 - p_3)]^{y_{110}} \\ & \times [(1 - \alpha)(1 - p_1)p_2 p_3 + \alpha_2(1 - p_1)p_2]^{y_{011}} \\ & \times [(1 - \alpha)p_1(1 - p_2)(1 - p_3) + \alpha_2 p_1(1 - p_2)]^{y_{100}} \\ & \times (1 - \alpha)^{y_{101} + y_{010}} [p_1(1 - p_2)p_3]^{y_{101}} [(1 - p_1)p_2(1 - p_3)]^{y_{010}} \\ & \times [(1 - \alpha)(1 - p_1)(1 - p_2)p_3 + \alpha_1(1 - p_1)p_3]^{y_{001}} \\ & \times [(1 - \alpha)(1 - p_1)(1 - p_2)(1 - p_3) + \alpha_1(1 - p_1)(1 - p_3) \\ & + \alpha_2(1 - p_1)(1 - p_2) + \alpha_4(1 - p_1)]^{N - y_{000}}. \end{aligned}$$

In the context of epidemiology or public health with lists, generally the capture attempts are not time-ordered, and the dependence between first and third lists may be present. One possibility is to assume the second-order interaction is zero leading to model TBM2. The corresponding three marginal probabilities are then $p_Y = p_1$, $p_Z = \alpha_1 p_1 + (1 - \alpha_1)p_2$, and $p_W = \alpha_3 p_1 + \alpha_2 p_2 + (1 - \alpha_2 - \alpha_3)p_3$. The likelihoods for TBM1 and TBM2 and some probabilities are derived in the appendix of Chatterjee and Bhuyan (2020b). Maximum likelihood estimators of the parameters and asymptotic variances need to be found numerically. The method was applied to malaria incidence in the Netherlands and to a census coverage study on urban adult black males. An extensive simulation study was also carried out comparing the new methods with two other available ones, model M_{tb} described below and a log-linear model.

12.4 Several Samples

The Petersen method can be extended from two samples to s samples. With unique identifiers, after each sample is taken, we observe the identifier numbers of those tagged, tag the untagged, and return the sample to the population. We then build up a capture history of each tagged individual caught during the whole experiment. The aim is to estimate the size N of the population.

Otis et al. (1978) in a comprehensive booklet introduced a suite of eight closed population models that initially established a general notation for the models that we describe as M_0 , M_t , M_b , M_{tb} , M_h , M_{th} , M_{bh} , and M_{tbh} . Here the subscript “ t ” represents temporal changes in which the probability of capture is the same for all individuals but varies from sample to sample, “ b ” represents behavioral (e.g., trap) response, and “ h ” represents heterogeneity in which each individual has the same probability of capture in any sample, but that probability varies from individual to individual. Let p_{hi} be the probability that individual h is captured in sample i ($i = 1, \dots, s$). Then the general likelihood function (omitting constants) is

$$\prod_h \prod_i p_{hi}^{x_{hi}} (1 - x_{hi})^{1-p_{hi}}.$$

where $x_{hi} = 1$ if individual h is caught in sample i and $x_{hi} = 0$ otherwise. The eight models that follow we will refer to as Pollock’s models from his unpublished PhD thesis. Some of these have more than one way of being expressed.

Model M_0 $p_{hi} = p$ for all $h = 1, 2, \dots, N$ and all $i = 1, 2, \dots, s$.

Model M_t $p_{hi} = p_i$ for all h .

Model M_b $p_{hi} = p$ for first capture and $p_{hi} = c$ for recaptures.

Model M_h $p_{hi} = p_h$ for all i .

Model M_{tb} $p_{hi} = p_i$ for first capture and c_i for recaptures.

Model M_{bh} $p_{hi} = p_h$ for first capture and c_h for recaptures.

Model M_{th} p_{hi} unconstrained.

Model M_{tbh} p_{hi} unconstrained for first capture and c_{hi} unconstrained for recaptures.

A further list of these categories of models was given by Chao (2001) who gave a helpful table showing the log-linear/logistic forms of all the models. Using the previous notation, x_{hi} will give an $N \times s$ data matrix. For a multiplicable form M_{tbh} , we have

$$p_{hi} = \begin{cases} p_h e_i, & \text{until first capture} \\ \tau p_h e_i, & \text{for any recapture.} \end{cases}$$

Taking logarithms gives us

$$\log(p_{hi}) = \alpha_i + \beta_j + \alpha z_{hi},$$

where $\alpha = \log(\tau)$ and $z_{hi} = 1$ if animal h has been captured before the i th sample, and 0 otherwise. The other models of Chao (2001) are special cases of the above linear model.

King and McCrea (2019, chapter 2) considered the above models but also added a suite of models by including a category of group membership. They considered the case where additional information may be recorded on individuals observed within the study such as sex, age group, or breeding status. In such a model, we let $\mathcal{G} = \{1, \dots, G\}$ denote the set of G mutually exclusive and exhaustive individual physical characteristics (or groups) within a population. This leads to observed heterogeneity. They also considered the case where there is unobserved heterogeneity with k homogeneous (unobserved) subpopulations, but it is unknown which mixture or sub-population each individual belongs to. This type of model, discussed below, is referred to as a finite mixture model. They also considered a sequence of models using covariates and logistic linear models. McCrea and Morgan (2015: chapter 3) also considered some of the above models including mixture models, Bayesian models, and spatial capture-recapture models.

Kéry and Schaub (2012, chapter 6) discussed the above models, with a focus on data augmentation (parameter-expanded data augmentation) as well as with the addition of log-linear individual covariate models. The computer package **WinBUGS** was used along with Bayesian MCMC techniques, and extensive computer code was provided. Some practical data examples were given. They made the important comment that “what used to be called model t, b, or h is in fact not a single model, but in reality a whole family of models: there are various ways in which temporal, behavioral, or individual effects may be modeled.” We shall see this variation later in this chapter.

Methods of estimation for each of the above models as well as some variations will be given below, and there are further developments in the next chapter where Bayesian methods are also considered. But first we highlight a particular conditional model that has been used very successfully in various places in this book.

12.4.1 Multinomial Model

Let a_ω be the number of individuals with a particular capture history $\omega = (\omega_1, \omega_2, \dots, \omega_s)'$. For example, 10340...0 represents those individuals caught in the first, third, and fourth samples only. This can be expressed more briefly as simply 134 so that ω is then a nonempty subset of the integers $\{1, 2, \dots, s\}$. Thus our example becomes a_{134} for the number caught with that history; also r is the total number of different individuals caught. Let p_i be the probability of capture of an individual in the i th sample, and assume individuals act independently. Then P_ω , the probability that an individual chosen at random from the population with history ω , is the same for each individual, and the individuals may be regarded as N

independent “trials” from a multinomial experiment. Therefore the joint probability function of the random variables $\{a_\omega\}$ is

$$f(\{a_\omega\}) = \frac{N!}{\prod_\omega a_\omega!(N-r)!} Q^{N-r} \prod_\omega P_\omega^{a_\omega}, \quad (12.7)$$

where $Q = 1 - \sum_\omega P_\omega$ and r , a sufficient statistic for N , is the number of different individuals caught throughout the whole experiment. There are two ways of analyzing this model, either directly or using the following conditional model.

12.4.2 Conditional Method

As the conditional method is used particularly in this chapter and in other chapters as well, it is appropriate to give details of the method (cf Appendix A.3.2). If $P = \sum_\omega P_\omega$, Darroch (1958) showed that

$$\begin{aligned} f(\{a_\omega\}) &= f(\{a_\omega\}|r)f(r) \\ &= \frac{r!}{\prod_\omega a_\omega!} \prod_\omega \left(\frac{P_\omega}{P} \right)^{a_\omega} \times \binom{N}{r} P^r (1-P)^{N-r}. \end{aligned} \quad (12.8)$$

If P_ω is some function of an unknown vector parameter θ , say $\theta = (p_1, \dots, p_s)'$, then using the conditional model $f(\{a_\omega\} \mid r)$, it transpires that the maximum likelihood estimate of θ , $\hat{\theta}_C$, say, is virtually the same as the unconditional estimate from the full distribution $f(\{a_\omega\})$. Using the binomial distribution $f(r)$ of r with parameters N and P , Fienberg (1972) then obtained the conditional maximum likelihood estimate \hat{N}_C of N as

$$\hat{N}_C = [r/\hat{P}_C],$$

where $\hat{P}_C = P(\hat{\theta}_C)$ and $[r/\hat{P}_C]$ is \hat{P}_C rounded down to the nearest integer. The conditional and unconditional estimates have the same asymptotic distributions (Sanathanan, 1972a). The conditional distribution can be used for testing the hypothesis $P_\omega = P_\omega(\theta)$. It can also be used by applying standard large sample maximum likelihood theory to it to obtain the asymptotic variance-covariance matrix conditional on r and then make this matrix unconditional by using the fact that r is binomial. Frequently, it is helpful to use the parameter $Q = 1 - P$, the probability of not being caught.

Since the case $s = 2$ has already been discussed in detail, the conditional probability function is then given by

$$f(a_{10}, a_{01}, a_{11} | r) = \frac{r!}{a_{10}! a_{01}! a_{11}!} (p_1 q_2)^{a_{10}} (q_1 p_2)^{a_{01}} (p_1 p_2)^{a_{11}} (1 - Q)^{-r},$$

where $Q = q_1 q_2$.

The conditional model has been used by Huggins (1989, 1991) and Alho (1990) to develop some special cases of the above six models using linear logistic regression and possible covariates. By conditioning on just the captured individuals, characteristics of individuals not captured are not required, a plus when covariates are used. The use of covariates also helps reduce the number of unknown parameters. The conditional likelihood has now become a widely used approach in a variety of settings, as seen in this book. It was incorporated in the MARK program (White & Burnham, 1999) to analyze data from closed populations, and the function F.huggins.estim in a R package mra (McDonald, 2010) is also available. Huggins' models are discussed in Sect. 12.5.1.

12.5 Pollock's Model Categories

We shall begin by looking at each of the individual models involved, one at a time. Before doing so, we refer to Xi et al. (2008) who considered the minimum capture proportion for a reliable estimate (RE) of population size N in capture-recapture models. Their criterion is

$$\text{RE}(\hat{N}) = \frac{\sqrt{\text{MSE}(\hat{N})}}{N} \leq a,$$

for some specified value a , where MSE is the mean square error. For the models to be considered, we only have the asymptotic RE as $N \rightarrow \infty$, and estimates are assumed to be asymptotically unbiased, so that the criterion becomes

$$\text{a.var}(\hat{N}) \leq a^2 N^2,$$

where a.var denotes the asymptotic variance.

For a capture-recapture (CR) model with model k , M_k , say, it is generally found that $\text{a.var}(\hat{N}_k) = N \psi_k(\pi)$, where π is the average capture proportion of the population captured at least once during the study. We now have as the criterion for the minimum capture proportion (MCP)

$$\pi_k^* = \min\{\pi : \psi_k(\pi) \leq a^2 N\}.$$

Since $\psi_k(\pi)$ is generally a decreasing function of π , π_k^* can be obtained by solving the equation

$$\psi_k(\pi) = a^2 N.$$

Looking at models below, the authors found that although asymptotic values are required, simulation showed that the results are adequate for quite small N . For simple models, the results are adequate for $n = 20$, but for more complicated models, larger population sizes are required before the asymptotic results provide a good approximation. A reasonable value of a is $a = 0.1$. The authors gave values of MCP for a number of discrete and continuous models, which we now consider.

The first model is the oldest and most well-known and is usually referred to as the Schnabel (1938) census. Special cases of this were considered previously such as two samples (Lincoln-Petersen method) and three samples (triple-catch method).

12.5.1 Model M_t

In this model, we assume that $p_{hi} = p_i$ so that the probability depends only on time. We begin with some notation. Let

$p_i = (1 - q_i)$, the probability an individual is caught in sample i ,

n_i = size of the i th sample, ($i = 1, 2, \dots, s$),

m_i = number of tagged individuals in n_i ,

$u_i = n_i - m_i$,

$M_i = \sum_{j=1}^{i-1} u_j$, ($i = 1, 2, \dots, s + 1$),

= number marked in the population just before sample i is taken.

Since there were no tagged individuals in the first sample, $m_1 = 0$, $M_1 = 0$, $M_2 = u_1 = n_1$ and we define M_{s+1} (= r , say) as the total number of marked individuals in the population at the end of the experiment, i.e., the total number of different individuals caught throughout the experiment. We describe two major versions of this model, and the notation used follows that of Seber (1982, chapter 4). We recall the general model (12.7), namely,

$$f(\{a_\omega\}) = \frac{N!}{\prod_\omega a_\omega!(N-r)!} Q^{N-r} \prod_\omega P_\omega^{a_\omega},$$

where $Q = 1 - \sum_{\omega} P_{\omega}$. Assuming that for any individual the events “caught in the i th sample ($i = 1, 2, \dots, s$)” are independent of one another, our example becomes $\mathcal{Q} = \prod_{i=1}^s q_i$ and

$$P_{134} = p_1 q_2 p_3 p_4 q_5 \cdots q_s = \frac{p_1 p_3 p_4 Q}{q_1 q_3 q_4}, \quad \text{etc.}$$

Darroch (1958) showed that (12.7) reduces to

$$f(\{a_{\omega}\}) = \frac{N!}{\prod_{\omega} a_{\omega}!(N-r)!} \prod_{i=1}^s p_i^{n_i} q_i^{N-n_i}. \quad (12.9)$$

From the above assumptions, the $\{n_i\}$ are independent binomial variables, so that

$$f(\{n_i\}) = \prod_{i=1}^s \binom{N}{n_i} p_i^{n_i} q_i^{N-n_i}$$

and the joint probability function of the $\{a_{\omega}\}$ conditional on fixed sample sizes $\{n_i\}$ (i.e., the sample sizes are chosen in advance) is

$$f(\{a_{\omega}| \{n_i\}\}) = \frac{N!}{\prod_{\omega} a_{\omega}!(N-r)!} \prod_{i=1}^s \binom{N}{n_i}^{-1}. \quad (12.10)$$

If ∇ denotes a first difference, then by setting $\nabla \log f(\{a_{\omega}\}| \{n_i\}) = 0$ and using the fact that $\nabla \log N! = \log N$ etc., Darroch showed that the maximum likelihood estimate \widehat{N} of N for the model (12.10) is the unique root, greater than r , of the $(s-1)$ th degree polynomial given by

$$\left(1 - \frac{r}{N}\right) = \prod_{i=1}^s \left(1 - \frac{n_i}{N}\right). \quad (12.11)$$

He also found for model (12.10) the asymptotic bias and asymptotic variance $V[\widehat{N}]$ as $N \rightarrow \infty$ and $n_i \rightarrow \infty$ such that n_i/N remains constant, as well as confidence intervals for N . Here $V[\widehat{N}]$ is estimated by

$$v[\widehat{N}] = \left[\frac{1}{\widehat{N}} + \frac{s-1}{\widehat{N}} - \sum_{i=1}^s \left(\frac{1}{\widehat{N}-n_i} \right) \right]^{-1}.$$

When $s = 2$ the above Eq. (12.10) has the solution $\widehat{N} = n_1 n_2 / m_{12}$, the Petersen estimate. Equation (12.10) was first obtained by Chapman (1952) assuming random

samples without replacement. If we sum on the $\{a_\omega\}$ in (12.10) for each sample, we get the probability function of the $(\{m_i\})$, namely,

$$\begin{aligned} f(m_2, \dots, m_s | \{n_i\}) &= \prod_{i=2}^s \binom{M_i}{m_i} \binom{N - M_i}{n_i - m_i} / \binom{N}{n_i} \\ &= \frac{\prod_{i=2}^s \binom{M_i}{m_i}}{\prod_{i=1}^s u_i!} \cdot \frac{N!}{(N - r)!} \prod_{i=1}^s \binom{N}{n_i}^{-1}. \end{aligned} \quad (12.12)$$

This leads to the same Eq. (12.11).

From Sect. 12.5, the MCP (Minimum Capture Proportion) to achieve the criterion $\leq a$ (Xi et al., 2008, 246) is the solution π^* of $\psi_t(\pi) = a^2 N$, where

$$\psi_t(\pi) = \frac{\pi}{1 - \pi} - \sum_i^s \frac{p_i}{1 - p_i}.$$

Random Samples

In the previous discussion, we considered estimates for the model (12.9). Darroch (1958) investigated (12.8) in detail, which is the case when the sample sizes are random, often the more realistic situation. He found that as far as point and interval estimation of N is concerned, there is no difference for large samples between the two cases of fixed and random sample sizes, the reason being that the maximum likelihood estimate \hat{N} is almost the same in both cases, and in estimating the variance of \hat{N} , one effectively replaces $E[n_i]$ by random n_i . For later reference, Darroch found that the asymptotic variance of \hat{N} from (12.9) is

$$V(\hat{N}) = N \left[\frac{1}{q_1 q_2 \cdots q_s} + s - 1 - \sum_{i=1}^s \frac{1}{q_i} \right]^{-1}. \quad (12.13)$$

Seber (1982, chapter 4) had an extensive discussion of the above models and mentioned a number of regression methods along with examples. Also a number of tests for various departures from the underlying assumptions were given. He also discussed sequential and inverse sampling versions of the Schnabel method in Seber (1982, Sections 4.2, 4.3, and 12.8.5).

If overdispersion is present for the situation where we have several multinomial distributions, e.g., covering several years, Farzana et al. (2020) used a quasi-likelihood approach to give a new estimator. This estimator compared favorably with other standard estimators, especially with sparse data.

Known Removals

In capture-recapture experiments for animals, there are sometimes accidental losses through trapping, and some may be deliberately removed for further study. Suppose d_i individuals from the i th sample are removed, and let M_i be the number of tagged individuals alive in the population before the i th sample is taken. Here N is now the

initial population size. Assuming the $\{n_i\}$ to be fixed parameters, Chapman's model (12.10) becomes

$$f(m_2, \dots, m_s | \{n_i\}) = \prod_{i=2}^s \frac{\binom{M_i}{m_i} \binom{N-M_i-D_i}{n_i-m_i}}{\binom{N-D_i}{n_i}},$$

where $D_i = \sum_{j=1}^{i-1} d_j$, the total removal up to but not including the i th sample. We find that the maximum likelihood estimate \widehat{N}_D of N is now the unique root greater than r of the polynomial of

$$\frac{N-r}{N} = \prod_{i=1}^s \left\{ \frac{N-D_i-n_i}{N-D_i} \right\}, \quad i = 1, \dots, s$$

where r is the total number of different animals caught during the whole experiment, *including* the ones not returned. Further details are given by Seber (1982, section 4.1.4).

Regression Methods

Apart from logistic regression and log-linear models, ordinary regression models relating to animal abundance are not featured so much these days. However, they can be useful informally in providing some visual graphical information. For example, using the catchability theory, Seber (1970, 1982, 151) showed that for the two-sample case

$$E[(n_1 + 1)(n_2 - m_{12}/(m_{12} + 1)] \approx k(N - n_1).$$

Making use of this equation, he obtained the regression plot

$$Y_i = k(N - M_i) + \epsilon_i$$

proposed by Marten (1970), where $Y_i = u_i(M_i + 1)/(m_i + 1)$, M_i is the number of tagged individuals in the population just before the i th sample, and $E[\epsilon_i] = 0$. Methods for fitting this model are given by Seber (1982, 85). Note that the variances of the ϵ_i will vary, depending on $M_i n_i$, though in many cases, they will not be too different. A departure from $k = 1$ will indicate that variable catchability is involved, though we still have to determine whether it is due to an inherent variation in catchability or due to tagging affecting future catchability.

Incomplete Detectability

This topic is included here as the model M_t is the one that has been used most frequently to incorporate detectability. For example, we have the two or multiple observer models in aerial censusing and, in general, various encounter models such as camera trapping.

Adding Known Number of Tagged

In some situations, a convenient method with a population (e.g., homeless people) is to add a known number of tagged individuals (“plants”) to the target population in study. Norris and Pollock (2001) considered some earlier applications of this method where it was generally assumed that the plants and the others had the same catchability distributions with just two samples and using the Petersen method. The authors considered several samples, and p_h was the probability capture for individual h for the target population and p_h^* for the planted population. Different distributions F and F^* , respectively, were assumed for the two populations, and individuals were assumed to be independent. A joint likelihood was then set up for the two sets of data, and models M_h and M_{bh} discussed below were considered.

Looking at one application, an annual Homeless Outreach Population Estimate (HOPE) survey carried out in New York City pays planted individuals to dress themselves as members of the target homeless population, before joining them on the streets and observing whether or not they are enumerated (Hopper et al., 2008). Goudie and Gormley (2013) noted that enumerators can fail to classify some plants as homeless when the individuals concerned, although appropriately attired, were judged to look too young and healthy to be homeless. It is indeed a key assumption of the plant-capture approach that the appearance and behavior of the plants match those of the target population. They considered the following generalization of the M_t model with a targeted population of N plus R plants added to the target population giving an augmented population. Let n_i^+ be the fixed number of those that are seen in the i sample ($i = 1, \dots, s$) from the augmented population and n_i from the targeted population. We then have the hypergeometric distribution

$$f_i^{(R)}(n_i) = \binom{N}{n_i} \binom{R}{n_i^+ - n_i} / \binom{N + R}{n_i^+}.$$

They then applied Eq. (12.10) to this model and the corresponding one with $R = 0$ to get a combined model in which duplicate capture histories for an individual are avoided. Maximum likelihood estimation for N was then used.

12.5.2 Model M_0

To deal with this special case (Goudie et al., 2007), we simply set each $p_i = p$ in the M_t model described above. Otis et al. (1978) found from simulation results that estimates from this model are not robust with respect to any type of variability. For example, any heterogeneity in capture probabilities leads to a significant negative bias in the estimation of N . The authors gave estimates and variance estimates for p and N , and, given the model to be true, they concluded from simulation results that the bias of \hat{N} is negligible for values of $p \geq 0.10$ and $s \geq 5$. For smaller probabilities of capture, however, positive relative biases of 15–20% can occur. Darroch (1958, 355) showed that the asymptotic variance for \hat{N} followed from (12.13) simply by

setting each q_i equal to $1 - p$. He concluded from this that, asymptotically, no information is gained by using the knowledge that p_i is constant. One might as well use model M_t .

12.5.3 Model M_b

This model allows an animal to exhibit a behavioral response to being captured such as becoming “trap addicted” or “trap shy” after being first captured. It assumes that once caught, the probability of capture does not change in the future. The model has probability function

$$f(\{a_\omega\}) = \frac{N}{\prod_\omega a_\omega!(N-r)!} p^r (1-p)^{N-r-M.} c^{m.} (1-c)^{M.-m.},$$

where p is the probability of catching an unmarked individual and c is the probability of catching a marked individual in a sample. Here r is the number of different individuals caught throughout the experiment, M_i is the number of marked in the population prior to sample i , $M. = \sum_{i=1}^s M_i$, and $m. = \sum_{i=1}^s m_i$, the total number of individuals (not necessarily different) that are caught. The first thing to notice about this probability function is that the capture-recapture data are independent of the estimation of N and p . From the second part of the probability function, we get that $\hat{c} = m./M.$ is the maximum likelihood estimator of c . Therefore, if we allowed the probability of capture to vary with the number of times captured, it would not affect the first part of the likelihood, only the part involving c .

Otis et al. (1978, Appendix D) suggested using a search method to find the maximum likelihood (ML) estimator \hat{N}_b , and then the ML estimator of p follows from

$$\hat{p} = \frac{M_{s+1}}{s\hat{N}_b - M.}.$$

The simplest way of dealing with this model is to use the removal model and treat each individual as “removed” when tagged. This has the advantage of maintaining the same trapping effort, i.e., the same number of individuals per trap; otherwise, this would decrease as the population is deleted with the usual removal method. Hence, from Sect. 10.1.1, the asymptotic variance of \hat{N}_b is

$$V[\hat{N}_b] = \frac{N(1-q^s)q^s}{(1-q^s)^2 - (ps)^2 q^{s-1}}.$$

The above authors stated that provided the model assumptions were satisfied, simulation results of experiments found that \hat{N}_b was a good estimator provided

$p \geq 0.2$. The authors gave tables indicating that biases of 15–20% could occur when $p = 0.1$ or N is moderate, say 400. However, the frequency of failure, which occurs when $\sum_{i=1}^s (s+1-2i)n_i < 0$ (Seber, 1970), is substantial (20%) when $p \leq 0.1$.

The MCP (minimum capture proportion) to achieve criterion $\leq a$ from Sect. 12.1 is (Xi et al., 2008, 246) the solution π^* of $\psi_t(\pi) = a^2 N$ where

$$\psi_t(\pi) = \left[\frac{\pi}{1-\pi} - \frac{s^2 \{1 - (1-\pi)^{1/s}\}^2}{(1-\pi)^{1/s} \pi} \right]^{-1}.$$

12.5.4 Model M_{tb}

This model is the same as M_b except that the probability of capture of an unmarked individual is now p_i in sample i and the probability of capturing a marked individual in sample i is c_i . The probability function of the $\{a_\omega\}$ is now

$$f(\{a_\omega\}) = \frac{N!}{(\prod_\omega a_\omega!)(N-r)!} \prod_{i=1}^s p_i^{u_i} (1-p_i)^{N-M_{i+1}} c_i^{m_i} (1-c_i)^{M_i-m_i},$$

where M_i is the number marked just prior to the i th sample and m_i caught in sample. Once again we can ignore the second component involving the c_i and use a generalized version of the removal method. Unfortunately, we now have $s+1$ parameters for the first component but only s u_i 's, so we need to introduce one constraint, say two p_i 's are equal, which is unrealistic. Another type of restriction considered is setting $c_i = \theta p_i$ (Chao et al., 2000), but this is a strong assumption and not likely to be true. Behavioral models have been difficult to deal with, and we now digress and look at more recent additions.

First-Order Markov Chain

Yang and Chao (2005) assumed that the capture-recapture experiment was conducted on T occasions (samples), started at time $t = 0$, with all animals behaving independently. They built in both temporary (lasting one occasion) and permanent effects of behavior into their first-order Markov chain model, which we now consider.

We assume, without loss of generality, that the animals caught in the experiment are labeled from 1 to r and those not captured labeled from $r+1$ to N . The number of observed captures with histories ω are again described by a_ω . With a slight change in notation ($i \rightarrow t$ and $s \rightarrow T$), we assume the experiment is carried out over T sampling occasions at time points $t = 1, \dots, T$. Let $x_{ht} = 1$ with probability p_{ht} if animal h is captured at time t , and 0 otherwise, and let the observed capture histories (1s and 0s) be expanded as an $r \times T$ matrix.

Let $y_{ht} = 1$ if individual h is marked at time t , and 0 otherwise. This means that $y_{ht} = 1$ if animal h has been captured at least once in $(1, \dots, t)$. The state of an animal at time t is determined by the capture status (capture/noncapture) and marking status (marked/unmarked by time t). Because $x_{ht} = 1$ implies $y_{ht} = 1$, the state $(x_{ht}, y_{ht}) = (1, 0)$ is not possible so that the state $(1, 0)$ is ruled out, leaving three states $\{a = (0, 0), b = (0, 1), c = (1, 1)\}$. The bivariate Markov chain assumes that (x_{ht}, y_{ht}) is a discrete-time stochastic process with state space $S = \{a, b, c\}$ for any animal with an initial state a , that is, $\Pr(x_{h0}, y_{h0} = a) = 1$. If s_j is the state at time j , then

$$\Pr[(x_{ht}, y_{ht}) = s_t | (x_{h,t-1}, y_{h,t-1})] = \Pr[(x_{ht}, y_{ht}) = s_t | (x_{h,t-1}, y_{h,t-1} = s_{t-1})].$$

If an animal h is in state a at time $t - 1$, the probability of making a transition into state a at time t is defined as

$$P_{aa}(h, t) = \Pr[(x_{ht}, y_{ht}) = a | (x_{h,t-1}, y_{h,t-1}) = a].$$

The transition probabilities $P_{ab}(h, t)$, $P_{ba}(h, t)$, \dots , $P_{cc}(h, t)$ are similarly defined. Then $P_{ab}(h, t)$, $P_{ba}(h, t)$, and $P_{ca}(h, t)$ are all zero as the corresponding events are not possible. Also $P_{aa}(h, t) + P_{ac}(h, t) = 1$, $P_{bb}(h, t) + P_{bc}(h, t) = 1$, and $P_{cb}(h, t) + P_{cc}(h, t) = 1$.

Define $n_{aa}(t)$ to be the number of transitions from state a at time $t - 1$ to state a at time t in the *observed* capture history matrix, and similarly define $n_{ac}(t)$, $n_{bb}(t)$, $n_{bc}(t)$, $n_{cb}(t)$, and $n_{cc}(t)$. Let $n_{aa} = \sum_{t=1}^T n_{aa}(t)$, with the other n_{uv} similarly defined and $n_{ac} = r$. The likelihood function now turns out to be

$$\begin{aligned} L(N, P_{ac}, P_{bc}, P_{cc}) &= \frac{N!}{(N-r)! \prod_\omega a_\omega!} \times \left[(1 - P_{ac})^{T(N-r)+n_{aa}} (P_{ac})^{n_{ac}} \right. \\ &\quad \times \left. (1 - P_{bc})^{n_{bb}} (P_{bc})^{n_{bc}} \times (1 - P_{cc})^{n_{cb}} (P_{cc})^{n_{cc}} \right]. \end{aligned}$$

Let $Q = (P_{aa})^T = (1 - P_{ac})^T$ be the probability that an animal is not captured in the experiment, and let $\boldsymbol{\eta} = \{P_{ac}, P_{bc}, P_{cc}\}$ be the set of transition probabilities. Since the above likelihood is for a multinomial distribution, the conditional approach of Appendix A.3.2 can be used, and we can factorize

$$L(N, \boldsymbol{\eta}) = L_b(N, \boldsymbol{\eta}) L_c(\boldsymbol{\eta}),$$

where

$$L_b(\boldsymbol{\eta}) = \binom{N}{r} (1 - Q)^r Q^{N-r},$$

and

$$L_c(\boldsymbol{\eta}) = \frac{r!}{(1 - Q)^r \prod_{\omega} a_{\omega}!} \times [(1 - P_{ac})^{n_{aa}} (P_{ac})^{n_{ac}} \\ \times (1 - P_{bc})^{n_{bb}} (P_{bc})^{n_{bc}} \times (1 - P_{cc})^{n_{cb}} (P_{cc})^{n_{cc}}].$$

The above equations provide a general framework, and conditional maximum likelihood estimates are given in the authors' Appendix for N (\widehat{N} , say) and the P_{uv} . An asymptotic expression for the variance of \widehat{N} is provided. Several special cases of the model are considered, including extensions to all of Pollock's models involving a behavior effect b . Covariates were also considered briefly.

General Behavioral Model

Farcomeni (2011) extended the above behavioral models to include a variety of behavioral effects. Let $x_{hi} = 1$ if individual h ($h = 1, \dots, N$) is caught in sample i ($i = 1, \dots, s$) with probability p_{hi} , and 0 otherwise. Let $\mathbf{x}_h = (x_{h1}, \dots, x_{hs})'$. Allowing for general dependence of x_{hi} on the entire previous capture history, let $\Pr(x_{h1} = 1) = p_1$ and

$$\Pr(x_{hk} = 1 | x_{h,k-1}, \dots, x_{h1}) = p_k(x_{h,k-1}, \dots, x_{h1}), \quad k > 1.$$

These conditional probabilities are arranged in lexicographical order in a vector

$$\mathbf{p} = (p_1, p_2(0), p_2(1), p_3(00), p_3(01), \dots, p_s(1, \dots, 1))',$$

where the leftmost index runs fastest and corresponds to the most recent capture occasion. The number of free parameters above is $2^s - 1$, and the joint probabilities can be computed using a simple chain rule. We find that the number of unknown parameters is equal to the number of observations so that a "saturated" model can be fitted using a log-linear parameterization for the s -dimensional contingency table. Assuming independent capture histories, the likelihood for the above model is

$$L(N, \mathbf{p}) = \prod_{h=1}^N p_1^{x_{h1}} (1 - p_1)^{1-x_{h1}} \\ \times \prod_{k=2}^s p_k(x_{h,k-1}, \dots, x_{h1})^{x_{hk}} (1 - p_k(x_{h,k-1}, \dots, x_{h1}))^{1-x_{hk}},$$

and sufficient statistics are the counts of all rows of the matrix $\mathbf{X} = (x_{hi})$ with equal capture histories. We denote these counts by n_g , where the capture history g lives in the space of observable capture histories \mathcal{G} . Then, if n is the number of different

individuals captured, $L(N, \mathbf{p})$ can be expressed in the form (using a conditional likelihood)

$$\begin{aligned} L(N, \mathbf{p}) &= L_r(N, \mathbf{p}) \times L_c(\mathbf{p}) \\ &= \binom{N}{r} (1 - Q)^r Q^{N-r} \times \frac{r!}{\prod_{g \in \mathcal{G}} n_g!} \frac{\prod_{g \in \mathcal{G}} \Pr(\mathbf{x}_h = g)^{n_g}}{(1 - Q)^r}, \end{aligned}$$

where the probability that an animal is not captured during the experiment is

$$Q = \Pr(\mathbf{x}_h = \mathbf{0}) = (1 - p_1) \prod_{k=2}^s [1 - p_k(0, \dots, 0)].$$

Conditional estimators of N and Q are obtained by solving two equations,

$$\widehat{N} = \left[\frac{r}{1 - \widehat{Q}} \right],$$

where $[\cdot]$ denotes the integer part and

$$\widehat{Q} = \frac{\widehat{N} - \sum_{h=1}^r x_{h1}}{\widehat{N}} \prod_{k=2}^s \left\{ \frac{\widehat{N} - r + \sum_{h=1}^r \prod_{j=1}^k (1 - x_{hj})}{\widehat{N} - r + \sum_{h=1}^r \prod_{j=1}^{k-1} (1 - x_{hj})} \right\}.$$

Closed-form expressions were given for the estimates of the p_k . Farcomeni (2011) discussed a wide variety of constrained forms of the above likelihood, depending on how behavioral effects are modeled, and included the previous model of Yang and Chao (2005) as a special case, as well as higher-order Markov chain models.

Fegatelli and Tardella (2012) also looked closely at various behavioral models where capture probabilities can depend on a temporary or enduring effect from previous captures and noted some likelihood problems in using the conditional models such as unbounded estimates for the finite size of a population. They considered a generalized k -th order Markov chain model as well as Farcomeni's (2011) model with various linear constraints on \mathbf{p} . In addition, they referred to the parallel removal model with its own problems (Seber, 1970), which they extended to provide conditions to guarantee the finiteness and uniqueness of the conditional maximum likelihood estimator of N . They showed that a fully Bayesian analysis overcomes the likelihood failure phenomenon.

Farcomeni (2016) pointed out that the above approach by Farcomeni (2011) has two strong limitations: firstly, there is no way of including observed or unobserved heterogeneity; secondly, the set of possible constraints is large, but there are no suggestions as to how to explore the model class. He endeavored to get around

these limitations by firstly including observed and unobserved heterogeneity, as in the continuous case (Farcomeni & Scacciatelli, 2013) considered in Sect. 12.5.2.

Let $\mathbf{x}_h = (x_{h1}, \dots, x_{hs})'$ ($h = 1, \dots, N$) denote the usual binary capture history for individual h and $i = 1, \dots, s$ capture occasions, let r be the number of different individuals captured, and let $p(\mathbf{x})$ denote the probability of a capture history \mathbf{x} . A fully general parameterization is given by the chain rule, so that

$$p(\mathbf{x}) = \Pr(x_{h1})\Pr(x_{h2} | x_{h1}) \cdots \Pr(x_{hs} | x_{h,s-1}, \dots, x_{h1}).$$

Let $p_1 = \Pr(x_{x1} = 1)$ and, for $a_i = \{0, 1\}$, let

$$p_i(a_{i-1}, \dots, a_1) = \Pr(x_{hi} = 1 | x_{h,i-1} = a_{i-1}, \dots, x_{h1} = a_1).$$

These conditional capture probabilities are arranged in lexicographical order in a vector

$$\mathbf{p}' = (p_1, p_2(0), p_2(1), p_3(0, 0), \dots, p_s(1, \dots, 1))$$

of $2^s - 1$ elements. Various equality constraints can be expressed in the form $\mathbf{C}\mathbf{p} = \mathbf{0}$ (usually contrasts) to give a range of possible M_{tb} models. More generally, Farcomeni (2016) introduced the model M_{hotb} , which is the model M_{tbb} discussed below with observed heterogeneity and covariates and unobserved heterogeneity (M_h models). A one-to-one reparameterization was used, namely,

$$\text{logit}(p_i(a_{i-1}, \dots, a_1)) = \beta_{ia_1, \dots, a_{i-1}},$$

where $\mathbf{C}\mathbf{p} = \mathbf{0}$ if and only if $\mathbf{C}\boldsymbol{\beta} = \mathbf{0}$. Let \mathbf{z}_{hi} be a subject and time-specific vector of covariates for individual h . One way of dealing with observed heterogeneity is to rely on a logistic reparameterization and use the conditional likelihood, conditional on those captured at least once. Unobserved heterogeneity for individual h can be allowed for using θ_h as a random effect with some distribution. For $i = 1, \dots, s$, the general model is now

$$\text{logit}(p_i(a_{i-1}, \dots, a_1)) = \beta_{ia_1, \dots, a_{i-1}} + \theta + \boldsymbol{\gamma}' \mathbf{z}.$$

Then, given the constraints $\mathbf{C}\boldsymbol{\beta} = \mathbf{0}$, the author gave a conditional log-likelihood function and obtained maximum likelihood estimates using a constrained EM algorithm. A Horwitz-Thompson estimate of N and its variance formula were given, followed by model fitting and model selection. The method was applied to two sets of data, HIV data and snowshoe hares data. A special case of the above model was given by Farcomeni (2015).

12.5.5 Model M_h

Burnham, in an unpublished PhD thesis (Oregon State University, 1972), was the first to consider heterogeneous models. This model assumes $p_{hi} = p_h$ ($h = 1, 2, \dots, N$) and that each animal has its own capture probability, which remains constant over the capture occasions and not being altered by a previous capture. This model cannot be used as it stands as there are $N + 1$ unknowns, namely, N and the p_h , so that further structure or assumptions need to be added for it to be used. Because of these difficulties, such an experiment should be devised, where possible, to minimize heterogeneity as much as possible. Although focused on open populations, Jeyam et al. (2017) provided a test for positive association for detecting heterogeneity and compared it with tests due to Leslie and Carothers.

The original approach has been to assume that the p_h are a sample from some density function $g(p)$. From Otis et al. (1978, Appendix E), we define x_{hi} , the random variable taking the value 1 when animal h is caught in sample i and 0 otherwise. It transpires that a sufficient statistic for the data are the capture frequencies f_1, f_2, \dots, f_s , where f_i is the number of animals caught exactly i times during the experiment. If r is the number of different individuals caught, the probability distribution used is then the multinomial distribution

$$\Pr(f_1, \dots, f_s) = \binom{N}{N-r, f_1, \dots, f_s} \pi_0^{N-r} \prod_{i=1}^s \pi_i^{f_i},$$

where $r = \sum_{i=1}^s f_i$ and

$$\pi_i = \int_0^1 \binom{s}{i} p^i q^{s-i} g(p) dp, \quad i = 0, 1, \dots, s.$$

If one assumes a parametric form for g with two parameters, then maximum likelihood theory can be used to obtain an estimator for N . Unfortunately, there is a problem with this approach as there is a type of non-identifiability with N . Two very similar distributions can lead to very different estimates of N , which means that inference for N is highly model dependent. For example, for two distributions, we can have the π_i the same for $i = 1, \dots, s$ but the π_0 are different, which means different \hat{N} . In fact Link (2003) said: "Thus even with very large samples, the analyst will not be able to distinguish among reasonable models of heterogeneity, even though these yield quite distinct inferences about population size." Restrictions are required on the class of possible distributions.

Fortunately, we can obtain an estimator of the form $\hat{N}_H = \sum_{i=1}^s a_i f_i$, where the constants a_i are generated by considering various jackknife estimators. These are developed by application of the generalized jackknife statistic (Gray & Schucany,

1972) to the naive estimator r and assuming the bias of r as an estimator of N is expressible as a power series in $1/s$. For example, the first order one is

$$\widehat{N}_{H1} = r + \frac{s-1}{s} f_1 = \left(1 + \frac{s-1}{s}\right) f_1 + \sum_{i=2}^s f_i.$$

For a k th order jackknife we get

$$\widehat{N}_{Hk} = \sum_{i=1}^s a_{ik} f_i,$$

which, from the multinomial distribution of the f_i given above, has an approximate variance estimator of

$$\widehat{\text{var}}[\widehat{N}_{Hk}] = \sum_{i=1}^s (a_{ik})^2 f_i - \widehat{N}_{Hk}.$$

Jackknife expressions for $k = 1, \dots, 5$ are given by Otis et al. (1978, Appendix E).

Schofield and Barker (2014) mentioned two challenges in dealing with models like M_h . Firstly, N appears as the population size, but it also determines the number of possible capture probabilities p_h . This means that the dimension of the vector of these probabilities is not fixed but can vary as estimation progresses. Secondly, the method of analysis is important as N may not be included as a basic parameter, but rather as a derived parameter. A Bayesian approach to the problem is given in the next chapter.

The M_h model has been found useful for estimating species richness as heterogeneity will exist among the species (e.g., Boulinier et al., 1998). For example, if ψ_i is the probability of occurrence of species i and p_i is the probability of detection of the species, then the most basic model (Kéry & Royle, 2008a) is given by $\text{logit}(\psi_i) = \alpha_i$, $\text{logit}(p_i) = \beta_i$, where $\alpha_i \sim N(\mu_\alpha, \sigma_\alpha^2)$ and $\beta_i \sim N(\mu_\beta, \sigma_\beta^2)$, with the assumption that the random effects are mutually independent. The authors used the method of data augmentation (Appendix A.5) to carry out their analysis.

Stoklosa et al. (2011) mentioned that the conditional model described in Appendix A.3.2 was complicated to implement with covariates and introduced a partial likelihood method based on conditioning on the first captured time of each individual. By comparing efficiencies with the conditional model, it was found that the partial method did lose some efficiency, but there was more flexibility in that standard software could be used. Their main focus was on the M_h model.

From Sect. 12.5, the MCP (minimum capture proportion) to achieve criterion $\leq a$ is the solution π^* of $\psi_t(\pi) = a^2 N$, where $\psi_t(\pi)$ is given by Xi et al. (2008, 247) for the case when $g(p)$ is the Beta distribution.

12.5.6 Model M_{th}

To use this model, some structure needs to be imposed such as $p_{hi} = p_h e_i$, as suggested by Otis et al. (1978), where i could be replaced by t the time of the i th sample. They also used the conditional model and required that the density function $f(p)$ of the p_h be completely specified because of the number of unknown parameters. Such an approach is an example of a mixed model where fixed and random effects are combined. For example, Huggins and Hwang (2011) considered the model

$$\log \left(\frac{p_{hi}}{1 - p_{hi}} \right) = \alpha_h + \beta_i,$$

where β_i ($i = 1, \dots, s$) are fixed time effects and the α_h ($h = 1, \dots, N$) are random heterogeneity effects, being regarded as independent and identically distributed random variables. Here the distribution of α_h is F_{gamma} for some parameters γ , for example, the gamma distribution (Sanathanan, 1972b) or the normal distribution (Coull & Agresti, 1999).

King et al. (2016) considered the following model with n distinct individuals observed, s samples, and population size N , where individuals are numbered $h = 1, \dots, n$ for the observed and $h = n + 1, \dots, N$ for the unobserved. Let $\mathbf{x}_i = \{x_{hi} : i = 1, \dots, s\}$ denote the capture history for individual h , where $x_{hi} = 1$ if individual h is observed/captured in sample i , and 0 otherwise, and $\mathbf{X} = \{\mathbf{x}_i\}$. If p_{hi} is the capture probability of individual h in sample i , then the authors used the model $p_{hi} = g(\boldsymbol{\theta}, \boldsymbol{\varepsilon}_h)$ for some function g , where $\boldsymbol{\theta}$ denotes the model parameters associated with the capture probabilities and $\boldsymbol{\varepsilon}_h$ corresponds to the individual heterogeneity for individual h . Let $\boldsymbol{\varepsilon} = \{\boldsymbol{\varepsilon}_h : h = 1, \dots, N\}$, where $\boldsymbol{\varepsilon}$ is assumed to be a function of the parameters $\boldsymbol{\eta}$ and that the individual heterogeneity terms and $\boldsymbol{\varepsilon}_h$ are independent of each other conditional on $\boldsymbol{\eta}$. The conditional density function for $\boldsymbol{\varepsilon}_h$ is defined to be $f(\boldsymbol{\varepsilon}_h | \boldsymbol{\eta})$, with joint density function

$$f_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon} | N, \boldsymbol{\eta}) = \prod_{h=1}^N f(\boldsymbol{\varepsilon}_h | \boldsymbol{\eta}).$$

Let $\boldsymbol{\varepsilon}_h = (\boldsymbol{\varepsilon}_h^{(O)}, \boldsymbol{\varepsilon}_h^{(U)})$, where O is for observed and U for unobserved. We similarly write $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}^{(O)}, \boldsymbol{\varepsilon}^{(U)})$. It is assumed that the capture histories of the individuals are independent of each other, given the capture probability model parameters, $\boldsymbol{\theta}$,

and individual heterogeneity terms $\boldsymbol{\varepsilon}$. We find that the marginal data likelihood can be expressed in the form

$$\begin{aligned} f_m(\mathbf{X}, \boldsymbol{\varepsilon}^{(O)} | N, \boldsymbol{\theta}, \boldsymbol{\eta}) &= \int_{\boldsymbol{\varepsilon}_1^{(U)}} \dots \int_{\boldsymbol{\varepsilon}_N^{(U)}} f_C(\mathbf{X}, \boldsymbol{\varepsilon} | N, \boldsymbol{\theta}, \boldsymbol{\eta}) d\boldsymbol{\varepsilon}_N^{(U)} \dots d\boldsymbol{\varepsilon}_1^{(U)} \\ &\propto \frac{N!}{(N-n)!} \prod_{h=1}^N \int_{\boldsymbol{\varepsilon}_h^{(U)}} f(\mathbf{x}_h | \boldsymbol{\theta}, \boldsymbol{\varepsilon}_h) f(\boldsymbol{\varepsilon}_h | \boldsymbol{\eta}) d\boldsymbol{\varepsilon}_h^{(U)}, \end{aligned}$$

where $f_C(\mathbf{X}, \boldsymbol{\varepsilon} | N, \boldsymbol{\theta}, \boldsymbol{\eta})$ corresponds to the complete data likelihood (i.e., the joint probability density function of the capture histories and individual effects). Three examples were considered by the authors, namely, including continuous individual covariates, M_h -type models with unobserved individual random effects, and spatial capture-recapture models. Computational methods and model fitting, including some difficulties, were described.

Stoklosa and Huggins (2012) used a conditional likelihood approach with time dependence and heterogeneity. If p_{hi} is the probability that individual h is captured in the i th sample, a generalized additive model was used for p_{hi} in the form $p_{hi} = H(\eta_{hi})$. Here $H(\cdot, \cdot)$ is a link function and

$$\eta_{hi} = \varphi + \sum_{j=1}^J f_j(x_{hj}) + \boldsymbol{\beta}' \mathbf{z}_h + \alpha_i.$$

This is expressed in terms of continuous covariates x_{hj} using B-splines f_j , categorical covariates \mathbf{z} (e.g., sex) for individuals, and random time effects α_i . A penalized likelihood was used, and computations were given in an Appendix.

Other methods for the M_{th} model are described below, such as the coverage model approach of Sect. 12.8.

12.5.7 Model M_{bh}

Although possibly one of the most useful models, it is also one of the most difficult ones to use. We begin by assuming that (p_h, c_h) , the probabilities of individual h being caught for the first time or being caught at least once, respectively, are a random sample from some joint distribution function $G(p, c; \boldsymbol{\theta})$. Otis et al. (1978) obtained the multinomial distribution

$$\Pr(\{u_i\} | g(p, c; \boldsymbol{\theta}) = \frac{N!}{(\prod_{i=1}^s u_i!) (N-r)!} (\prod_{i=1}^s \pi_i^{u_i}) (\pi_{s+1})^{N-r}, \quad (12.14)$$

where π_i is a function of the moments of $G(p, c; \theta)$, and $\pi_{s+1} = 1 - \sum_{i=1}^s \pi_i$. However, the authors considered basing the estimation of N solely on the “removal” approach where once an individual is caught, it is regarded as being removed from the population. They began by transforming the parameters $\{\pi_i\}$ into the set $\{\bar{p}_i\}$ by using the relationships

$$\pi_i = (1 - \bar{p}_1)(1 - \bar{p}_2) \cdots (1 - \bar{p}_{i-1})\bar{p}_i, \quad i = 1, 2, \dots, s.$$

Here \bar{p}_i is a conditional probability that represents the average first capture probability of those members of the population that have not yet been captured at the time of the i th sample. The authors argue that it is not unreasonable to assume $\bar{p}_1 > \bar{p}_2 > \dots > \bar{p}_s$ and $(\bar{p}_1 - \bar{p}_2) > (\bar{p}_2 - \bar{p}_3) > \dots > (\bar{p}_{s-1} - \bar{p}_s)$, so that the larger differences in average first capture probability occur initially. For example, both conditions hold for a Beta distribution.

We use (12.14) as a basis for the generalized removal method now described. For $k = 1, 2, \dots, s-2$, Otis et al. (1978) proposed the following:

- (i) Assume that $\bar{p}_k = \bar{p}_{k+1} = \dots = \bar{p}_s = p$ and that the $\bar{p}_1, \dots, \bar{p}_{k-1}$ differ. This reduces the number of parameters involved in the estimation of N to $k+1$. To use maximum likelihood estimation, we abbreviate (12.14) by leaving off $g(p, c; \theta)$ for convenience and write

$$f(\{u_i\}) = \prod_{i=1}^s f(u_i | u_1, u_2, \dots, u_{i-1}),$$

where $f(u_i | u_1, u_2, \dots, u_{i-1})$ is the conditional distribution of the i th “removal,” given the values of the previous removals.

- (2) Estimate N by the maximum likelihood method.
- (3) Choose the smallest value of k that produces a sufficient fit to the data, and take as the estimate of N the estimate associated with this value of k . The authors measured the fit of the data $\{u_i\}$ using the usual chi-square goodness-of-fit test but with a significance level of 0.2 because of the seriousness of Type II errors. If significance levels of all the tests (one for each value of k) are less than 0.20, the authors chose the value of k corresponding to the largest achieved significance level.

For a given value of k , the asymptotic sampling variance $\text{var}[\hat{N}_{\text{bh}}]$ of \hat{N}_{bh} is N times

$$\left\{ \frac{\sum_{i=1}^s \pi_i}{1 - \sum_{i=1}^s \pi_i} - \sum_{i=1}^{k-1} \frac{\bar{p}_i^2}{\pi_i(1 - \bar{p}_i)} - \frac{(s - k + 1)^2 \bar{p}^2}{(1 - \bar{p}) \left[\prod_{i=1}^{k-1} (1 - \bar{p}_i) + \sum_{i=1}^s \pi_i - 1 \right]} \right\}^{-1}$$

The failure criterion for any value of k is $\sum_{j=k}^s (s + k - 2j)u_j \leq 0$. If this happens, a different value of k needs to be considered.

12.5.8 Model \mathbf{M}_{tbh}

Because of the complications involved, this model was not considered initially to be useful in practice, as strong assumptions were needed to reduce the number of unknown and unidentifiable parameters. However, other methods are described in this chapter, which use covariates to reduce the number of parameters. This leads us to consider in more detail some models proposed by Huggins and colleagues in Sect. 12.5.11.

12.5.9 Choosing the Number of Samples

Fegatelli and Farcomeni (2016) considered the problem of determining the smallest number of capture samples s so that the expected length of the profile confidence interval is less than or equal to a fixed threshold. We recall that the general likelihood function for the above models is

$$\prod_h \prod_i p_{hi}^{x_{hi}} (1 - p_{hi})^{1-x_{hi}},$$

where $x_{hi} = 1$ if individual h is caught in sample i and $x_{hi} = 0$ otherwise. We can express this briefly as $L(N, \mathbf{P}; \mathbf{T})$, where \mathbf{P} is the matrix of the p_{hi} and \mathbf{T} is a multidimensional sufficient statistic. Using a profile likelihood approach, let $\ell(N; T) = \sup_{\mathbf{P}} \log(L(N, \mathbf{P}; T))$, and let $N_\ell(T)$ and $N_u(T)$ denote the limits of the confidence interval, as the solutions to the equation in N of

$$2(\ell(\tilde{N}; T) - \ell(N; T)) = z_{\alpha/2}^2,$$

where $z_{\alpha/2}$ is the upper $\alpha/2$ quantile of a standard normal distribution and $(1 - \alpha)$ is the desired coverage of the confidence interval. The width of the confidence interval is approximately $N_u(T) - N_\ell(T)$, which is needed to determine the required sample size. We have just sketched the method, and extensive details are given by the authors for all of the above eight models including “failure” conditions for some models.

12.5.10 Model Selection

In addition to the usual methods of CR model selection, Bartolucci and Lupparelli (2008) introduced another method called the focused information criteria (FIC), which used the conditional likelihood approach of Appendix A.3.2. It leads to the smallest mean-squared error (MSE) of the resulting estimator of the population

size and is based on an index which, up to a constant term, is equal to the asymptotic MSE of the estimator, extending the results of Hjort and Claeskens (2003).

12.5.11 Huggins' Models

We recall that p_{hi} is the probability that individual h ($h = 1, 2, \dots, N$) is captured in sample i ($i = 1, \dots, s$). Huggins (1989) introduced a procedure for estimating population size by modeling the p_{hi} in terms of observable covariates such as age, sex, etc., by constructing a likelihood conditional on the captured individuals for parameter estimation (the “conditional model”), which avoids information for those not captured. Huggins (1991) and Hwang and Huggins (2011) considered the following models, using logistic regression, which are similar to the Pollock categories above.

$$M_0: \text{logit}(p_{hi}) = \beta_0.$$

$M_t:$ $\text{logit}(p_{hi}) = \beta_0 + \beta_i$ with $\beta_s = 0$. This logistic model allows the capture probabilities to vary only by time and is a re-parameterization of model M_t of Otis et al. (1978).

$M_b:$ $\text{logit}(p_{hi}) = \beta_0 + \beta_b z_{hi}$. For example, we can define $z_{hi} = 1$ if individual h has been captured before sample/time i and $z_{hi} = 0$, otherwise. This parameterization allows the capture probabilities to vary according to an individual's capture history. The capture probability changes once an individual is caught. Hwang and Huggins (2007) and Huggins and Hwang (2007) gave semiparametric and nonparametric methods, respectively, for the model.

$M_{tb}:$ $\text{logit}(p_{hi}) = \beta_0 + \beta_i + \beta_b z_{hi}$ with $\beta_s = 0$. This model allows the capture probabilities to vary according to time i and past behavior b .

$M_h:$ Huggins (2001) noted the shortcomings of some models proposed in the literature, including some identifiability issues relating to random effects. A covariate model used was

$$\text{logit}(p_{hi}) = \beta_0 + \beta_{\text{sex}} + \beta_{\text{age}} + \beta_{\text{wt}} \times \text{weight}(h),$$

where the covariate β_{sex} is the effect of sex, β_{age} the effect of age, β_{wt} the effect of one unit of weight, and weight (h) the weight of individual h . This model accounts for heterogeneity resulting only from differences in sex, age, and weight. It was also considered by Huggins (1989), and Xi et al. (2009) assumed that the covariates are subject to measurement error and may be missing completely at random for each capture. White and Cooch (2017) considered the above special case of M_h for an encounter model for individuals seen at least once that might be used for camera trapping, for example.

$M_{th}:$ $\text{logit}(p_{hi}) = \alpha x_h + \beta_i$, where x_h is the body weight of individual h caught in one of the samples (Huggins & Hwang, 2011).

M_{bh} : $\text{logit}(p_{hi}) = \beta_0 + \beta_{\text{sex}} + \beta_{\text{age}} + \beta_{\text{wt}} \times \text{weight}(h) + \beta_b z_{hi}$.

M_{tjh} : $\text{logit}(p_{hi}) = \beta_0 + \beta_{\text{sex}} + \beta_{\text{age}} + \beta_{\text{wt}} \times \text{weight}(h) + \beta_i + \beta_b z_{hi}$. An alternative parametric model was given by Huggins and Hwang (2011).

A computational package called VGAM (vector generalized additive model) or handling models like the above has been provided by Yee et al. (2015).

12.5.12 Test of Model Assumptions

Otis et al. (1978) gave a number of tests for testing individual model assumptions. To begin with, we need to test whether the population is actually closed, and such an omnibus test is not available as existing tests may indicate non-closure even when a population is closed if other effects such as trap response are present. Some specific tests exist such as the closure.test secr (cf. <https://www.otago.ac.nz/density/html/closure.html>). Stanley and Burnham (1999) presented a closure test for time-specific data that, in principle, tests the null hypothesis of the closed-population model M_t against the open-population Jolly-Seber model as a specific alternative. This test is chi-square and can be decomposed into informative components that can be interpreted to determine the nature of closure violations. The test is most sensitive to permanent emigration and least sensitive to temporary emigration and is of intermediate sensitivity to permanent or temporary immigration. Stanley and Richards (2005) introduced CloseTest, a Windows-based program that computes the Otis et al. (1978) and Stanley and Burnham (1999) closure tests for capture-recapture data sets. Information on CloseTest features and where to obtain the program are provided.

In model fitting, there are two types of test. The first is a goodness-of-fit test to see whether a particular model fits the data. The second is a nested hypothesis test to see if a simpler model does better than a more complex model. In the theory above, we could start with the most complex model such as M_{tjh} and then sequentially test to see if a simpler model will do just as well or better. We could also start at the other end with the simplest model M_0 and expand it while there is an adequate fit using, for example, the so-called Score test.

12.6 Mixture Models

Pledger (2000) and colleagues originally developed a number of closed population models using the method of mixtures, which especially deal with heterogeneity. The idea is that there is an unknown number of underlying groups in the population, each with its own catchability distribution. She began with the standard model of assuming that the capture of animal h in sample i is a Bernoulli trial with probability p_{hi} of capture. If r distinct animals are caught overall, the data form the $r \times s$ matrix

of data with element $x_{hi} = 1$ if animal h is caught in sample i , and 0 otherwise. Row h of the matrix is then the capture history of animal h , namely, ω'_h . We then add $N - r$ rows of zeros to include the $N - r$ animals not captured to give the matrix \mathbf{X} . We now have the joint probability function

$$\Pr(\mathbf{X}; N, \{p_{hi}\}) = \frac{N!}{(\prod_{\omega} n_{\omega}!)(N - r)!} \prod_{h=1}^N \prod_{i=1}^s \{p_{hi}^{x_{hi}} (1 - p_{hi})^{1-x_{hi}}\},$$

where n_{ω} is the number of observed animals with capture history ω .

Pledger (2000) considered a broad mixture model called $M_{t \times b \times h_G}$ to allow for variable group catchability. The model assumes that there is a fixed number, G , of animal groups within each of which there is homogeneity of capture. It is not known which animals are in which group, and each animal is assumed to come independently from group g with probability π_g ($g = 1, \dots, G$, $\sum_g \pi_g = 1$). The probability of capturing animal h in sample i is assumed to take the form θ_{ibg} with g replacing h and modeled with a linear logistic formulation as an analysis of variance model with main effects and interactions given by

$$\text{logit}(\theta_{ibg}) = \mu + \tau_i + \beta_b + \eta_g + (\tau\beta)_{ib} + (\tau\eta)_{ig} + (\beta\eta)_{bg} + (\tau\beta\eta)_{ibg}.$$

Here μ is a constant unknown parameter, τ_i (for time) is a fixed main effect depending on the sample i , β_b is a fixed behavior main effect (where $b = b_{hi} = 1$ if animal h was not caught before sample i ; otherwise, $b = 2$), and η_g (for heterogeneity) is a random animal main effect. Parameters $(\tau\beta)_{12}$ or $(\tau\beta\eta)_{12g}$ are not needed, as no recaptures can occur in sample 1.

To make the parameters identifiable, as with the usual analysis of variance models, constraints are needed such as setting to zero any parameters with $i = 1$, $b = 1$, and $g = 1$. Also, with no parameters for recapture at sample 1, we set $(\tau\beta)_{22} = (\tau\beta\eta)_{22g} = 0$ for all $g = 1, \dots, G$. The numbers of independent parameters associated with the terms of the model are then: μ (1); τ_i ($s-1$); β_b (1); η_g ($G - 1$); $(\tau\beta)_{ib}$ ($s - 2$); $(\tau\eta)_{ig}$ ($G - 1$) ($s - 1$); $(\beta\eta)_{bg}$ ($G - 1$); and $(\tau\beta\eta)_{ibg}$ ($G - 1$) ($s - 2$). These $2Gs - G$ independent parameters together with N and the $G - 1$ independent π parameters for group membership give $2Gs$ independent parameters for model $M_{t \times b \times h_G}$.

The corresponding likelihood function (ignoring the term $(\prod_{\omega} n_{\omega}!)$ in what follows) is

$$L(N, \{\pi\}, \{\theta\} | \mathbf{X}) \propto \frac{N!}{(N - r)!} \prod_{h=1}^N \sum_{g=1}^G \left[\pi_g \prod_{i=1}^s \left\{ (\theta_{ibg})^{x_{hi}} (1 - \theta_{ibg})^{1-x_{hi}} \right\} \right].$$

One submodel is the main effects model M_{t+b+g} in which only the constant and main effects terms are retained. Various hierarchical submodels of the above logistic

model are given by Pledger (2000, her Table 1). If an interaction term is present, so are all its associated main effects and lower-order interactions.

The models without heterogeneity, namely, M_0 , M_t , M_b , and M_{tb} , give the same maximum likelihood estimates as those of Otis et al. (1978). We now consider some submodels with heterogeneity beginning with the simplest M_{hg} with the capture probability θ_g . This leads to the model

$$\text{logit}(\theta_g) = \mu + \eta_g$$

with likelihood

$$L(N, \{\pi_g\}, \{\theta_g\}; \mathbf{X}) = \frac{N!}{(N-r)!} \prod_{j=0}^r \left[\left\{ \sum_{g=1}^G \pi_g \theta_g^j (1-\theta_g)^{s-j} \right\}^{f_j} \right],$$

where f_j is the number of animals with j captures so that $f_0 = N-r$. There are $2G$ independent parameters and s minimal sufficient statistics f_1, \dots, f_s , so we need $2G \leq s$ for estimation. A sequence of models with one, two, three, or more animal groups M_{h_1} ($= M_0$), M_{h_2} , M_{h_3} , etc. can be fitted and compared using likelihood ratio tests. Pledger noted that with a fixed number of animal groups, multimodality in the likelihood function might be a problem. Brooks et al. (1997) recommended obtaining profile likelihoods plotted against the mixing probability (π_1) in mixture problems with two components M_{h_2} . In doing this Pledger found that multimodality of M_{h_2} was fairly rare.

Another submodel, a variation on M_{th} , is $M_{t \times hg}$, along with a main effects version M_{t+hg} and a proportional version if the main effects are assumed additive on a log scale. Agresti (1994) used a main effects model with finite mixtures for a Poisson-based model (described below) of M_{th} .

With the interactive model $M_{b \times hg}$, we have the likelihood

$$\begin{aligned} L(N, \{\pi_g\}, \boldsymbol{\theta}; \mathbf{X}) &= \frac{N!}{(N-r)!} \left\{ \prod_{h=1}^r \sum_{g=1}^G \pi_g \theta_{1g} (1-\theta_{1g})^{a_h-1} \theta_{2g}^{r_h} (1-\theta_{2g})^{s-a_h-r_h} \right\} \\ &\quad \times \left\{ \sum_{g=1}^G \pi_g (1-\theta_{1g})^s \right\}^{N-r}, \end{aligned}$$

where a_h is the sample number at which animal h is first captured and r_h is its number of recaptures.

Pledger (2000) noted that a major problem with the previous models is the large number of parameters assigned to the time factor. Also we saw above that there is a major problem in fitting M_{tb} because of unidentifiable parameters. She said that both

problems are solved if time (as well as heterogeneity) is modeled as a random effect. This leads to the general model $M_{tv \times b \times h_G}$ in which samples are partitioned into V groups ($v = 1, \dots, V$) and the animals are partitioned into G groups giving a two-dimensional cross-partition of the capture matrix X . Each animal has probability π_g of belonging to group g , and each sample has probability λ_v of being in sample group v ($\sum \pi_g = 1$ and $\sum \lambda_v = 1$). For animal group g and sample group v , the probability of capture is modeled as

$$\text{logit}(\theta_{vbg}) = \mu + \tau_v + \beta_b + \eta_g + (\tau\beta)_{vb} + (\tau\eta)_{vg} + (\beta\eta)_{bg} + (\tau\beta\eta)_{vbg},$$

This leads to the likelihood

$$\begin{aligned} L(N, \{\pi_g\}, \{\lambda_v\}, \boldsymbol{\theta}; \mathbf{X}) \\ = \frac{N!}{(N-r)!} \sum_{v_s=1}^V \lambda_{v_s} \dots \sum_{v_2=1}^V \lambda_{v_2} \sum_{v_1=1}^V \lambda_{v_1} \\ \times \prod_{h=1}^N \sum_{g=1}^G \left[\pi_g \prod_{i=1}^s \left\{ (\theta_{gbv_i})^{x_{hi}} (1 - \theta_{gbv_i})^{1-x_{hi}} \right\} \right], \end{aligned}$$

where v_i is the allocation of sample i . The $2GV$ capture probability parameters, together with $G - 1$ independent π parameters, $V - 1$ independent λ parameters, and N , give a total of $2GV + G + V - 1$ model parameters. If, as tends to happen, only two animal groups and two sample groups are needed, this means 11 parameters. By contrast, the model with fixed time effects and two animal groups has 4s parameters. Of particular value is the model $M_{tv \times b}$ as M_{tb} was not feasible. Various hypothesis test procedures were considered along with some difficulties. For example, one of the problems in comparing two nested models is when extra animal groups or sample groups are introduced as we move from the sub-model to the supermodel.

Pledger (2005) carried out a simulation study of various mixture distributions including the use of three mixture models considered by Dorazio and Royle (2003), with the beta-binomial described in detail below. She considered 18 generating distributions using simulation and some real data sets and used as a measure of heterogeneity the ratio $\eta = \sigma^2 / (\mu(1 - \mu))$ suggested by Dorazio and Royle (2003), as well as the skewness of the distribution. Her emphasis was on fitting the beta and two-point mixture models to investigate the importance of true distribution shape. She commented that all models are wrong (which we know in general), but some are better than others. It is best to consider distributions with a variety of shapes as well as discrete and continuous forms, and she found that a discrete model with a two-point support can often provide sufficient variability. She concluded that despite the problems raised in her article, “heterogeneity models are working well for many data sets, and provide distinct improvement over the homogeneous models” in spite of higher standard errors and wider profile likelihood intervals for N . Some of her

comments about the use of a beta (BB) distribution were replied to by Dorazio and Royle (2005b), who also emphasized caution in choosing a mixture distribution.

12.6.1 *Models of Norris and Pollock*

Norris and Pollock (1995, 1996a,b) also developed some of the above models by fitting and estimating F , a distribution function for the catchabilities p_{h_g} , given $N = n$. Their model M_h is similar to M_{h_G} above except that they use a nonparametric maximum likelihood along with the EM algorithm to obtain an estimate F_n of F , given $N = n$, which is discrete. Then one finds the maximum for the pair (n, F_n) , say (n^*, F_{n^*}) leading to N^* , the nonparametric estimate of N . Bootstrapping from (n^*, F_{n^*}) was used to estimate conditional biases and standard deviations of the estimates. Model M_{bh} was investigated using a similar approach. The Akaike information criterion (AIC) can be used to select a model when comparing with other Pollock models (Norris & Pollock, 1995). The models are readily extended to include covariates.

12.6.2 *Models of Pollock and Otto*

Pollock and Otto (1983) also considered some of the above models, especially M_h and M_{bh} . They assumed that the catchabilities for the pairs of unmarked and marked (p_i, c_i) for $1, \dots, N$ were a random sample from some joint distribution function $G(p, c)$ ($0 \leq p \leq 1, 0 \leq c \leq 1$) so that the probability function of the a_ω (the number with capture history ω) is

$$\Pr(\{a_\omega\}) = \frac{N!}{\prod_\omega a_\omega!(N-r)!} \left[\prod_\omega P_\omega^{a_\omega} \right] (\text{E}[(1-p)^s])^{N-r},$$

where P_ω is the probability of capture history ω . For example, if ω is the capture history of being captured in all s samples,

$$P_\omega = \text{E}[pc^{s-1}] = \int_0^1 \int_0^1 pc^{s-1} dG(p, c).$$

This model includes models M_0 , M_b , M_h , and M_{bh} . If we knew the distribution G exactly and $q = 1 - p$, then an obvious estimator of N is given by $N - r \approx Nq^s$, the number not caught at all, leading to the practical estimator

$$\widehat{N}_1 = \frac{r}{1 - \text{E}[q^s]}.$$

For model M_b, $G(p, c)$ degenerates to two values, p and c , giving us

$$\widehat{N}_1 = \frac{r}{1 - \widehat{q}^s}.$$

This model has already been discussed above, and we now turn to the models with heterogeneity.

For M_h and M_{bh}, and looking at \widehat{N}_1 above, we need to consider

$$E[q^s] = \int_0^1 q^s g(p) dp,$$

where $g(p)$ is the marginal density function of p , the probability of first capture. However, we need the distribution of p given the probability of capture at least once, and this leads to the so-called weighted distribution (cf., Patil & Rao, 1978)

$$f^W(p) = \frac{(1 - q^s)g(p)}{(1 - E[q^s])},$$

that we now use. If the p_h 's are known exactly, an unbiased estimate of $(1 - E[q_h^s])^{-1}$ is the sample version $\sum_{h=1}^r (1 - q_h^s)^{-1}/r$ using the properties of weighted distributions. This leads to the weighted unbiased estimate

$$\widehat{N}_W = \sum_{h=1}^r (1 - q_h^s)^{-1},$$

a result originally derived by Overton (1969). To use \widehat{N}_W we need point estimators for the q_h for all animals captured, and these will depend on the model used.

For model M_h, the number of times i individual h is captured is binomial (s , p_h) so that the obvious estimator of p_h is the sample proportion i/s . \widehat{N}_W now becomes

$$\widehat{N}_h = \sum_{h=1}^r \frac{1}{1 - (1 - \widehat{p}_h)^s} = \sum_{i=1}^s \frac{f_i}{1 - (1 - i/s)^s},$$

where f_i is the number of animals captured i times. This estimator takes the form

$$\widehat{N} = \sum_{i=1}^s a_{is} f_i, \tag{12.15}$$

a linear combination of the capture frequencies (which form a minimal sufficient statistic) with constants a_{is} only depending on i and s . Other estimators of this type are $r = \sum_i^s f_i$, a biased estimator that can be used as a first step in constructing a

jackknife estimator and the jackknife estimator \widehat{N}_j of Burnham and Overton (1979). The estimator (12.15) has expectation

$$\mathbb{E}[\widehat{N}] = N \sum_{i=1}^s a_{is} \pi_i$$

and variance

$$\text{var}[\widehat{N}] = N \sum_{i=1}^s a_{is}^2 \pi_i - (\mathbb{E}[\widehat{N}])^2/N.$$

Here $\pi_i = \mathbb{E}\left[^s \binom{s}{i} p^i (1-p)^{s-i}\right]$ with the expectation over the distribution $g(p)$.

For model M_{bh} , only the time to first capture can be used because recaptures are influenced by trap responses. The number of samples to first capture (i) follows a geometric distribution which for individual h is given by $g(i) = p_h(1-p_h)^{i-1}$. The maximum likelihood estimator of p_h is $1/i$ so that our estimator of N now becomes

$$\widehat{N}_P = \sum_{i=1}^s \left[\frac{u_i}{1 - (1 - 1/i)^s} \right].$$

This estimator is of the form

$$\widehat{N} = \sum_{i=1}^s b_{is} u_i,$$

a linear combination of the number of animals u_i “removed” by marking. We note that $r = \sum_i u_i$. Such estimators have expectation

$$\mathbb{E}[\widehat{N}] = N \sum_{i=1}^s b_{is} \pi_i^*$$

and variance

$$\text{var}[\widehat{N}] = N \sum_{i=1}^s b_{is}^2 \pi_i^* - (\mathbb{E}[\widehat{N}])^2/N,$$

where $\pi_i^* = \mathbb{E}[p(1-p)^{i-1}]$ with the expectation over the density $g(p)$.

Finally, Pollock and Otto (1983) provided a number of jackknife estimates of N for various models.

12.6.3 Identifiability

In Sect. 12.5.5, when looking at models involving heterogeneity, we ran into the problem of a form of nonidentifiability with N . The mixture models described above involving heterogeneity face the same problem, which was considered by Holzmann et al. (2006). The authors also looked at the additional identifiability of a family of distributions that arises when using mixture distributions generally. Some mixture distributions do not cause problems. For example, if the mixture distribution is discrete and involves G groups, then we have identifiability if $2G \leq s$. Also the uniform distribution on $(0, a]$ is identifiable for $s \geq 2$, as well as the Beta distribution for $s \geq 3$. Mao and You (2009) also discussed this problem, focusing on various model failures, and revisited the beta and two-point mixture models, with an emphasis on some biases.

12.6.4 Model of Dorazio and Royle

Dorazio and Royle (2003) introduced a beta-binomial mixture to allow for heterogeneous rates of capture. Let x_i be the number of times individual i ($i = 1, \dots, N$) is captured (encountered) in s samples. For r different individuals captured, $x_i \in \{1, \dots, s\}$, and $x_i = 0$ for the $N - r$ not captured. It is assumed that individual captures are independent and that the x_i are a random sample from the beta-binomial distribution with parameters s, α, β . This implies that capture probabilities vary among individuals, and this variation can be modeled as a $\text{Beta}(\alpha, \beta)$ distribution. If $\mathbf{x} = (x_1, \dots, x_n)'$ is the observed data, the density function of the x_i is

$$f(\mathbf{x}; s, \alpha, \beta) = \prod_{i=1}^r \binom{s}{x_i} \frac{\Gamma(\alpha + \beta) \Gamma(x_i + \alpha) \Gamma(s - x_i + \beta)}{\Gamma(\alpha) \Gamma(\beta) \Gamma(s + \alpha + \beta)}.$$

If n_x is the number of individuals observed in x of the s sampling occasions, then the above distribution can be expressed in terms of a vector of sufficient statistics $\mathbf{n} = (n_1, \dots, n_s)'$, namely,

$$g(\mathbf{n}; s, \alpha, \beta) = \prod_{x=1}^s \left\{ \binom{s}{x} \frac{\Gamma(\alpha + \beta) \Gamma(x + \alpha) \Gamma(s - x + \beta)}{\Gamma(\alpha) \Gamma(\beta) \Gamma(s + \alpha + \beta)} \right\}^{n_x}.$$

If $n_0 = N - r$, the number of individuals not caught, then the complete likelihood was shown to

$$\begin{aligned} L(N, \alpha, \beta \mid \mathbf{n}) \\ = \frac{N!}{\prod_{x=0}^s n_x!} \prod_{x=0}^s \left\{ \binom{s}{x} \frac{\Gamma(\alpha + \beta) \Gamma(x + \alpha) \Gamma(s - x + \beta)}{\Gamma(\alpha) \Gamma(\beta) \Gamma(s + \alpha + \beta)} \right\}^{n_x} \end{aligned}$$

$$\begin{aligned}
&= \frac{N!}{(N-r)! \prod_{x=1}^s n_x!} \left\{ \frac{\Gamma(\alpha + \beta) \Gamma(\alpha) \Gamma(s + \beta)}{\Gamma(\alpha) \Gamma(\beta) \Gamma(s + \alpha + \beta)} \right\}^{N-r} g(\mathbf{n}; s, \alpha, \beta) \\
&= \frac{c N!}{(N-r)!} \left\{ \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \Gamma(\beta) \Gamma(s + \alpha + \beta)} \right\}^N \\
&\quad \times \{\Gamma(\alpha) \Gamma(s + \beta)\}^{N-r} \prod_{x=1}^s \{\Gamma(x + \alpha) \Gamma(s - x + \beta)\}^{n_x},
\end{aligned}$$

where $c = \prod_{x=1}^s \binom{s}{x}^{n_x} / \prod_{x=1}^s n_x!$ depends only on the data. Maximum likelihood estimates have to be found numerically, and Dorazio and Royle (2003) used the following transformation of the parameters $\mu = \alpha/(\alpha + \beta)$, $\tau = \alpha + \beta$, and $n_0 = N - r$ giving the log-likelihood function

$$\begin{aligned}
\ell(\mu, \tau, n_0 | \mathbf{n}) &= h(n_0 + n + 1) - h(n_0 + 1) + (n + n_0) \\
&\quad \times [h(\tau) - h(\tau\mu) - h\{\tau(1 - \mu)\} - h(s + \tau)] \\
&\quad + n_0[h(\tau\mu) + h\{s + \tau(1 - \mu)\}] \\
&\quad + \sum_{x=1}^s n_x [h(x + \tau\mu) + h\{s - x + \tau(1 - \mu)\}],
\end{aligned}$$

where $h(\cdot) = \log \Gamma(\cdot)$ and the constant c is omitted. A further 1–1 transformation is even better, namely, $\eta = \text{logit}(\mu)$, $\theta = \log(\tau)$, and $\phi = \log(n_0)$, as it places all parameters on a similar scale, ensures that estimates of μ , τ , and n_0 are confined to their admissible ranges without requiring constraints during optimization, and improves the condition of Hessian matrices required for computation, thus reducing any identifiability problems. Computational details and some examples are given by the authors as well as a discussion about identifiability.

12.6.5 Model of Tounkar and Rivest

Let p_h be the probability of capture for individual h ($h = 1, \dots, N$) and x_h the number of times it is captured during the s samples with \mathbf{z}_h covariates. Then $x_h \sim \text{Binomial}(s, p_h)$, and let $x_h > 0$ for $h = 1, \dots, r$, with r the number of different individuals captured. It is assumed that $p_h = f(\mathbf{z}'_h \boldsymbol{\beta}; a_h)$, where a_h is a random covariate with parameter $\boldsymbol{\alpha}$ to allow for unobserved heterogeneity. We then have (Tounkara & Rivest, 2015)

$$f(x_h | \boldsymbol{\alpha}, \boldsymbol{\beta}) = \binom{s}{x_h} \text{E}[p_h^{x_h} (1 - p_h)^{s-x_h}], x_h = 0, \dots, s,$$

where $E[\cdot]$ is the average over the unobserved a_h . Since the covariates are only available for the n observed individuals, we consider the probability of being caught at least once, namely,

$$\pi_h(\boldsymbol{\alpha}, \boldsymbol{\beta}) = 1 - f(x_h = 0 | \boldsymbol{\alpha}, \boldsymbol{\beta}) = 1 - E[(1 - p_h)^s],$$

and we end up with

$$f(x_h | x_h > 0, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{f(x_h | \boldsymbol{\alpha}, \boldsymbol{\beta})}{\pi_h(\boldsymbol{\alpha}, \boldsymbol{\beta})}, x_h = 1, \dots, s.$$

The likelihood function is

$$L(\boldsymbol{\alpha}, \boldsymbol{\beta}_0) = \prod_{h=1}^r f(x_h | x_h > 0, \boldsymbol{\alpha}, \boldsymbol{\beta})$$

from which can be obtained maximum likelihood estimates $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$ and the Horvitz-Thompson estimator

$$\hat{N} = \sum_{h=1}^r 1/\pi_h(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}).$$

Various models were considered for $f(x_h | \boldsymbol{\alpha}, \boldsymbol{\beta})$ and the distribution for the a_h .

12.7 Log-Linear Poisson Models

Fienberg (1972) applied the log-linear model to capture-recapture data for closed populations using a 2^s contingency table with one cell having a structural zero corresponding to one unobservable observation, namely, the number not caught at all ($\omega = (0, 0, \dots, 0)'$). By using the standard approach of defining main effects and interactions, the door was opened to a variety of models. Linear models have been applied to capture-recapture data for both open and closed models by Professor Richard Cormack in a number of papers, e.g., Cormack (1981, 1985, 1989, 1993, 1994), and some of this work is summarized by Seber and Schofield (2019, chapter 10).

A key aspect of these models is that they are based on a Poisson model given in the next section. This follows from considering the capture process as either a Poisson process with individuals being randomly distributed or as a limit of a binomial model with s tending to infinity and the probability of catching an individual tending to zero such that the product is a constant.

12.7.1 Using the Conditional Model

Following Huggins and Hwang (2011), if n_{ω} is the number of individuals with capture-recapture record $\omega = (\omega_1, \omega_2, \dots, \omega_s)'$ with each ω_i equal to 1 or 0 in an incomplete contingency table, the n_{ω} are assumed to be independent Poisson variables with means $\mu_{\omega} = E[n_{\omega}] = NP_{\omega}(\theta)$. Here $P_{\omega}(\theta)$ is the probability that an individual has a capture history ω , and θ is a set of model parameters. Recalling that $P(\theta) = \sum_{\omega} P_{\omega}(\theta)$ and a_{ω} is the number of individuals with capture history ω , we have that the joint probability function of the a_{ω} is

$$\begin{aligned} f(\{a_{\omega}\}) &= \prod_{\omega} \frac{\exp(-NP_{\omega}(\theta))(NP_{\omega}(\theta))^{a_{\omega}}}{a_{\omega}!} \\ &= \frac{r!}{\prod_{\omega} a_{\omega}} \prod_{\omega} \left\{ \frac{P_{\omega}(\theta)}{P(\theta)} \right\}^{a_{\omega}} \times \frac{\exp(-NP(\theta))(NP(\theta))^r}{r!} \end{aligned} \quad (12.16)$$

The first term above is the same as the first term of the conditional model of (12.8) in Sect. 12.4.2. If we treat N as continuous, we get the same estimate as the conditional estimate $\hat{\theta}$ of θ . The maximum likelihood estimate of N is then, from the second term of (12.16),

$$\hat{N} = r/P(\hat{\theta}).$$

The next step is to model $\ell_{\omega} = \log \mu_{\omega}$. For example, for model M_t , we get

$$\mu_{\omega} = N \prod_{i=1}^s p_i^{\omega_i} q_i^{1-\omega_i} = N \prod_{i=1}^s q_i \prod_{i=1}^s \left(\frac{p_i}{q_i} \right)^{\omega_i}.$$

Then $\log \mu_{\omega} = \gamma + \omega' \beta$, where $\gamma = \log N + \sum_{i=1}^s \log q_i$ and β has i th element $\beta_i = \log(p_i/q_i)$. We define the vector ℓ to have elements ℓ_{ω} so that $\ell = X\beta$ with $\theta = (\gamma, \beta')$ and X is a design matrix with the first column being 1's and having the rows corresponding to the appropriate ω . By fitting a log-linear Poisson model, we can estimate θ , the p_i from the β_i , and then the expected number of individuals not captured, namely, $\exp(\gamma) = \mu_0 = N \prod_{i=1}^s q_i$, finally leading to an estimate of N . Standard errors were discussed in Cormack and Jupp (1991) and Rivest and Lévesque (2001). An R package is available for the implementation (Baillargeon & Rivest, 2007).

12.7.2 Pollock's Models in Log-Linear Form

Pollock's models M_0 , M_h , M_{th} , and M_b can also be treated using standard log-linear models. For example, from Rivest and Lévesque (2001), we have for model M_0 and

n_{ω} the number with capture history ω ,

$$\mu_{\omega} = E[n_{\omega}] = Nq^s \prod_{i=1}^s (p/q)^{\omega_i},$$

and

$$\log \mu_{\omega} = \log(N) + x_1 \log(p) + (s - x_1) \log(q),$$

where the design matrix has two columns x_1 and $x_2 = \sum \omega_j$ is the number of times a unit is caught. Hence

$$\log(\mu_{\omega}) = x_0\gamma + x_1\beta,$$

where $\gamma = \log(Nq^s)$ and $\beta = \log(p/q)$. Also $\mu_0 = Nq^s = \exp(\gamma)$, so that we can estimate p from $\hat{\beta}$ and then N from $\hat{\gamma}$.

Model M_t is similar to M_0 and was described in the previous section. The model assumes that the s capture occasions are independent, which may not be the case. An extension to the time model to include heterogeneity, a version of model M_{th} , was given by Agresti (1994) and Darroch et al. (1993), who incorporated an interaction term, thus adding a term ω_{s+1} , where

$$\omega_{s+1} = \sum_{j>k} \omega_j \times \omega_k.$$

This gives us the model

$$\begin{aligned} \log \mu_{\omega} &= \gamma + \boldsymbol{\omega}' \boldsymbol{\beta} + \omega_{s+1}\alpha \\ &= x_0\gamma + \sum_{i=1}^s x_i\beta_i + x_{s+1}\alpha. \end{aligned}$$

A model M_h follows from M_{th} by setting all the β_i equal to β .

With model M_b , Rivest and Lévesque (2001) defined x_0 as above and then defined $x_1 = \min\{i : w_i = 1\} - 1$, the number of capture occasions before the first capture, and $x_2 = \sum(w_i) - 1$, the number of catches after the first one. This gives us

$$\log(\mu_{\omega}) = x_0 \log(Np) + x_1 \log(q) + x_2 \log \frac{c}{1-c} + (s - x_1 - 1) \log(1 - c).$$

When estimating N and p , c is a nuisance parameter, and a set of sufficient statistics for N and p is u_1, \dots, u_s , where u_i is the number of unmarked caught in sample i . Setting $\mu_i = E[u_i] = Nq^{i-1}p$, we can express the model M_b as

$$\log \mu_i = \gamma + (i - 1)\beta, \quad i = 1, \dots, s, , \quad (12.17)$$

where $\gamma = \log(Np)$ and $\beta = \log q$. The expected number not caught in the experiment is then $\mu_0 = Nq^s = \exp(\gamma + s\beta)/(1 - \exp(\beta))$, and we can get estimates of $N = e^\gamma/(1 - e^\beta)$, p , and μ_0 from estimates of γ and β .

When there is heterogeneity, as in M_{bh} , those caught early are not representative of those caught later, say after s_0 or more occasions. We assume that the first s_0 u -values do not follow (12.17), and let N_0 be the size of the population that has not been captured after s_0 samples. Then $N = N_0 + \sum_{i=1}^{s_0} E[u_i]$, and $\log \mu_{i+s_0} = \gamma + (i-1)\beta$, for $i = 1, \dots, s-s_0$, where $\gamma = \log N_0 p$, $\beta = \log(q)$, and p now stands for the probability of capture in samples s_0+1, \dots, s . Once γ and β are estimated from the log-linear model, we can estimate N from

$$N = \sum_{i=1}^{s_0} E[u_i] + \exp \gamma / (1 - \exp \beta),$$

where $E[u_i]$ can be estimated by u_i .

Asymptotic biases and variances, some bias corrections, and computational methods are discussed by Rivest and Lévesque (2001) for the above models. Before leaving this topic, we note that all the above models, including those of Pollock, can be incorporated into a general multinomial log-linear model as shown by Evans and Bonett (1994). Such an approach is useful for developing a computer package. Bias corrections are also given by Evans and Bonett (1994).

12.8 Coverage Models

We begin by considering a nonparametric approach using the idea of sample coverage introduced by I. J. Good and A. M. Turing (Good, 1953) that has been used in species and population size estimation. It has been applied to capture-recapture models with heterogeneous capture by Chao, Lee et al. (1992) and Lee (1996).

12.8.1 Heterogeneity

We begin with model M_h . In a random-effects model, the capture probabilities p_h of individuals are assumed to follow some distribution determined by two parameters. In a fixed-effects model, it is assumed that all the heterogeneity effects are characterized by the population mean and coefficient of variation of the p_h 's, namely,

$$\bar{p} = \sum_{h=1}^N p_h / N \tag{12.18}$$

and the coefficient of variation

$$\gamma = \frac{[\sum_{h=1}^N (p_h - \bar{p})^2 / N]^{1/2}}{\bar{p}}. \quad (12.19)$$

It is assumed that animals act independently and the fixed-effects approach is used.

Following Huggins and Hwang (2011, section 3.3), let x_h be the number of times individual h is captured in the experiment. If $I(x_h > 0)$ is an indicator function taking the value of 1 when $x_h > 0$ is true, and 0 otherwise, then the sample coverage is defined by

$$C = \sum_{i=1}^N p_h I(x_h > 0) / \sum_{h=1}^N p_h, \quad (12.20)$$

the proportion of the total probabilities that turn up during the experiment. Then

$$E[C] \approx 1 - \left(\sum_{h=1}^N s p_h q_h^{s-1} / \sum_{h=1}^N s p_h \right),$$

If f_k is the number of individuals caught exactly k times, then a moment estimator of the above expression is

$$\hat{C} = 1 - f_1 / \left(\sum_{k=1}^s k f_k \right).$$

When s is small, the estimator may be biased, and a further bias correction of \hat{C} can be found in Chao, Lee et al. (1992). Since $C = r/N$ if all the p_h are equal, as with no heterogeneity, an estimator of N is $\widehat{N}_0 = r/\hat{C}$. However, with heterogeneity, it has a negative bias which, asymptotically, is given by

$$E[\widehat{N}_0] \approx N - (E[f_1]/E[C])\gamma^2.$$

A bias corrected sample coverage estimator is then

$$\widehat{N} = (r/\hat{C}) + (f_1/\hat{C})\hat{\gamma}^2,$$

where

$$\hat{\gamma}^2 = \max \left\{ \left[s \widehat{N}_0 \sum_{k=1}^s k(k-1) f_k \right] / \left[(s-1) \left(\sum_k f_k \right)^2 \right] - 1, 0 \right\}.$$

Using a Poisson model, Chao (1987) applied the Cauchy-Schwarz inequality to establish a simple lower-bound estimator $\widehat{N}_L = r + f_1^2/(2f_2)$, if s is not too small.

The estimator has been shown to perform well in species richness estimation and is called the Chao 1 estimator in the ecology literature. Using the delta method, Chao also obtained an asymptotic variance estimator for it, namely,

$$\widehat{\sigma}_L^2 = f_2[0.25(f_1/f_2)^4 + (f_1/f_2)^3 + 0.5(f_1/f_2)^2].$$

Treating $\log(\widehat{N}_L - r)$ as approximately normal, an approximate 95% confidence interval for N is

$$[r + (\widehat{N}_L - r)/d, r + (\widehat{N}_L - r)d],$$

where

$$d = \exp\{1.96[\log(1 + \widehat{\sigma}_L^2/(\widehat{N}_L - r)^2)^{1/2}\}.$$

Chao (1989) focussed on the case of sparse data when individuals are caught only once or twice in the s samples, which enabled her to obtain a more direct derivation of the previous estimator in the form $r + (s - 1)f_1^2/(2sf_2)$, which reduces to \widehat{N}_L for large s .

12.8.2 Coverage and Pollock's Models

Having considered the important case of heterogeneity, we now look at estimators for all of Pollock's models given by Lee and Chao (1994) using the idea of coverage. Recalling that p_{hi} is the probability that an individual h is captured in sample i , the authors used the model M_{tbh} with $p_{hi} = p_h e_i$ for any first capture and $p_{hi} = c_h e_i^*$ for any recapture. Model M_{tb} then follows from this by setting $p_h = p$ and $c_h = c$. For M_{th} , $p_{hi} = p_h \alpha_i$, where α_i is an unknown time effect of the i th sample on capture. The effects are in a “relative sense” so that p_h and e_i are defined only up to a multiplicative constant in M_{tbh} . This means that \bar{p} of (12.18) is not uniquely defined, but the coefficient of variation (12.19) and the sample coverage (12.20) are, in this case. For model M_{tbh} ,

$$E[r] = N - \sum_{h=1}^N \prod_{i=1}^s (1 - p_h e_i),$$

and

$$E[C] = 1 - \sum_{h=1}^N p_h \prod_{i=1}^s (1 - p_h e_i) / \sum_{h=1}^N p_h.$$

Chao, Lee et al. (1992) proved that

$$\frac{E[r]}{E[C]} = N - \frac{N\bar{p}\sum_{i=1}^s(e_i)\prod_{k\neq i}^s(1-\bar{p}e_k)}{E[C]}\gamma^2 + R_1, \quad (12.21)$$

where R_1 is a term involving only the third and fourth central moments of the p_h 's. Replacing the e_i 's by α_i 's in (12.21), we have the same conclusion for model M_h . For models M_h and M_{bh} , (12.21) reduces to

$$\frac{E[r]}{E[C]} = N - \frac{Ns\bar{p}(1-\bar{p})^{s-1}}{E[C]}\gamma^2 + R_2, \quad (12.22)$$

where R_2 is a lengthy expression. We now consider some estimators of N beginning with model M_{tbh} .

The relevant statistics for this model are the unmarked captured in the samples, namely, the $\{u_i\}$. Defining C_k as the sample coverage of the first k samples ($k \leq s-1$), then for just those k samples (12.22) without the remainder term becomes

$$N \approx \frac{E[r_k]}{E[C_k]} + \frac{N\bar{p}\sum_{i=1}^k(e_i)\prod_{j\neq i}^k(1-\bar{p}e_j)}{E[C_k]}\gamma^2.$$

Lee and Chao (1994) then derived the following estimator based on the first k samples:

$$\widehat{N}(k) = \frac{r_k}{\widehat{C}_k} + \frac{A_k}{\widehat{C}_k}\widehat{\gamma}_k^2,$$

where

$$A_k = (u_{k+1}/e_{k+1})\sum_{j=1}^k(e_{j+1}u_j/u_{j+1}),$$

$$\widehat{C}_k = 1 - \frac{u_{k+1}/e_{k+1}}{u_1/e_1}, \quad k = 1, \dots, s-1,$$

$$\widehat{\gamma}_k^2 = \max\{\widehat{N}_0(k)[u_1 - u_2(e_1/e_2)]/u_1^2 - 1, 0\},$$

and $\widehat{N}_0(k) = r_k/\widehat{C}_k$. The estimator $\widehat{N}(k)$ ($k = 1, \dots, s-1$) is independent of the scale of the relative effects e_i . The authors suggested choosing $k = s-1$ if $u_s/e_s < u_1/e_1$, since all the data are then used. If $u_s/e_s \geq u_1/e_1$ and s is large, they suggested choosing k such that $u_{k+1}/e_{k+1} < u_1/e_1$ ($k \leq s-2$) and discarding the last $s-k-1$ observations u_{k+2}, \dots, u_s . The authors discussed variance estimation and gave a list of all the estimators for the remaining seven models.

Lee (1996) revisited all the models with a main focus on M_{tb} and M_{tjh} and avoiding the problem of estimating the ratio of the relative time effect between the first sample and other samples (e_1/e_r) required in the above theory. She began with the model M_{tjh} , which assumed $p_{hi} = p_{he_i}$ for any first capture and $p_{hi} = b_{he_i}$ for any recapture, treating (p_1, \dots, p_N) and e_1, \dots, e_s as fixed parameters, except in the derivation of the variance of estimators. She used the idea of the conditional sample coverage defined as the sample coverage for dropping out the animals which are captured in the first sample and provided a number of estimators of N . Her paper should be consulted for details.

12.8.3 Coverage and Dual Record System

The dual record system was considered in Sect. 12.2.6, and we now briefly consider (Chao & Tsay, 1998) with reference to the two types of dependency previously considered. We recall that we used the data matrix \mathbf{X} where $x_{hi} = 1$ if individual h is listed in sample i and zero otherwise. We can write $x_{hi} = I(\text{individual } h \text{ is listed in sample } i)$, where $I(A)$ is the usual indicator function. The authors defined $\mu_i = N^{-1} \sum_{h=1}^N E[x_{hi}]$ (the average inclusion probability for list i) and the coefficient of covariation (CCV) of samples i and j as

$$\gamma_{ij} = \frac{1}{N} \sum_{h=1}^N E[(x_{hi} - \mu_i)(x_{hj} - \mu_j)] / (\mu_i \mu_j).$$

Here the magnitude of γ_{ij} measures the degree of dependence of samples i and j . In the independent case, all CCVs are zero. In a similar fashion, we can define the CCV for samples i_1, \dots, i_r as

$$\gamma_{i_1 \dots i_r} = \frac{1}{N} \sum_{h=1}^N E[(x_{hi_1} - \mu_{i_1}) \dots (x_{hi_r} - \mu_{i_r})] / (\mu_{i_1} \dots \mu_{i_r}).$$

From Sect. 12.2.1 (measure of independence)

$$N = \frac{E[n_1]E[n_2]}{E[m_{12}]}(1 + \gamma_{12}),$$

which shows the importance of γ_{12} and the need to be able estimate it.

Only Heterogeneity Between Individuals Exists

If there is no dependence between captures of an individual, then conditional on individual h , the capture histories $\{x_{h1}, x_{h2}, \dots, x_{hs}\}$ are independent. Since the

probability $\Pr(x_{hi} = 1 | history)$ does not depend on *history*, $\Pr(x_{hi} = 1) = p_{hi}$, $\mu_i = \sum_h p_{hi}/N$, and

$$\begin{aligned}\gamma_{ij} &= \frac{1}{N} \sum_{h=1}^N [(p_{hi} - \mu_i)(p_{hj} - \mu_j)] / (\mu_i \mu_j) \\ &= \frac{1}{N} \sum_{h=1}^N \frac{p_{hi} p_{hj}}{\mu_i \mu_j} - 1.\end{aligned}$$

This class of model includes three widely used models:

1. Rasch (1961) model. This is given by

$$\Pr(x_{hi} = 1) = \exp(v'_h + a'_i) / [1 + \exp(v'_h + a'_i)],$$

which can reparameterized as

$$\Pr(x_{hi} = 1) = v_h a_i / (1 + v_h a_i),$$

where v_h denotes the heterogeneity effect of individual h and a_i denotes the sample effect of the i th list. The values of v_h are defined only up to a multiplicative constant. In a fixed-effects model, all the v_h 's are treated as parameters, whereas in a random-effects model, the v_h ($h = 1, 2, \dots, N$) are a random sample from a distribution. In the latter case, we use the conditional probability $\Pr(x_{hi} = 1 | v_h)$.

2. Generalized Rasch (1961) model (Darroch et al., 1993)

$$\Pr(x_{hi} = 1) = \begin{cases} v_h a_i / (1 + v_h a_i), & i = 1, 2, \dots, k, \\ v_h^* a_i / (1 + v_h^* a_i), & i = k + 1, \dots, s. \end{cases}$$

Here the patterns of heterogeneity effects are different in two separate groups, and the model can be extended to more than two types of groups.

3. Model M_{th}. Here we have the model $\Pr(x_{hi} = 1) = p_h e_i$ and, for all i and j ,

$$\gamma_{ij} = \left[\sum_h (p_h - \bar{p})^2 / N \right] / \bar{p}^2,$$

where $\bar{p} = \sum_h p_h / N$. Similarly we have for all i, j and k ,

$$\gamma_{ijk} = \left[\sum_h (p_h - \bar{p})^3 / N \right] / \bar{p}^3.$$

Only List Dependence Exists

We now have $\Pr(x_{hi} = 1 | \text{all } h)$ is independent of h so that $\Pr(x_{hi} = 1) = p_i$, the probability of being listed in sample i . Also $\Pr(x_{hi} = 1, x_{hj} = 1) = P_{ij}$, and there is a similar expression for P_{ijk} . Then

$$\gamma_{ij} = \frac{P_{ij}}{p_i p_j} - 1,$$

and list independence occurs if and only if $P_{ij} = p_i p_j$ or $\gamma_{ij} = 0$.

Having highlighted the difference between the two types of problem, we have a difficulty with the two sample cases as we have four parameters μ_i , μ_j , N , and γ_{ij} , but only data in three cells of the 2×2 table of capture data. Clearly, a third sample is needed. In this case, Chao and Tsay (1998) provided an estimate of N via sample coverage, and Tsay and Chao (2001) considered the case of more than three samples.

12.9 Martingales and Estimating Functions

For the model M_t , Yip (1991) used a martingale based on the estimating equation

$$D_i = (N - M_i)m_i - M_i u_i$$

which, conditional on the history of the process up to and including sample $i - 1$, has mean zero. He obtained a number of estimators of N of which the optimal one is the solution of

$$\sum_{i=1}^s \frac{1}{N - n_i}(N - M_i)m_i - M_i u_i = 0,$$

and obtained its asymptotic properties and a variance estimate. One of the estimators of N was

$$\widehat{N} = \sum_{i=1}^s M_i n_i / \sum_{i=1}^s m_i$$

originally obtained by Schnabel (1938), where M_i is the number tagged just prior to the i th sample. Using the method of moments for martingales, it was also obtained by Yip (1989) and Huggins (1989, 2004), who also considered a discrete version of the continuous case, building on the work of Becker (1984).

12.9.1 Combined with Coverage

Having discussed martingale and coverage methods above, it is perhaps not surprising that the two methodologies have been combined. Chao, Yip et al. (2001b) provided methods for analyzing all of Pollock's models subject to certain assumptions beginning with the most general model M_{tbh} . This is defined as $p_{hi} = p_{he_i}$ until first capture and ϕp_{he_i} for any recapture. It is a strong assumption as it assumes that the recapture probabilities are proportional to the initial capture probabilities. However, it has the advantage that all the six possible submodels are readily obtained by selectively applying some constraints such as $e_i = 1$ (model M_{bh}), $p_h = 1$ (model M_{tb}), $\phi = 1$ (model M_{th}), $e_i = 1, \phi = 1$ (model M_{h}), $p_h = p$ and $e_i = 1$ (model M_{b}), and $p_h = 1$ and $\phi = 1$ (model M_{t}).

Let $\boldsymbol{\theta}$ denote the vector of parameters involved in the model, and let \mathbf{g}_i be an unbiased estimating function associated with sample i . The estimating functions considered by the authors are $\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_s$, where the weight associated with \mathbf{g}_i is allowed to be a function of $\boldsymbol{\theta}$ and the past up to and including sample $i - 1$. We define E_c , var_c , cov_c , and Cov_c to denote, respectively, expectation, variance, covariance, and variance-covariance matrix conditional on the past up to and including sample $i - 1$. The optimal estimating equation is given by (see Liang and Zeger (1995) for necessary background)

$$\mathbf{g} = \sum_{i=1}^s \mathbf{D}'_i \mathbf{V}_i^{-1} \mathbf{g}_i = \mathbf{0},$$

where $\mathbf{D} = E_c[\partial \mathbf{g}_i / \partial \boldsymbol{\theta}]$ and $\mathbf{V}_i = \text{Cov}_c[\mathbf{g}_i]$, the conditional covariance matrix of the elements of \mathbf{g}_i . Here “optimality” is in the sense that the corresponding root $\boldsymbol{\theta}$ of $\mathbf{g} = \mathbf{0}$ has minimum asymptotic variance in the class of weighted linear functions of $\mathbf{g}_1, \dots, \mathbf{g}_s$. The quantities $\mathbf{D}'_i \mathbf{V}_i^{-1}$ play the role of the optimal weights.

Chao, Yip et al. (2001b) listed the appropriate equations for each of the seven models, so we will look briefly at just two models. We first begin with M_t , which does not involve sample coverage and which we have already referred to briefly above.

Model M_t

For each sample i , we have

$$\mathbf{g}_i = \begin{pmatrix} g_{1i} \\ g_{2i} \end{pmatrix} = \begin{pmatrix} u_i - (N - M_i)e_i \\ m_i - M_i e_i \end{pmatrix}.$$

Here $\boldsymbol{\theta}$ consists of the parameters N and the e_i , and the authors end up with $s + 1$ equations

$$\sum_{i=1}^s [(N - M_i)(1 - e_i)]^{-1} [u_i - (N - M_i)e_i] = 0, \quad n_i - Ne_i = 0 \quad (i = 1, \dots, s),$$

leading to

$$\sum_{i=1}^s [(N - M_i)(N - n_i)]^{-1} [M_i u_i - (N - M_i)m_i] = 0$$

to be solved for N . The above equation was originally derived by Yip (1991), who expressed $M_i u_i - (N - M_i)m_i$ as a martingale difference and $[(N - M_i)(N - n_i)]^{-1}$ as the optimal weight associated with the martingale difference.

Model M_h

We now look at model M_h , which involves the sample coverage C_j of samples $1, 2, \dots, j$ and the coefficient of variation γ of the p_h . The sample coverage is defined to be

$$C_j = \sum_{h=1}^N p_h I\left(\sum_{i=1}^j x_{hi} > 0\right) / \sum_{h=1}^N p_h,$$

with $C_0 = 0$, where $I(\cdot)$ is the usual indicator function and $x_{hi} = 1$ if individual h is caught in sample i , and 0 otherwise. Here $\theta = (N, \bar{p})'$. We now have for each sample i the estimating equations (Huggins & Chao, 2002)

$$\mathbf{g}_i = \begin{pmatrix} g_{1i} \\ g_{2i} \end{pmatrix} = \begin{pmatrix} u_i - (N - M_i^*)\bar{p} \\ m_i - M_i^*\bar{p}_i \end{pmatrix},$$

where $M_i^* = NC_{i-1}$ ($i \geq 2$, $M_1^* = 0$). Conditional on past records up to and including sample $i - 1$ and conditional on all the p_h , we have

$$E[u_i] = (N - NC_{i-1})\bar{p}, \quad \text{and} \quad E[m_i] = NC_{i-1}\bar{p},$$

along with two variance expressions. We recall that $\bar{p} = \sum_{h=1}^N p_h/N$. Making some variance approximations and assuming C_{i-1} and M_i^* are known, the optimal estimating equations now take the form

$$\sum_{i=1}^s (1 - C_{i-1})^{-1} [u_i - (N - M_i^*)\bar{p}] = 0 \quad \text{and} \quad \bar{p} = (sN)^{-1} \sum_{i=1}^s n_i.$$

Estimates of M_i^* (which involves an estimate of γ) and C_i are given that lead to the estimator of N

$$\hat{N} = \frac{\sum_{i=1}^s (1 - \hat{C}_{i-1})^{-1} \hat{M}_i^*}{\sum_{i=1}^s (1 - \hat{C}_{i-1})^{-1} (1 - su_i / \sum n_j)}.$$

Further details are given by Huggins and Chao (2002) and by Chao, Yip et al. (2001b) for all the models. They used a non-parametric bootstrap to estimate variances and confidence intervals for the various estimators of N .

12.10 Time-to-Detection Method

Alldredge et al. (2007b) developed a time-to-detection method, which is essentially a removal method, for estimating abundance from a point-count survey and for estimating detection probabilities (cf. Sect. 4.3.5). It was specifically designed to account for variation in detection probabilities associated with the singing rates of birds. The model accounts for both availability and detection by modeling the *combined* probability that a bird sings during the count and the probability that it is detected when it sings. The model requires dividing the count interval into several intervals, say K , and recording detections of birds in each interval. It requires observers to accurately track individual birds through the count. The detection (“capture”) history for each bird consists of one of 2^{K-1} possible detection histories. Zeros in a history represent either individuals who did not sing during a given interval (i.e., were not available) or individuals who sang but were not detected by the observer. This approach with a redefinition of “capture” can be incorporated into all of Pollock’s models but allowing for the different locations, as follows:

- Model M_0 : Equal detection probability for all individuals among all intervals.
- Model M_t : Equal detection probability for all individuals but different detection probabilities among intervals.
- Model M_b : Equal probability of first detection for all individuals among all intervals and a unique probability of subsequent detections that is equal for all individuals among all subsequent intervals.
- Model M_{tb} : Equal probability of first detection for all individuals but different among intervals and a unique probability of subsequent detection that is equal for all individuals but different among subsequent intervals.
- Model M_h : Unique probability of detection for each individual that remains constant among all intervals.
- Model M_{th} : Unique probability of detection for each individual that differs among intervals.
- Model M_{bh} : Unique probability of first capture for each individual that remains constant among intervals and a unique probability of subsequent detection that remains constant among intervals.
- Model M_{tjh} : Unique probability of first capture for each individual that differs among periods and a unique probability of subsequent detection that differs among intervals.

Models M_b and M_{bh} are equivalent to the removal models given by Farnsworth et al. (2002). Covariates can also be included in all the above models to account for individual differences in detection probabilities among birds, such as detection

distance, time of day, time of year, weather conditions, singing rate, and so on. Also several species were considered at the same time, and use was made of any commonalities among the species to reduce the number of parameters to be estimated. Alldredge et al. (2007c) carried out a field simulation using simulated bird songs to investigate the time-to-detection method using four 2-minute intervals. They also used the two-point mixture models to introduce heterogeneity where there is a proportion of the animals with one capture probability and the remainder of the animals in the population with another capture probability. Errors due to wrong interval recording and double counting were determined, and the method was found to generally work well, particularly when birds were located by their calls.

Alldredge et al. (2008) carried out a similar kind of study to evaluate distance sampling (using three different detectability functions) and independent double-observer methods. They found that common assumptions about the accuracy of distance estimation, and the ability of observers to accurately map and match observations, are often not met on auditory point counts. Covariates should be included in these analyses to allow for observer and distance effects, as well as for factors like habitat and ambient noise.

12.11 Prior Detection Effects

Survey methods that account for detection probability often require repeated detections of individual birds or repeated visits to a site to conduct counts or collect presence-absence data, e.g., repeat counts and time of detection methods. These initial encounters may influence detection probabilities for subsequent encounters. For example, observers may be more likely to re-detect a species or individual once they are aware of the presence of that species or individual at a particular site. This can lead to an increase in the probability of redetection. This idea is reminiscent of what can happen in capture-recapture experiments where an initial capture can affect later captures due to trap addiction or trap shyness.

Riddle et al. (2010) discussed and considered the capture-recapture models $M_{b(tod)}$ (observer-based behavioral effect at the individual level), $M_{t(tod)}$ (time effect), and $M_{0(tod)}$ (constant probability of detection) from time-to-detection (tod) data (Sect. 12.10). In the case of repeated occupancy (occ) models, they considered the capture-recapture models $M_{b(occ)}$, $M_{t(occ)}$, and $M_{0(occ)}$ from (occ) data. It was found that probability of subsequent detections increased in all data sets after an initial detection. Further details and discussion about prior detection effects for various models were given by the authors.

12.12 Using DNA Tags

Boulanger et al. (2008) introduced a two-sample Petersen experiment based on two methods of collecting DNA material from grizzly bears and using separate pooling of data. The first method obtained baited, grid-based, hair snag samples, while the second used hair from un-baited bear-rub trees. Hair snag captures were treated as the initial session with rub tree captures as the recapture session leading to an estimator in program MARK that treats hair snag and rub tree samples as successive sessions. One reason for two data collecting methods is that the first method should be of sufficient sampling intensity to produce high recapture rates, which can be expensive to conduct and can lead to low levels of precision in large-scale projects. In contrast, since rubbing is a natural behavior of grizzly and black bears range-wide, collecting hair from bear rubs found on trails is much cheaper to collect. A short piece of barbed wire can improve such hair collection, which did not appear to discourage rubbing. Ruell et al. (2009) used a multi-sample capture-recapture using DNA from scats for a bobcat population. The sampling period was short to maintain closure, and the genetic error was very low.

Kohn et al. (1999) described how the genotyping of feces can be used to estimate abundance of coyotes from feces collected along paths or roadways where coyotes, like most carnivores, often defecate and mark territorial boundaries. DNA is extracted from the feces, and species identity and sex are determined by mitochondrial DNA and the Y chromosome using multilocus typing. A single item of feces is obtained by choosing it at random from the sample for typing. To reduce effort, it was decided arbitrarily that when only one new genotype was discovered in 30 consecutively analyzed feces, the typing would stop. The total number of genotypes was incremented if the newly found genotype was unique.

Assuming that each unique genotype represents a different animal, a plot of $y = (ax)/(b + x)$ is carried out, where y is the cumulative number of unique genotypes and x is the number of feces sampled. When $x \rightarrow \infty$, $y \rightarrow a$ the asymptote of the curve, which is our estimate of N the population size. The method mirrors that used in estimating the number of species using the “species accumulation curve” (Sect. 5.5.6). The parameters a and b and their confidence interval were estimated using nonlinear regression (Seber & Wild, 1989). The authors also compared their results with those using capture-recapture (CR) analysis and telemetry.

One difference between CR and the above noninvasive DNA method is that the latter can be used for just a single sample. Petit and Valiere (2006) compared the population size estimate for the above noninvasive estimate with the CR estimate and a Bayesian estimate using simulation. They found that without heterogeneity, the DNA method performed as well as the classical CR. The Bayesian estimate presented the best compromise between low mean-squared error and a 95% confidence interval encompassing the parametric value of N in most simulations.

DNA tags are considered in more detail in Chap. 14 where we look at spatial models, usually combined with camera trap methods. Misidentification with DNA tags is discussed in the next section.

12.13 Misidentification

We consider two types of misidentification, namely, , when using DNA tags or photographic identification. In the past, several authors have tackled this problem with DNA tags (e.g., Link et al., 2010, using a Bayesian approach, and Yoshizaki et al., 2011) and photo identification Yoshizaki et al. (2009). They ran into likelihood problems, which were resolved by Vale et al. (2014) referring to a model $M_{t,\alpha}$. This was an extension of M_t described above, where with each time t ($t = 1, \dots, T$), an individual is detected (captured) and identified correctly with probability α . Therefore, associated with each time t is a capture probability p_t , with three following possible outcomes for each animal:

- (1) Not captured, with probability $(1 - p_t)$ (code 0).
- (2) Captured and correctly identified, with probability αp_t , called a *sound capture* (code 1).
- (3) Captured and misidentified, with probability $(1 - \alpha)p_t$, called a *faulty capture* (code 2).

This means we have $J = 3^T$ true, unobservable (latent) possible capture histories λ_j ($j = 1, \dots, J$) for each animal. An identification may be defined as “correct” or “sound” if an accurate genotype is obtained or if a photograph adequately captures the individual’s distinguishing features. In practice, it will not be known whether the identification is sound or not. It is assumed that the outcome at time t is independent of outcomes at other times and for other animals. For example, if $T = 5$, the latent history $\lambda_j = 10,122$ denotes an animal with sound captures at times 1 and 3, noncapture at time 2, and faulty captures at times 4 and 5.

Let λ_{jt} be the history λ_j at time t . The latent variable underlying the observed data is the vector of frequencies $\mathbf{x} = (x_1, \dots, x_J)'$, where x_j is the number of the N animals whose true capture history is λ_j and \mathbf{x} is multinomial with index N . Model $M_{t,\alpha}$ makes the assumption that each faulty capture generates a unique error, giving rise to an observed history with exactly one entry. For example, animals with a history $\lambda = 10,122$ generates three observed histories: 10100, 00010, and 00001. Any observed history with exactly one entry is called a *unit history*, while a history whose single entry derives from a faulty capture (a 2 in the latent history) is called a *ghost history*. An observed history with more than one entry is termed a *duplicate history*.

Excluding the unobservable zero history, there are $K = 2^T - 1$ possible observed histories, which are written as $\omega_1, \dots, \omega_K$, with corresponding observed frequencies f_k giving the vector $\mathbf{f} = (f_1, \dots, f_K)'$. Here the f_k do not refer to numbers of animals but to numbers of observed histories. We therefore have that the latent histories are indexed by j ($j = 1, \dots, J$), observed histories by k ($k = 1, \dots, K$), capture occasions by t ($t = 1, \dots, T$), and from Link et al. (2010)

$\mathbf{f} = \mathbf{A}'\mathbf{x}$ for a known matrix \mathbf{A} . For observed history ω_k , let $|\omega_k| = \sum_{t=1}^T \omega_{kt}$ be the number of 1-entries in ω_k , e.g., if $\omega_k = 101000$, then $|\omega_k| = 2$. Let

$$n_t = \sum_{k:\omega_{kt}=1} f_k, \text{ the numbers of capture at time } t.$$

$u_t =$ the number of unit histories that occur at time t .

$$U = \sum_{t=1}^T u_t = \sum_{k:|\omega_k|=1} f_k, \text{ number of unit histories in the observed data.}$$

$$D = \sum_{k:|\omega_k|\geq 2} f_k, \text{ number of duplicate histories in the observed data.}$$

$$d_t = \sum_{k:|\omega_k|\geq 2} f_k \omega_{kt}, \text{ number of captures within the } D \text{ duplicate histories}$$

that occur at time t .

$$C = \sum_{t=1}^T d_t = \sum_{t=1}^T n_t - U, \text{ the total captures in the } D \text{ duplicate histories.}$$

Here, all captures in D histories are sound, by definition. However, animals contributing to these histories may also contribute to ghost histories. The C captures are all known to be sound, by assumption.

We calculate the probability of the data by partitioning over choices for the number of sound captures among the u_t uncertain captures at time t . Define the latent variable $\mathbf{r} = (r_1, \dots, r_T)'$ such that r_t is the number of the u_t uncertain captures that are sound. Let $r. = \sum_{t=1}^T r_t$ be the total number of animals with exactly one sound capture, noting that any of these animals may have additional faulty captures at other times. Let \mathcal{R} be the feasible set of all \mathbf{r} compatible with the data \mathbf{f} and parameter N , so that $\mathcal{R} = \{\mathbf{r} : r. \leq N - D \text{ and } r_t \in \{0, \dots, u_t\} \text{ for } t = 1, \dots, T\}$. For $N \geq \max\{D, N_1, \dots, n_T\}$, we have the likelihood

$$L(N, \{p_t\}, \alpha; \mathbf{f})$$

$$\begin{aligned} &= \left\{ \alpha^C \prod_{t=1}^T p_t^{n_t} (1 - p_t)^{N-n_t} \right\} \\ &\times \left\{ \sum_{\mathbf{r} \in \mathcal{R}} \frac{N! \alpha^{r.} (1 - \alpha)^{U-r.}}{(\prod_{k:|\omega_k|\geq 2} f_k!) r_1! \cdots r_k! (N - D - r.)!} \right. \\ &\times \left. \prod_{t=1}^T \binom{N - d_t - r_t}{u_t - r_t} \right\}. \end{aligned}$$

Maximum likelihood estimates were obtained using the above likelihood. Further details were given by the authors.

With regard to photographic identification using natural markings for tags, misidentification can occur. Tancredi et al. (2013) considered this problem and assumed there were two catalogues of photographs obtained during two different sampling occasions. Applying the M_t model for two samples, they used a matching model for individual values based on the multivariate normal distribution.

12.14 Sampling One at a Time

With some populations it may be only feasible to sample and tag (or identify) one individual at a time so that the above theory applies by setting $n_i = 1$ for all i . As well as captures, animal “signs” can also be used, for example, photos, hair from snares, dwellings, and scats deposited in continuous space and time (see also the examples to papers given by Miller et al., 2005). Originally a discrete model was considered by Craig (1953) for the study of butterfly populations and developed further by Darroch (1958) (cf. Seber, 1982, 136–139). Since then, continuous time models have been developed where, in addition to tagging the animal, the exact capture time may be recorded with each capture regarded as a trapping occasion. Becker (1984) constructed an estimating equation approach for this situation that was built on by others. For example, Hwang et al. (2002) extended this work further and considered a modification of the model M_{tb} in which the first capture intensity is proportional to the recapture intensity. The special cases of M_t , M_b , and M_0 were then considered. They found that their likelihood approach showed some degree of robustness with respect to the proportional-constant assumption.

Wright et al. (2009) introduced a model in which DNA genotyping error resulting from allelic dropout was incorporated into a mark-recapture type model for field sampling of DNA. Here they used DNA in shed items such as feces, hair, or feathers, and their description of field sampling was based on an urn-sampling model proposed by Miller et al. (2005). This model is implemented in a software called CAPWIRE (henceforth the CAPWIRE model), which treats DNA fragments as drawn one at a time from the population. Some of problems with DNA matching are described by Seber and Schofield (2019, section 2.1.4.).

Motivated by field sampling of DNA fragments, Barker et al. (2014) developed a model based on Poisson sampling where, in addition, the sampling time may not be observed. They showed that previously described models correspond to partial likelihoods from their Poisson model, and their use may be justified through arguments concerning S- and Bayes ancillarity of discarded information. They mentioned the following features of the process: (i) the animal signs sampled in space and time are representative of all the animal signs in the study area; (ii) sampling may occur in distinct sessions (e.g., days), but within each session, we have sampling in continuous-time or space; (iii) the number of signs collected is not predetermined but is considered a random outcome of the sampling experiment;

(iv) during a prescribed sampling period, an individual may be represented in more than one sample, and this may not be known until samples have been processed to determine identities. We now briefly consider their models.

Beginning with just one session, let T be the number of individual samples taken. The first concern is to reconcile two different approaches to field sampling of DNA, namely, conditioning on T or not. Let $y_j \in \{1, \dots, N\}$ denote the identity of the individual appearing in sample j ($j = 1, \dots, T$). Let $\mathbf{y} = (y_1, \dots, y_T)'$ be the ordered (by sample) collection of identities. Any one individual may be represented more than once in the T samples, where T is regarded as random.

Let c_i ($i = 1, \dots, N$) denote the number of times individual i is identified, let $\mathbf{c} = (c_1, \dots, c_N)'$, and let n be the number of distinct signs (individuals) seen. Using the indicator function $I(\cdot)$ we see that $c_i = \sum_{j=1}^T I(y_j = i)$ and the statistic $T = \sum_{i=1}^N c_i$. Let t_{ij} ($j = 1, \dots, c_i$) denote the times or locations of the sample in which individual i was identified. Then the latent times $\{t_{ij}\}$ (summarized as \mathbf{t}) are modelled for each i and $j = 1, \dots, c_i$ as a realization of an individual-specific Poisson process with intensity $\lambda_i(t)$ for $0 \leq t \leq \tau$, where τ is the duration of the study. This formulation is sufficiently general to accommodate session effects. For example, if the sampling intensity is constant within a session but varies between sessions, then $\lambda_i(t)$ is a step function with steps at the start of each new session.

Let \mathcal{C} denote the set of indices of the individuals that appear in any of the T samples and $\bar{\mathcal{C}}$ the indices of those individuals not sampled. Then, assuming independence among individuals and noting that n , \mathbf{c} , and \mathbf{y} are summaries of the t_{ij} , the joint probability function of the ordered sample times is, from Cox and Hinkley (1974, 15) or Ross (2010, theorem 5.2, exercise 81(a)),

$$\begin{aligned} f(\{t_{i1}, \dots, t_{ic_i}, c_i\}_{i \in \mathcal{C}}, n, |\boldsymbol{\lambda}, N) &= \prod_{i \in \bar{\mathcal{C}}} e^{-v_i} \times \prod_{i \in \mathcal{C}} \left\{ e^{-v_i} \prod_{j=1}^{c_i} \lambda_i(t_{ij}) \right\} \\ &= \prod_{i=1}^N \frac{e^{-v_i} v_i^{c_i}}{c_i!} \times \prod_{i \in \mathcal{C}} c_i! \frac{\prod_{j=1}^{c_i} \lambda_i(t_{ij})}{v_i^{c_i}}, \end{aligned} \quad (12.23)$$

where the second term is $f_2(\mathbf{t}^{\text{obs}}; \mathbf{c}, \boldsymbol{\lambda})$, $v_i = \int_0^\tau \lambda_i(u) du$, and $\boldsymbol{\lambda} = \{\lambda_i(\cdot)\}$.

The first factor of (12.23) represents the joint distribution of N exchangeable Poisson observations c_i with parameter v_i , while the second factor is the joint distribution of n exchangeable realizations of individual-specific vectors of sampling times conditional on \mathbf{c} . The distribution (12.23) is expressed in terms of the latent indexing of \mathcal{C} and $\bar{\mathcal{C}}$. To find an observed data likelihood in terms of \mathbf{t} , we need to account for the fact that indexing of the individuals in (12.23) is latent. Without loss of generality, the individuals can be indexed by the order in which they appear in the samples. Barker et al. (2014) use \mathbf{t}^{true} when conditioning on the latent indexing and \mathbf{t}^{obs} when conditioning on indexing by order of appearance in samples. For any observed history, the authors showed that the associated count of true histories that could generate it is $N!/(N-n)!$. This means that we can simply premultiply (12.23)

by this factor to get

$$f(\mathbf{c}, \mathbf{t}^{\text{obs}} | \boldsymbol{\lambda}, N) = f_1(\mathbf{c}; N, \mathbf{v}) \times f(\mathbf{t}^{\text{obs}}; \mathbf{c}, \boldsymbol{\lambda}, \mathbf{v}). \quad (12.24)$$

We see that the maximum likelihood estimator of N can be obtained just from $f_1(\mathbf{c}; N, \mathbf{v})$.

If the sample times are unobserved or unrecorded, the authors obtained the density function

$$f(\mathbf{y}^{\text{obs}} | \mathbf{v}, N) = \frac{N!}{(N-n)!} \times \frac{1}{T!} \times \prod_{i=1}^N e^{-v_i} v_i^{c_i}. \quad (12.25)$$

They then used the above theory with a few modifications to obtain expressions for all of Pollock's models discussed above.

We consider just the M_t model where it is assumed that $\lambda_i(t) = \lambda(t)$ and $v_i = v$ for all i . Under the special case M_0 we have $\lambda(t) = \lambda$ and $v_i = v$ so that for M_t and M_0 , (12.25) can be written as

$$\begin{aligned} f(\mathbf{y}^{\text{obs}} | v, N) &= \frac{e^{-Nv}(Nv)^T}{T!} \times \frac{N!}{(N-n)!} \left(\frac{1}{N}\right)^T \\ &= f(T; N, v) \times f(\mathbf{y}^{\text{obs}}; N, T). \end{aligned} \quad (12.26)$$

The term $f(\mathbf{y}^{\text{obs}}; N, T)$ in (12.26) is proportional to the CAPWIRE likelihood of Miller et al. (2005) and is a partial likelihood (Cox, 1975) obtained by conditioning on T . Further details, as well as the use of covariates, sampling in sessions, and computational methods involving Markov chain Monte Carlo implemented by Gibbs sampling (cf. Barker & Link, 2013) were considered by Barker et al. (2014).

Rehman et al. (2016) considered a model for one-at-a-time sampling, which was further discussed by Schofield (2017) with a rejoinder by Rehman et al. (2016).

12.15 Continuous-Time Models

When cameras traps are used to “trap” animals, the recordings (detections) arrive in a continuous stream but tend to be grouped into “occasions” for analysis, as with discrete trapping times where there are well-defined sampling occasions. This grouping leads to a loss of information and some subjectivity in defining “occasion.” The use of continuous data has been steadily evolving, and Liu et al. (2018) gave a snapshot of the earlier work done on the subject before year 2000 when martingales with estimating equations (Becker, 1984; Lin & Yip, 1999) and coverage methods were used (Yip & Chao, 1996). Initially only one animal was caught at each trapping occasion and the time of capture noted. Wilson and Anderson (1995) used

simulation to compare two previous estimators for model M_t using a general frailty model for a Poisson intensity function

$$\lambda(t) = b_i h_i \lambda_t,$$

where b_i and h_i are the respective behavior and the heterogeneity effects for individual i , and λ_t is a time effect. Wang and Yip (2002) introduced a combined multiplicative and additive frailty model

$$\lambda(t, \mathbf{z}) = \lambda_0(t) \exp[\gamma' \mathbf{x}(t)] + \boldsymbol{\beta}' \mathbf{w}(t)$$

with covariates \mathbf{x} and \mathbf{w} .

12.15.1 *Markov Model*

Wileyto et al. (1994) considered capture-recapture for insect populations and proposed a four-state discrete time Markov process. Their method provided an estimate from a single trap observation by allowing subjects to mark themselves. The estimate of the unknown population size was based on the assumption of a closed population and a simple Markov model in which the rates of marking, capture, and recapture were assumed to be equal. They pointed out that in the case of insect populations, mark-recapture models have not been very successfully used because they usually involve intense labor and they often provide few or no recaptures. Also, the initial stage involves mass rearing of insects for release or capture with arduous care to maintain subjects in good condition. When trapping is used to monitor insects in agricultural systems, release may be unacceptable to industry. The method required sequential visits to the traps in order to obtain an estimate.

Wileyto et al. (2000) used passive self-marking and concurrent trapping and developed a continuous-time dynamic model of the marking and trapping process-marking. Alpizar-Jara and Smith (2008) also considered a continuous-time model, extending Wileyto et al. (1994). As we shall see below, it is not necessary to group data.

12.15.2 *General Frailty Models*

We begin with a unified likelihood-based approach for estimating population size in continuous-time capture-recapture experiments with frailty developed by Xu et al. (2007). Given a study of duration τ , which was short enough so that the population of size N was effectively closed, let $N_i(t)$ denote the number of times individual i is captured in $[0, t]$, a continuous-time counting process with an intensity $\lambda_i(t)$.

A multiplicative form of the most general capture-recapture model, M_{thb} , has the frailty model

$$\lambda_i(t) = \beta \gamma_i \phi^{r_i(t)} \lambda_0(t),$$

where $r_i(t) = 1$ if individual i is observed/captured at least once by time t in $[0, \tau]$, and 0 otherwise; $\lambda_0(t)$ is an arbitrary non-negative time-varying function in $[0, \tau]$; the γ_i and ϕ are positive values representing the effects heterogeneity and behavioral response (after first capture), respectively; and β represents an overall capture intensity. It is assumed that in each case, the parameter reverts to unity in the case of homogeneity, e.g., $\phi = 1$, $\gamma_i = 1$ for all i (see the submodels below), and $\lambda_0(t)$ is assumed to have average value 1. For heterogeneity, it is further assumed that $\gamma_1, \dots, \gamma_N$ are sampled from a gamma distribution $\text{Gamma}(\alpha, \alpha)$, namely,

$$f(x) = \frac{x^{\alpha-1} \exp[-\alpha x] \alpha^\alpha}{\Gamma(\alpha)}$$

with mean 1 and variance $1/\alpha$. We then have the following submodels:

- (1) $M_0 : \lambda_i(t) = \beta$.
- (2) $M_t : \lambda_i(t) = \beta \lambda_0(t)$.
- (3) $M_b : \lambda_i(t) = \beta \phi^{r_i(t)}$.
- (4) $M_h : \lambda_i(t) = \beta \gamma_i$.
- (5) $M_{th} : \lambda_i(t) = \beta \gamma_i \lambda_0(t)$.
- (6) $M_{bh} : \lambda_i(t) = \beta \gamma_i \phi^{r_i(t)}$.
- (7) $M_{tb} : \lambda_i(t) = \beta \phi^{r_i(t)} \lambda_0(t)$.
- (8) $M_{thb} : \lambda_i(t) = \beta \gamma_i \phi^{r_i(t)} \lambda_0(t)$.

Given capture times t_{ij} for individual i and time j , the authors then gave likelihood equations for each of the above models and obtained maximum likelihood estimates iteratively. A variation on this model with covariates was given by Xu et al. (2007), who assumed no behavioral response and ignored a time effect. An earlier covariate model of Hwang and Chao (2002) had for model M_{thb}

$$\lambda_i(t) = \lambda_0(t) \exp[\boldsymbol{\beta}' \mathbf{z}_i + \alpha r_i(t)],$$

where $\alpha = \log(\phi)$ with $\phi = 1$ for no behavioral effect and the \mathbf{z}_i are covariates. Duan et al. (2009) considered a similar continuous model but with a focus on estimating the covariate distribution. They used, for the i th individual captured,

$$\lambda(t, \boldsymbol{\beta} | \mathbf{z}_i) = \lambda_0(t) \exp[\boldsymbol{\beta}' \mathbf{z}_i],$$

where \mathbf{z}_i is a vector of covariates.

12.15.3 Some Model Variations

Bell et al. (2003) developed a continuous capture-recapture model using unequal capture intervals for edible crabs. More than 3000 crabs were marked with paint and sampled using baited traps of standard commercial design, and fishing-effort data was incorporated in the model.

Farcomeni and Scacciatelli (2013) combined previous models to include not only covariates but also unobserved heterogeneity to estimate the number of cannabis users in Italy. Unlike the current models, they explicitly distinguished and allowed for two different sources of individual heterogeneity. They called the model M_{hob} , where h stands for the unobserved heterogeneity, o for the observed heterogeneity, t for the time-heterogeneity, and b for a behavioral response to capture, specified by

$$\lambda_i(t) = \rho_i \exp[\beta' \mathbf{z}_i] \phi^{r_i(t)} \omega(t).$$

Here ρ_i is a subject-specific frailty term corresponding to unobserved heterogeneity, β is a vector of log hazard-ratio coefficients describing the observed heterogeneity, and $\omega(t)$ is an unspecified nonnegative baseline function describing the time-heterogeneity. All the parameters, except for the β , are assumed to be strictly positive. A common choice for the distribution of ρ_i is $\text{Gamma}(\alpha, \alpha)$. As before, $N_i(t)$, $i = 1, \dots, N$, denotes the number of times the i th subject has been captured in the time interval $[0, \tau]$. The authors listed 16 submodels in a table. Given that one cannot measure covariates for subjects never captured, the only practical possibility for inference is via maximization of the conditional likelihood, that is, the likelihood obtained conditioning on the event that $N_i(\tau) > 0$ and then used a Horvitz-Thompson-type estimator for N . An EM algorithm was used to obtain maximum likelihood estimates. Full details are given in this extensive paper.

Liu et al. (2018) proposed an empirical likelihood (EL) approach for continuous capture-recapture assuming a nonhomogeneous Poisson process with a covariate-dependent intensity function $\lambda(t \mid \mathbf{z})$, where \mathbf{z} are time-independent covariates. Let $D(\mathbf{z}) = r_i(\tau) = 1$ if an object is captured at least once in time $[0, \tau]$, and 0 otherwise. Let

$$\pi(\mathbf{z}) = \Pr(D(\mathbf{z}) = 1) = 1 - \exp\left[\int_0^\tau \lambda(t \mid \mathbf{z}) dt\right],$$

and $\alpha = \Pr(D(\mathbf{z}) = 0)$. Let $f(N, \mathbf{z})$ be the distribution of (N, \mathbf{z}) ; $f_c(N, \mathbf{z})$ the conditional distribution, given $D(\mathbf{z}) = 1$; and n the random number of subjects captured throughout the study. Considering now the data we have $(n, N_1, \dots, N_n, z_1, \dots, z_n)$, where N_i is the number of times object i with covariates \mathbf{z}_i is captured, and the full likelihood L is composed of three parts. The first part is a binomial likelihood, the second part is a conditional parametric

likelihood or partial likelihood, and the third part is the marginal EL constructed from the covariate information, giving

$$L = f(n) f(\{\mathbf{z}_i\} \mid n) f(\{N_i\} \mid n, \{\mathbf{z}_i\}).$$

Let $A_i = \exp \left[- \int_0^\tau \lambda(t \mid \mathbf{z}_i) dt \right] / \pi(\mathbf{z}_i)$, then

$$f(n) = \binom{v}{n} (1 - \alpha)^n \alpha^{v-n},$$

where v is the population size,

$$f_c(\mathbf{z}_i, \dots, \mathbf{z}_n \mid n) = \prod_{i=1}^n f_c(\mathbf{z}_i) = \prod_{i=1}^n \pi(\mathbf{z}_i) f(\mathbf{z}_i) / (1 - \alpha),$$

and

$$f(N_1, \dots, N_n \mid n, \mathbf{z}_1, \dots, \mathbf{z}_n) = \prod_{i=1}^n A_i \exp \left[\int_0^\tau \log[\lambda(t \mid \mathbf{z}_i)] dN_i(t) \right].$$

Details are given by the authors. Models for missing covariates (including those missing at random) and those with measurement errors were given by Yip et al. (2005), Lee et al. (2016), and Liu et al. (2020), with the latter comparing their model with that of Liu et al. (2016).

A previous empirical likelihood approach for heterogeneity following the model of Alho (1990) and Huggins (1989) was given by Liu, Li et al. (2017b), who showed that the null distribution of the empirical likelihood ratio for the abundance was asymptotically chi-squared with one degree of freedom. Also, the maximum empirical likelihood estimator achieves semiparametric efficiency. Simulation studies showed that the empirical likelihood-based method is superior to the conditional likelihood-based method, as its confidence interval has much better coverage, and the maximum empirical likelihood estimator has a smaller mean-square error.

12.15.4 Models of Schofield et al.

Schofield et al. (2018) proposed a full range of capture-recapture models using the method and notation of Barker et al. (2014), described in the previous section. We repeat the notation and some equations for easy reference. Given a study of duration τ , let $t_{ij} \in (0, \tau)$ be the j th observed time of capture for the i th individual ($i = 1, \dots, N$, $j = 1, \dots, c_i$), and let T be the total number of observations during the study period. The vector $\mathbf{t}_i = (t_{i1}, \dots, t_{ic_i})'$ gives the ordered capture times for

individual i , with the collection $\mathbf{t}' = (\mathbf{t}'_1, \dots, \mathbf{t}'_N)$ arbitrarily ordered according to the number of observations for each individual. We have some summaries of \mathbf{t} .

n = Number of unique individuals observed during the study,

c_i = Number of observations for individual i ,

$\mathbf{c} = (c_1, \dots, c_n)'$,

y_h = Identity of the h th observation,

$\mathbf{y} = (y_1, \dots, y_T)'$,

s_h = Time of the h th observation,

$\mathbf{s} = (s_1, \dots, s_T)'$,

f_j = Number of individuals caught j times,

$\mathbf{f} = (f_1, \dots, f_T)'$.

We may not observe the capture times themselves but only have one or more of the summaries listed above. The models we consider below can be fitted using \mathbf{c} or \mathbf{y} and do not require the observed capture times \mathbf{t} . The capture times \mathbf{t}_i are modeled as realizations of a non-homogeneous Poisson process (NHPP) with intensity $\lambda_i(t)$, for $t \in [0, \tau]$). Assuming independence between individuals, a probability model for the observed times of captures is

$$f(\mathbf{t} | \boldsymbol{\lambda}, N) = \frac{N!}{(N-n)! \prod_{j=1}^T f_j!} \prod_{i=1}^N e^{-\nu_i} \prod_{i=1}^n \prod_{j=1}^{c_i} \lambda_i(t_{ij}), \quad (12.27)$$

where $\nu_i = \int_0^\tau \lambda_i(u)du$ can be thought of as a study-wide measure of catchability for individual i , and there is a combinatorial term due to the ordering of \mathbf{t} . We now factor the general model two ways to simplify model fitting.

Factorization I

The first factorization is given in Barker et al. (2014) and expressed the joint model for \mathbf{t} in terms of a model for the number of captures of each individual and a model for the times of capture conditional on number of captures, namely,

$$f(\mathbf{t} | \boldsymbol{\lambda}, N) = f(\mathbf{c} | \boldsymbol{\nu}, N) f(\mathbf{t} | \mathbf{c}, \boldsymbol{\lambda}), \quad (12.28)$$

where

$$f(\mathbf{c} | \boldsymbol{\nu}, N) = \frac{N!}{(N-n)! \prod_{j=1}^T f_j!} \prod_{i=1}^N \frac{e^{-\nu_i} \nu_i^{c_i}}{c_i!} \quad (12.29)$$

$$f(\mathbf{t} | \mathbf{c}, \boldsymbol{\lambda}) = \prod_{i=1}^n c_i! \left\{ \prod_{j=1}^{c_i} \frac{\lambda_i(t_{ij})}{v_i} \right\}. \quad (12.30)$$

Provided $\lambda(t)$ does not depend N , it is helpful to let $\delta_i(t) = \lambda_i(t)/v_i$ with the constraint $\int_0^\tau \delta_i(u)du = 1$ so that (12.30) becomes

$$f(\mathbf{t} | \mathbf{c}, \boldsymbol{\delta}) = \prod_{i=1}^n \left\{ c_i! \prod_{j=1}^{c_i} \delta_i(t_{ij}) \right\}. \quad (12.31)$$

We need just (12.29) for inference about N (and v). As the time of capture is not necessary for inference about N , no information is lost if times of capture are not recorded.

Factorization II

This factorization partitions the distribution of $\mathbf{t} = (\mathbf{y}, \mathbf{s})$ into terms for $\mathbf{y}|\mathbf{s}$ and \mathbf{s} , namely,

$$f(\mathbf{t} | \boldsymbol{\lambda}, N) = f(\mathbf{y}, \mathbf{s} | \boldsymbol{\lambda}, N) = f(\mathbf{y} | \mathbf{s}, \boldsymbol{\lambda}, N, T)(f(\mathbf{s} | \boldsymbol{\lambda}, N), \quad (12.32)$$

where the conditioning of T on \mathbf{y} is through \mathbf{s} . The model $f(\mathbf{y} | \mathbf{s}, \boldsymbol{\lambda}, N, T)$ describes the assignment of identity to each of the T observations. Under the assumptions of a Poisson process, the probability that observation j corresponds to individual i is $\lambda_i(s_j)/\sum_k \lambda_k(s_j)$, so that

$$f(\mathbf{y} | \mathbf{s}, \boldsymbol{\lambda}, N, T) = \frac{N!}{(N-n)! \prod_{j=1}^T f_j!} \prod_{j=1}^T \frac{\lambda_{y_j}(s_j)}{\sum_k \lambda_k(s_j)}, \quad (12.33)$$

where the combinatorial accounts for the arbitrary labeling of \mathbf{y} . If we assume a frailty model, such that $\lambda_i(t) = \gamma_i \lambda(t)$, this is further simplified to give

$$\begin{aligned} f(\mathbf{y} | \mathbf{s}, \boldsymbol{\lambda}, N, T) &= f(\mathbf{y} | \boldsymbol{\gamma}, N, T) \\ &= \frac{N!}{(N-n)! \prod_{j=1}^T f_j!} \prod_{i=1}^N \left(\frac{\gamma_i}{\sum_k \gamma_k} \right)^{c_i}, \end{aligned}$$

and \mathbf{y} and \mathbf{s} are independent. The model $f(\mathbf{s} | \boldsymbol{\lambda}, N)$ below is modeled as realizations of a single Poisson process where identities are not recorded and the intensity process is $\eta(t) = \sum_i \lambda_i(t)$, namely,

$$f(\mathbf{s} | \boldsymbol{\lambda}, N) = f(\mathbf{s} | \eta) = \prod_{j=1}^T \eta(s_j) \exp \left[- \int_0^\tau \eta(u) du \right]. \quad (12.34)$$

The above model provides no information about the abundance N due to S-ancillarity.

We now consider Pollock's models.

Models M_0 and M_t

Referring to (12.29), this model is overspecified as a likelihood for N ; there is no direct information about v_{n+1}, \dots, v_N , and further assumptions are necessary. Three such assumptions for $\lambda_i(t)$ are:

- (1) $\lambda_i(t) = \lambda$. We call this model M_0 as it is the continuous-time version of discrete model M_0 that has constant capture probabilities in each sampling occasion.
- (2) $\lambda_i(t) = \lambda(t)$. We call this model M_t as we allow the intensity to vary through time.
- (3) $\lambda_i(t)$ may differ from $\lambda_j(t)$, but $v_i = v_j$ for all i, j . All individuals can have different intensity functions provided all individuals have the same study-wide catchability v . This model is called M_t^* and includes the previous models.

Since v is constant, Eq. (12.29) is now

$$f(\mathbf{c} | v, N) = f(\mathbf{c} | N, T) f(T | N, v), \quad (12.35)$$

where $f(T | N, v)$ is Poisson(Nv) and

$$f(\mathbf{c} | N, T) = \frac{N!}{(N-n)! \prod_{j=1}^T f_j!} \frac{T!}{\prod_{i=1}^n c_i!} \left(\frac{1}{N}\right)^T. \quad (12.36)$$

This means that our inference regarding N does not depend on how catchability varies with time, since the same model is used irrespective of whether we assume a continuous-time model M_0 , M_t , or M_t^* (Hwang & Chao, 2002).

We can now either maximize the likelihood (12.36) for N (Hwang & Chao, 2002) or else take a Bayesian approach using Bayes formula to obtain the posterior distribution

$$f(N | \mathbf{c}, T) \propto \frac{N!}{(N-n)!} \left(\frac{1}{N}\right)^T f(N),$$

where $f(N)$ is the prior distribution for N . It is not necessary to use Markov chain Monte Carlo (MCMC) to evaluate this posterior as we can approximate it directly by assuming some upper value M for the abundance N . The resulting posterior, denoted $f_M(N | \mathbf{c}, T)$, can approximate the true posterior distribution to any degree of accuracy through the choice of M as it tends to $f(N | \mathbf{c}, T)$ as $M \rightarrow \infty$.

An alternative to the direct approach is to implement a Gibbs sampler for N and v based on (12.35). If it is assumed that $f(N) \propto N^{-1}$ and v is assumed to have a Gamma prior distribution with parameters α and β , then the full conditional

distributions are negative binomial and gamma for N and ν , respectively. For $f(N) \propto N^{-1}$, the authors showed in an Appendix that for large values of N

$$f(N | \mathbf{c}, T) \propto \left(\frac{1}{N}\right)^{T+1-n}.$$

The authors provided tests for heterogeneity and behavior for M_t .

Models M_h and M_{th}

Instead of assuming $v_i = \nu$ for all i , a less restrictive assumption is to assume that the v_i ($i = 1, \dots, N$) are independently distributed as $f(\theta_v)$, where f is assumed to have a gamma distribution with parameters $\rho^{-1}, \pi/(1-\pi)$, $\rho > 0, \pi \in (0, 1)$. By integrating over the v_i , we end up with $f(\mathbf{c} | \rho, \pi, N)$ being a product-negative binomial model for M_{th} . If $f(N) \propto N^{-1}$, then the full conditional distribution for N is negative binomial. To fit the model, independent priors were used for N, ρ , and π .

Models M_b and M_{tb}

The final set of assumptions we consider involve accounting for a behavior effect, namely, an effect of prior capture. Schofield et al. (2018), followed Hwang et al. (2002) and Xi et al. (2007) and (i) included behavior through a factor ϕ , $\phi > 0$ that is assumed to change the intensity after first capture and (ii) considered a frailty model, such that

$$\lambda_i(t) = \begin{cases} \gamma_i \lambda(t), & t \leq t_{i1}, \\ \phi \gamma_i \lambda(t), & t > t_{i1}. \end{cases} \quad (12.37)$$

Let $\mathbf{X} = (x_{ij})$ denote a $N \times (T + 1)$ matrix with $x_{ij} = 1$ if individual i has been seen prior to sample j , and $x_{ij} = 0$ otherwise. Let $\mathbf{z} = (z_1, \dots, z_T)'$ denote whether the individual caught in sample j has been caught before, with $z_j = x_{y_j j}$. Substituting (12.37) into (12.33) and (12.34), we get

$$f(\mathbf{t} | \boldsymbol{\gamma}, \lambda, \phi, N) = f(\mathbf{y} | \boldsymbol{\gamma}, \phi, N, T) f(\mathbf{s} | \boldsymbol{\gamma}, \lambda, \phi, N),$$

where

$$f(\mathbf{y} | \boldsymbol{\gamma}, \phi, N, T) = \frac{N!}{(N - n)! \prod_{j=1}^T f_j!} \prod_{j=1}^T \frac{\phi^{z_j} \gamma_{y_j}}{\delta_j}, \quad (12.38)$$

$$f(\mathbf{s} | \boldsymbol{\gamma}, \lambda, \phi, N) = \prod_{j=1}^T \delta_j \lambda(s_j) \prod_{j=1}^{T+1} \exp[-\delta_j \int_{s_{j-1}}^{s_j} \lambda(u) du]. \quad (12.39)$$

We have $\delta_j = \sum_{i=1}^N \phi^{x_{ij}} \gamma_i$, ($j = 1, \dots, T + 1$), with $s_0 = \mathbf{0}$ and $s_{T+1} = \tau$ (the total survey time). Here \mathbf{y} and \mathbf{s} are independent, and we cannot move between

$\mathbf{y}|N, T$ and $\mathbf{c}|N, T$ as we did for model M_t . Information about behavior arises from the order of observations (now given by \mathbf{y}) that is lost in the summary \mathbf{c} . As \mathbf{s} is S-ancillary for the parameters N, ϕ and $\boldsymbol{\gamma}$, it transpires that we can use (12.38) to make inferences about these parameters. However, to do so, we need to make further assumptions about $\boldsymbol{\gamma}$, and we choose $\gamma_i = 1$ for all i , giving us the continuous-time equivalent of model M_{tb} , namely,

$$f(\mathbf{y} | \phi, N, T) = \frac{N!}{(N-n)! \prod_{j=1}^T f_j!} \phi^R \prod_{j=1}^T \frac{1}{\xi_j}, \quad (12.40)$$

where R is the number of individuals caught at least twice, $\xi_j = N - m_j(1 - \phi)$, and $m_j = \sum_i x_{ij}$ being the number of individuals observed before sample j . Note that if $\phi = 1$ (no behavior effect), then the theory reduces to that for M_0 and M_t , as expected. The above discussion leaves out a lot of details, including the computational methods used, so that Schofield et al. (2018) and their supplementary materials need to be consulted.

Zhang and Bonner (2020) discussed the above modeling and argued (i) that the total number of captures per individual followed a Poisson distribution and not that the observations occur in continuous time and (ii) that there are continuous time processes for which the above results do not hold. They showed that the results hold even if the data are discretized, provided the number of captures per individual follows a Poisson distribution.

In responding, Schofield and Barker (2020) replied that their motivation for the study arose because the modeling of DNA recorded from badgers involved data that were generated in continuous time, but those times were not recorded (Wright et al., 2009). In their view, count data are rarely generated at discrete, instantaneous, time points. All familiar noninvasive sampling procedures involve continuous-time processes, for example, passive sensors (photo or acoustic) and DNA from hair and scat. In scat data, for example, the daily visits by the researchers to count scat are in discrete time, but the scat themselves are deposited by the animals in continuous time. The point of Barker et al. (2014) was to understand when these latent deposit times could be ignored.

Referring to (ii) above, Schofield and Barker (2020) agreed that there are counting processes for which we are not justified in modeling solely in terms of the counts and referred to the above behavioral models where data was modeled in terms of the ordered identification of individuals, the variable \mathbf{y} above, rather than the counts.

12.15.5 Random Removals

Yip and Wang (2002) considered a continuous capture-recapture process where some individuals are removed at random at known times t_h for subject h (e.g.,

accidental deaths). Let $N_h(t)$ be the number of times individual h has been captured by time t ($t \in [0, \tau]$), and let $y_h(t) = 1$ if individual h is at risk of being caught just before time t , and 0 otherwise. If $\mathbf{z}_h(t)$ is a vector of covariates, the authors built a model for the random and independently distributed triplets $(N_h(t), y_h(t), \mathbf{z}_h(t))$. It was assumed that the counting process had an intensity function $y_h(t)\lambda(t; \boldsymbol{\theta}, \mathbf{z}_h)$ so that $\lambda(t; \boldsymbol{\theta}, \mathbf{z}_h)$ is a parametric hazard function depending on unknown parameters $\boldsymbol{\theta}$ that represent the effects of the covariates. A conditional likelihood was used, and estimation details are given by the authors.

12.15.6 Self-Exciting (Self-Correcting) Processes

This section links up with previous ones involving a behavioral effect where the occurrence of an event at a time point either increases or decreases the probability of occurrence of a later event but in a much wider sense than previously described in terms of population changes. This means that the literature on the subject has developed differently. Self-exciting processes, also known as Hawkes processes (Hawkes, 1971), have been applied to environmental and seismological data, health and genomic studies, and crime, for example, while self-correcting (inhibition) processes have been applied mostly to earthquake data. Bonnet et al. (2021) combined both models by extending the Hawkes process. Following a similar line, but more appropriate to capture-recapture, Altieri et al. (2022) developed a model to assess the extent of illegal drug trafficking in Italy. Their contribution to the continuous-time capture-recapture literature is that of a “flexible and parsimonious behavioral effect, which has direct interpretation and applicability for both human and animal populations, embedded in a unified framework that can also take into account all other sources of heterogeneity.” We briefly outline the model.

Let $\{t_1, \dots, t_S\}$ be the sequence of times of a capture (or sighting) event for an individual, and assume a Poisson process for captures of individuals with intensity $\lambda(t)$. Let $N(t)$ be the number of events (captures) of an individual in the interval $[0, t]$, then $N(t) \sim \text{Poisson}(\Lambda(t))$, where $\Lambda(t) = \int_0^t \lambda(s)ds$. To allow for trap shyness or trap addiction, the authors modeled

$$\lambda(t) = \alpha \sum_{k:t_k < t} \exp[-\beta(t - t_k)] + \eta \exp[-\theta(N(t) - \eta t)],$$

with $\boldsymbol{\psi} = (\alpha, \beta, \eta, \theta)'$ are all positive ($\alpha < \beta$), θ is a parameter controlling the degree of trap-shyness and ηt a target for the number of events at time t . Referring now to the data, let $K_i = N_i(\tau)$ be the total captures observed for each individual i ($i = 1, \dots, n$), with individual subject-specific capture times $\{t_{i1}, \dots, t_{iK_i}\}$ over an

interval $[0, \tau]$. Assuming individual independence, the log-likelihood for individual i summed over i with $t_{i0} = 0$ and $t_{i,K_i+1} = T$ is

$$\ell(\lambda) = \sum_{i=1}^n \left[\sum_{k=1}^{K_i} \log(\lambda(t_{ik})) - \int_0^\tau \lambda_i(t) dt \right].$$

The authors then obtained a log-likelihood for the time-interaction process $\ell(\psi) = \sum_i \ell_i(\psi)$ where

$$\begin{aligned} \ell_i(\psi) &= \sum_{k=1}^{K_i} \log \left[\alpha \sum_{s: t_{is} < t_{ik}} \exp[-\beta(t_{ik} - t_{is})] + \eta \exp[-\theta(N_i(t_{ik}) - \eta t_{ik})] \right] \\ &\quad + \frac{\alpha}{\beta} \sum_{k=1}^{K_i} \{\exp[-\beta(T - t_{ik})] - 1\} \\ &\quad - \frac{1}{\theta} \sum_{k=0}^{K_i} \exp[-\theta k] \{\exp[\theta \eta t_{i,k+1}] - \exp[\theta \eta t_{ik}]\}. \end{aligned}$$

Finally, a conditional log-likelihood on positive captures is found to be

$$\ell^*(\mathbf{psi}) = \ell(\psi) - n \log \left[1 - \exp \left\{ -\frac{\exp[\theta \eta T] - 1}{\theta} \right\} \right].$$

This corresponds to model M_b , and a Horvitz-Thompson estimate of N is

$$\widehat{N} = \frac{n}{1 - \exp \left\{ -\frac{\exp[\widehat{\theta} \widehat{\eta} T] - 1}{\widehat{\theta}} \right\}},$$

where the denominator is an estimate of the probability that an individual is caught at least once. The authors then go on to modify the above $\lambda_i(t)$ to allow for heterogeneity as well as behavior and set up a general model M_{hotb} introduced in Sect. 12.15.3.

12.15.7 Single-Catch Spatial Model

Distiller and Borchers (2015) noted that the usual method of multi-catch traps, although robust with regard to density estimation, has a performance that deteriorates with high trap saturation and increasing density gradients. (Camera traps of course do not have this problem). They used a spatially explicit capture-recapture

model in which individual capture times were recorded. Although spatial models are considered in Chap. 14, the model is included here as it involves continuous time.

Let n unique individuals be caught over a survey of duration τ with an array of K traps. If release times are the same for all traps, then this leads to a natural definition of “occasion,” and the survey duration τ can be divided into L occasions. Animal density was modeled as a nonhomogeneous Poisson Process with hazard function (representing the mean capture rate per unit time) for the i th individual and the k th trap at time t , given by $h_k(t, \mathbf{x}_i; \boldsymbol{\theta})$. This can depend on both space (in terms of the distance from the trap to the activity center \mathbf{x}_i), time, and unknown vector $\boldsymbol{\theta}$ of hazard function parameters. For example, we have the commonly used time-invariant half-normal hazard function,

$$h_k(t, \mathbf{x}_i; \boldsymbol{\theta}) = \lambda_0 \exp[-d_k(x_i)^2/(2\sigma^2)],$$

where $\boldsymbol{\theta} = (\lambda_0, \sigma)$ and $d_k(x_i)$ is the distance of individual i from trap k . In the absence of other traps and other individuals, the “survivor function” for individual i at trap k over the whole survey (the probability of individual i not being caught in the trap by time τ) is

$$S_k(\tau, \mathbf{x}_i; \boldsymbol{\theta}) = \exp \left[- \int_0^\tau h_k(u, \mathbf{x}_i; \boldsymbol{\theta}) du \right].$$

The combined detection hazard over all traps at time t is

$$h.(t, \mathbf{x}_i; \boldsymbol{\theta}) = \sum_{k=1}^K h_k(t, \mathbf{x}_i; \boldsymbol{\theta}),$$

and the overall probability of capture in $[0, \tau]$ over all traps is

$$p.(\mathbf{x}_i; \boldsymbol{\theta}) = 1 - S.(\tau, \mathbf{x}_i; \boldsymbol{\theta}),$$

where

$$S.(\tau, \mathbf{x}_i : \boldsymbol{\theta}) = \exp \left[- \int_0^\tau h.(u, \mathbf{x}_i; \boldsymbol{\theta}) du \right]$$

is the overall survivor function. Let $D(\mathbf{x}; \boldsymbol{\phi})$ denote the density at a point \mathbf{x} in space, depending on the parameters $\boldsymbol{\phi}$, and the logarithm of this can be expressed as a linear regression in covariates, for example.

Considering now the data, let y_{ik} be the number of times the i th individual is caught at the k th trap, let $y_{ik} \leq L$ capture times be $\mathbf{t}_{ik} = (t_{ik1}, \dots, t_{iky_{ik}})'$ for individual i ($i = 1, \dots, n$), trap k ($k = 1, \dots, K$), and let the matrix $\mathbf{T} = \{\mathbf{t}_{ik}\}$ denote the set of all detection times. With single-catch traps, the survival function term needs to take account of traps having been taken out of action by catching

other individuals. Exposure to any particular trap falls to zero as soon as that trap catches any individual, and once an individual is caught in a particular trap, it cannot be caught in any other trap until it is released. To construct a likelihood with these features, we define indicator variables $a_k = 1$ if trap k is unoccupied at time t , and 0 otherwise, and $v_i(t) = 1$ if individual i is not in a trap at time t , and 0 otherwise. The hazard function for individual i for trap k at time t is then conveniently written as $v_i(t)a_k(t)h_k(t, \mathbf{x}_i; \boldsymbol{\theta})$, and the survivor function for individual i to time t is defined to be

$$S.(t, \mathbf{x}_i; \boldsymbol{\theta}) = \exp \left[- \int_0^t v_i(u) \sum_{k=1}^K a_k(u) h_k(u, \mathbf{x}_i; \boldsymbol{\theta}) du \right].$$

The likelihood for $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ is then

$$L(\boldsymbol{\theta}, \boldsymbol{\phi} | n, \mathbf{T}) = \frac{\exp[-\lambda(\boldsymbol{\theta}, \boldsymbol{\phi})]}{n!} \prod_{i=1}^n \left[\int_A D(\mathbf{x}_i; \boldsymbol{\phi}) S.(t, \mathbf{x}_i; \boldsymbol{\theta}) \prod_{k=1}^K \prod_{r=1}^{y_{ik}} h_k(t_{ikr}, \mathbf{x}_i; \boldsymbol{\theta}) d\mathbf{x}_i \right].$$

Here, $\lambda(\boldsymbol{\theta}, \boldsymbol{\phi}) = \int_A D(\mathbf{x}; \boldsymbol{\phi}) p.(\mathbf{x}; \boldsymbol{\theta}) d\mathbf{x}$, and the integral is over all possible activity center locations that could have led to a detection on the survey. The term $p.(\mathbf{x}; \boldsymbol{\theta})$ is the overall probability of being caught during the survey. Because of computational difficulties in computing this expression, it was replaced by

$$\tilde{\lambda}(\boldsymbol{\theta}, \boldsymbol{\phi}) = \int_A D(\mathbf{x}; \boldsymbol{\phi}) \exp \left[- \int_0^\tau \sum_{k=1}^K a_k(u) h_k(u, \mathbf{x}; \boldsymbol{\theta}) du \right] d\mathbf{x},$$

which depends on the observed $a_k(t)$. An estimator was obtained by maximizing the likelihood L using the above expression $\tilde{\lambda}(\boldsymbol{\theta}, \boldsymbol{\phi})$. The estimator bias and the coverage of a confidence interval estimator based on the observed data were evaluated by simulation. A different version of the above model was given by Borchers et al. (2014).

Dorazio and Karanth (2017) developed a hierarchical SCR model to estimate the spatial distribution and abundance of animals N detected with continuous-time recorders in the population region \mathcal{R} . We use their notation and assume there are K recorders with \mathbf{x}_k the location of the k th recorder, and the time interval $(0, T_k]$ is its continuous period of operation. Each camera is equipped with a motion sensor so that a photograph (detection) is only obtained during an encounter with one or more animals. The time and date of the encounter are normally recorded with each photograph. Let $\mathbf{t}_{ik} = (t_{ik1}, \dots, t_{iky_{ik}})'$ denote the observed sequence of detection times of the i th individual ($i = 1, \dots, n$) at camera k , where y_{ik} is the number of detections of individual i at camera k , and $t_{ikj} \in (0, T_k]$ if $y_{ik} > 0$, otherwise $\mathbf{t}_{ik} = \emptyset$. We then have the $n \times K$ matrix of camera-trap detections $\mathbf{Y} = (y_{ik})$ and

the block matrix of detection times $\mathbf{T} = (\mathbf{t}_{ik})$ ($i = 1, \dots, N$), a matrix of vectors; zero vectors for the unobserved $n_0 = N - n$ are attached to \mathbf{T} .

The spatial distribution of activity centers is assumed to be the realization of a Poisson process with intensity function $\lambda(\mathbf{s})$, the expected density of activity centers (number per unit area) at location $\mathbf{s} \in \mathcal{R}$. The total number of individuals is Poisson with mean $\Lambda(\mathcal{R}) (= \int_{\mathcal{R}} \lambda(\mathbf{s}) d\mathbf{s})$, and the expected density is modeled as $\log[\lambda(\mathbf{s})] = \boldsymbol{\beta}' \mathbf{v}(\mathbf{s})$ with covariates $\mathbf{v}(\mathbf{s})$. Then $N \sim \text{Poisson}(\Lambda(\mathcal{R}))$, and the joint density function of the \mathbf{s}_i ($i = 1, \dots, N$) can be shown to be

$$f(\{\mathbf{s}_i\} | N) = N! \prod_{i=1}^N \lambda(\mathbf{s}_i) / \Lambda(\mathcal{R}).$$

The goal now is to estimate the \mathbf{s}_i and N using the observed data. It is assumed that the sequence of detection times for the i th individual at camera k is also a Poisson process with intensity function based on the distance of an individual from a camera, namely,

$$\phi(t, \mathbf{s}_i, \mathbf{x}_k) = \psi_k \lambda(t) \exp[-\|\mathbf{s}_i - \mathbf{x}_k\|^2 / (2\sigma^2)],$$

for $t \in (0, T_k]$. To specify the baseline rate $\psi_k (> 0)$ it is assumed that $\log(\psi_k) = \boldsymbol{\alpha}' \mathbf{w}_k$, where \mathbf{w}_k are covariates for camera k , and we add $\log(\lambda(t)) = \boldsymbol{\delta}' \mathbf{z}(t)$ for temporal covariates $\mathbf{z}(t)$, provided that the values of the covariates are available for the entire period of each trap's operation.

Now the number of detections $y_{ik}(T_k)$ of individual i at camera k during time interval $(0, T_k]$ has a Poisson distribution

$$y_{ik}(T_k) | \mathbf{s}_i \sim \text{Poisson}(\Phi(T_k, \mathbf{s}_i, \mathbf{x}_k)),$$

where $\Phi(T_k, \mathbf{s}_i, \mathbf{x}_k) = \int_0^{T_k} \phi(t, \mathbf{s}_i, \mathbf{x}_k) dt$ is the expected number of detections of individual i at camera k during its period of operation. Then

$$f(\mathbf{t}_{ik} | y_{ik}, \mathbf{s}_i) = \prod_{j=1}^{y_{ik}} \phi(t_{ikj}, \mathbf{s}_i, \mathbf{x}_i) / \Phi(T_k, \mathbf{s}_i, \mathbf{x}_k).$$

which, allowing for the $\binom{N}{n}$ ways of observing n individuals, leads to the likelihood function for $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\alpha}', \sigma, \boldsymbol{\delta}')'$, namely,

$$f(\mathbf{Y}, \mathbf{T}, n, n_0, \{\mathbf{s}_i\} | \boldsymbol{\theta}) = \frac{\exp[-\Lambda(\mathcal{R})][\pi_0 \Lambda(\mathcal{R})]^{n_0}}{n_0!} \prod_{i=1}^n \lambda(\mathbf{s}_i) \prod_{k=1}^K f(y_{ik}, \mathbf{t}_{ik} | \mathbf{s}_i).$$

Here

$$f(y_{ik}, \mathbf{t}_{ik} | \mathbf{s}_i) = \frac{1}{y_{ik}!} \exp[-\Phi(T_k, \mathbf{s}_i, \mathbf{x}_k)] \prod_{j=1}^{y_{ik}} \phi(t_{ijk}, \mathbf{s}_i, \mathbf{x}_k),$$

and the probability of nondetection is

$$\pi_0 = \frac{1}{\Lambda(\mathcal{R})} \int_{\mathcal{R}} \lambda(\mathbf{s}) \prod_{k=1}^K \exp[-\Phi(T_k, \mathbf{s}, \mathbf{x}_k)] d\mathbf{s}.$$

A Bayesian approach is then applied using the unnormalized posterior distribution

$$f(\boldsymbol{\theta}, n_0, \{\mathbf{s}_i\} | \mathbf{Y}, \mathbf{T}, n) \propto f(\boldsymbol{\theta}) f(\mathbf{Y}, \mathbf{T}, n, n_0, \{\mathbf{s}_i\} | \boldsymbol{\theta}).$$

The case $\lambda(t) = \lambda$ was also considered. As well as using simulation, the method was applied to camera-trap detections of tigers living in and around the Nagarhole Tiger Reserve in India.

12.16 Conclusion

Capture-recapture methods for both open and closed populations have developed extensively in recent years, especially with the development of sophisticated computer programs and packages. There are now many different methods to estimate the abundance of closed populations. These include standard maximum likelihood methods, jackknife methods, coverage models, martingale estimating-equation models, log-linear models, logistic models, non-parametric models, and mixture models, which are all discussed in some detail. Because of the large amount of material, Bayesian methods are considered in the next chapter for convenience, as those methods are being used more. Covariates such as environmental variables are being used more, and with improved monitoring devices, including DNA methods, we can expect covariate methods to increase.

The two-sample capture-recapture model has been extensively used with a focus on variable catchability, use of two observers, which can also help with detectability problems, epidemiological populations using two lists (or later more lists), and dual record systems. For three or more capture-recapture samples, the glue behind the model development has been the setting out of eight particular model categories, due to Pollock, providing for a time factor, a behavioral factor, a heterogeneity factor, and combinations of these. A number of variations of these have also been developed by various researchers, including time-to-detection models. Heterogeneity has been the biggest challenge and, as well as various models, has also been considered using covariates or even stratification. Where possible underlying assumptions are tested. Finally, sampling one at a time and continuous models are considered in detail.

With this plethora of methods, the practitioner is left in a quandary. What methods are appropriate for what conditions and types of studies? What is needed here is a comparison of the various closed models with respect to both efficiency and robustness. Also, further research is needed on interval estimation, with intervals based on profile likelihoods becoming more popular and on model diagnostics.

Chapter 13

Capture-Recapture: Bayesian Methods



Matthew R. Schofield

13.1 Overview

The development of closed population capture-recapture models outlined in Chap. 12 focuses on classical estimation approaches. Alongside these are contributions that consider these models but make inference using a Bayesian approach. Many of these papers are not as well-known as their classical counterparts, particularly those published in the era before Markov chain Monte Carlo (MCMC). These early contributions typically focused on theoretical understanding and the mathematical manipulations and approximations required to find the posterior distribution. However, interest in Bayesian methods has increased with the availability of MCMC approaches, which has led to the rapid development of a wide class of models in a relatively short period of time. In this chapter, we focus on developments related to the Otis range of close population models (Otis et al. 1978). The use of Bayesian methods for extended models, e.g., spatial capture-recapture, is considered in Chap. 14 and other chapters.

13.2 Bayesian Statistics

The Bayesian ideal is that we use probability to describe uncertainty. When applied to estimation about an unknown parameter θ , uncertainty about θ is encoded through a probability mass function if θ is discrete or a probability density function if θ is continuous.

The prior distribution, denoted $f(\theta)$, describes our knowledge about θ before the experiment. Data y are collected and provide additional information about θ .

as described through a probabilistic model for the data $f(y|\theta)$. Bayes rule is used to update our prior knowledge, with the information obtained from the data. The resulting posterior distribution $f(\theta|y)$ contains our knowledge about θ given the observations y .

Bayes rule is

$$f(\theta|y) = \frac{f(y|\theta)f(\theta)}{f(y)} = \frac{f(y|\theta)f(\theta)}{\int_{\Theta} f(y|\psi)f(\psi)d\psi}, \quad (13.1)$$

where Θ is the support of θ . While this is mathematically simple to define, it can be challenging to implement. The integral in the denominator of (13.1) is difficult to evaluate for all but the simplest problems. This limited the application of Bayesian statistics throughout much of the twentieth century. Advances in computing power and algorithmic development since the late twentieth century have led to increased use of Bayesian approaches in applied statistics (cf. Seber & Schofield (2019, chapter 15 and Appendix) for more details on Bayesian computation).

The posterior distribution $f(\theta|y)$ contains all information about the parameters θ , conditional on the data y . Often we want to summarize our knowledge by a point or interval estimate of θ . This allows for comparison between Bayesian and classical estimators.

A principled approach to summarizing the posterior distribution is based on minimizing the loss (or equivalently maximizing the utility) we experience when comparing the estimate to the true value (see Berger 2010 for more details). Common loss functions include quadratic loss and absolute linear loss. The former quantifies the loss in terms of the squared Euclidean distance between the estimate and the true value. A particular posterior summary is obtained by minimizing any given loss function. For example, the posterior mean minimizes quadratic loss, while the posterior median optimizes absolute linear loss.

Interval estimates can be found using familiar notions of probability. The area under the posterior density gives a probability interval that summarizes our uncertainty about θ . Such intervals are referred to as credible intervals. A $100(1 - \alpha)\%$ credible interval is defined as an interval (θ_l, θ_u) where

$$\Pr(\theta_l < \theta < \theta_u) = 1 - \alpha.$$

There can be infinitely many such intervals. The central credible interval finds upper and lower endpoints θ_l, θ_u such that

$$\Pr(\theta > \theta_u) = \alpha/2$$

$$\Pr(\theta < \theta_l) = \alpha/2.$$

The idea of an interval estimate can be generalized to a region, which may consist of several intervals. For example, we may consider the highest posterior density (HPD) region \mathcal{P} , where

$$\Pr(\theta \in \mathcal{P}) = 1 - \alpha,$$

and

$$f(\theta_1|y) > f(\theta_2|y), \quad \forall \theta_1, \theta_2 : \theta_1 \in \mathcal{P}, \theta_2 \in \bar{\mathcal{P}}.$$

In other words, the HPD region is chosen so that any value of θ inside the region has higher posterior density than any value of θ outside the region.

13.3 Two-Sample Capture-Recapture

Suppose a random sample of M animals is captured without replacement, tagged, and released back into the population. Subsequently, a random sample of n animals is caught without replacement, of which m were tagged. Given M , n , and N , the random variable m has a hypergeometric distribution

$$f(m|n, M, N) = \frac{\binom{M}{m} \binom{N-M}{n-m}}{\binom{N}{n}}. \quad (13.2)$$

The maximum likelihood estimator is the integer part of the Lincoln-Petersen estimator given in (4.1). The model in (13.2) is model B of Darroch (1958) when there are two samples.

13.3.1 Roberts (1967)

The hypergeometric model was used as an example by Roberts (1967) while discussing stopping rules and Bayesian inference. Stopping rules are criteria that determine when data collection will cease. A stopping rule is informative if it provides information about the unknown parameters. The author assumes that sample size n is not fixed but has a binomial distribution with index N and capture probability p . This is an informative stopping rule, and the model for the data becomes

$$f(m, n|M, N, p) = \frac{\binom{M}{m} \binom{N-M}{n-m}}{\binom{N}{n}} \binom{N}{n} p^n (1-p)^{N-n}. \quad (13.3)$$

The maximum likelihood estimator of (13.3) is no longer the Lincoln-Petersen estimator. This is seen in a simple example where $M = 20$, $n = 20$, and $m = 10$. The Lincoln-Petersen estimate is $\hat{N} = 40$. The maximum likelihood estimator of N from (13.3) is 39. The model in (13.3) is similar to model A of Darroch (1958)

but differs in that Roberts (1967) included a model for the initial sampling of the M tagged animals.

The number of untagged individuals is $S = N - M$ before the second sample is conducted and is $S^* = S - n + m$ after the second sample. Roberts (1967) showed that the posterior distribution for S^* is beta-negative-binomial, provided a beta prior distribution is used for p and a discrete uniform prior is used for S . An interesting side note that is later highlighted by Castledine (1981), Garthwaite et al. (1995), and Schofield and Barker (2016) is that the posterior for N is unchanged irrespective of whether we adopt (13.2) or (13.3), provided that the prior for p is $f(p) = p^{-1}$. This is an example of what Severini (1995) later referred to as Bayes ancillarity. The prior distribution chosen for p determines whether the stopping rule is informative. We discuss this further in Sect. 13.11.2.

In specifying a discrete uniform prior for S , Roberts (1967) stated that they “explore a situation in which we know little” about S . In contrast, they assume that we may know a good deal about p :

“By knowing a good deal about p , we mean that for a given S , the marginal prior distribution of p is a beta density with parameters $[\alpha]$ and $[\gamma]$, both substantially greater than zero and substantially different from each other.”

We note that Roberts (1967) used a different parameterization of the beta distribution than that commonly used today. It was defined in terms of parameters α and γ , where

$$f(p|\alpha, \gamma) = \frac{\Gamma(\gamma)}{\Gamma(\alpha)\Gamma(\gamma-\alpha)} p^{\alpha-1} (1-p)^{\gamma-\alpha-1}, \quad 0 < p < 1, \alpha > 0, \gamma > \alpha.$$

The parameterization with shape parameters α and β is obtained by setting $\beta = \gamma - \alpha$.

Roberts (1967) considered an example where $M = 41$, $n = 32$, and $m = 8$ and used the posterior mean as a point estimate for abundance, minimizing quadratic loss. Three estimates were compared:

1. $\hat{N} = 240$: the posterior mean when using the prior $\alpha = 2, \gamma = 25$.
2. $\hat{N} = 202.5$: the posterior mean when using the prior¹ $\alpha = 0, \gamma = 0$.
3. $\hat{N} = 164$: the MLE of the hypergeometric model in (13.2).

While Roberts (1967) commented on the difference between the two Bayesian estimates with regard to the effect of the prior density chosen, nothing was said about the difference between the MLE and Bayesian estimates.

The difference between estimates 2 and 3 is largely attributable to the estimator used to summarize the posterior distribution. For the data of Roberts (1967), the posterior mass function when $\alpha = \gamma = 0$ (the prior associated with the estimate $\hat{N} = 202.5$) is given in Fig. 13.1. The posterior mass function is skewed and highly

¹ Technically the beta distribution with $\alpha = 0$ and $\gamma = 0$ is undefined. What is meant is that the prior is $f(p) \propto p^{-1}(1-p)^{-1}$, the limit of a beta distribution as $\alpha \rightarrow 0, \gamma \rightarrow 0$.

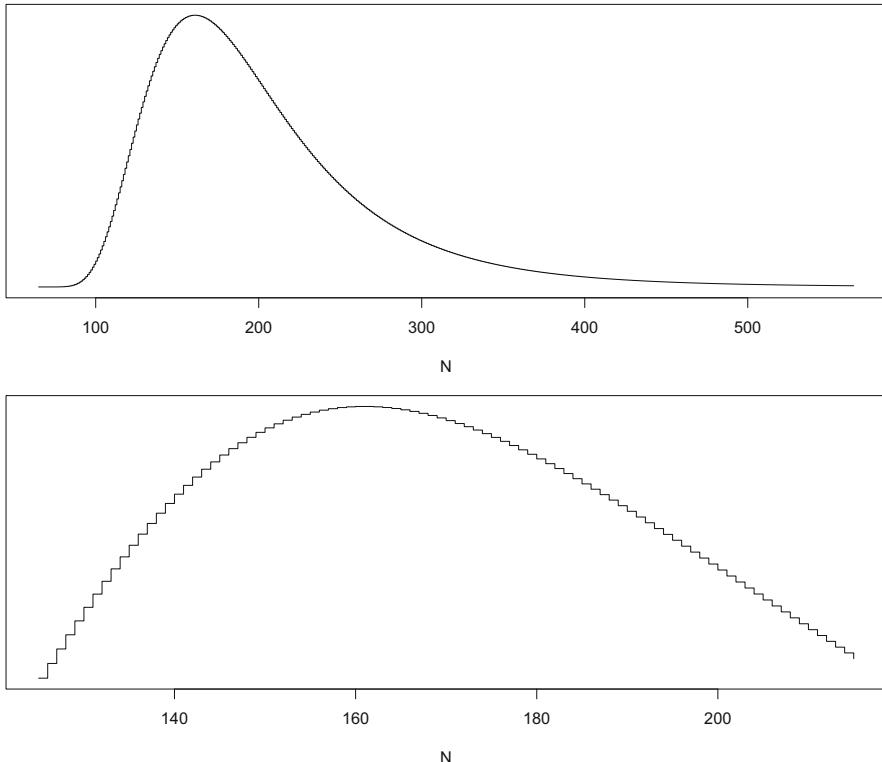


Fig. 13.1 The posterior distribution for N when fitting the model of Roberts (1967) and using a beta prior with parameters $\alpha = \gamma = 0$. The lower plot is a zoomed-in version of the upper plot to better show the location of the mode. The mode of the distribution is 161, while its mean is 202.5

variable, and Roberts (1967) did not consider any measure of variation. The mode of the posterior distribution is 161, similar to the MLE.

13.3.2 Other Approaches

Freeman (1973) also considered a two-sample capture-recapture experiment to determine an optimal stopping rule for the second sample. Among other results, the author showed that if the prior distribution for S is beta-negative-binomial, then the posterior distribution for S^* is also a beta-negative-binomial distribution. Closed form expressions are available for the posterior mean and variance of S^* , and equivalently for N . In determining a point estimate, Freeman (1973) assumed quadratic loss for N , $L(N, \hat{N}) = (N - \hat{N})^2$ and commented that other loss functions such as $(N - \hat{N})^2/N$ or $(N - \hat{N})^2/N^2$ were preferable, but not analytically

tractable. These other loss functions were used by Smith, P. (1988 and 1991) and Madigan and York (1997). For example, the latter used the loss function $L(N, \hat{N}) = (N - \hat{N})^2/N^2$, which corresponds to a point estimate

$$\hat{N} = \frac{E[N^{-1}|\text{data}]}{E[N^{-2}|\text{data}]},$$

where $E[N^{-x}|\text{data}]$ is the posterior expectation of N^{-x} ,

$$E[N^{-x}|\text{data}] = \int N^{-x} f(N|\text{data}) dN.$$

Gaskell and George (1972) also looked at the relationship between frequentist and Bayesian approaches, although they used a binomial model (rather than hypergeometric) to derive the Bayesian estimator. This assumes that sampling is with replacement; in most situations this is a reasonable approximation only if the population is large.

An observation we have not seen elsewhere is that if we start with the hypergeometric model in (13.2), the mode of the posterior distribution is available in closed form for certain prior distributions. The posterior mode is equivalent to the Lincoln-Petersen estimator (12.1) if an improper uniform prior is used for N . Alternatively, the posterior mode is identical to the Chapman estimator if the prior $f(N) = (N + 1)^{-1}$ is assumed (Simon Bonner, pers. comm.).

13.4 Model M_t

Model M_t is a multinomial model for data caught across s closed sessions. In particular, the model for a_ω , the count of individuals with capture history ω , can be written as

$$f(\mathbf{a}_\Omega | \mathbf{p}_s, N) = \underbrace{\frac{N!}{\prod_\omega a_\omega!(N - r)!} \prod_{i=1}^s \binom{N}{n_i}^{-1}}_{f(\mathbf{a}_\Omega | \mathbf{n}, N)} \underbrace{\binom{N}{n_i} p_i^{n_i} (1 - p_i)^{N - n_i}}_{f(\mathbf{n} | \mathbf{p}_s, N)}, \quad (13.4)$$

where n_i is the number of individuals caught in sample i , r is the total number of unique individuals caught throughout the study, and p_i is the capture probability in sample i . We use \mathbf{a}_Ω to represent the collection of capture histories $\{a_\omega\}$, where the subscript Ω defines the set of possible capture histories. We use $\mathbf{p}_s = (p_1, \dots, p_s)$ to represent the collection of capture histories. The subscript denotes that the collection has s elements. Finally, we use $\mathbf{n} = (n_1, \dots, n_s)$ to represent the collection of sample sizes.

The first term in (13.4) is Darroch's (1958) model B, while his full model is model A. Castledine (1981) specified the posterior distribution for N when p_i have independent beta prior distributions with parameters α and β , namely,

$$f(\mathbf{p}_s) = \prod_{i=1}^s \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} p_i^{\alpha-1} (1 - p_i)^{\beta-1}.$$

It is possible to integrate over \mathbf{p}_s in closed form, leading to the integrated likelihood

$$f(\mathbf{a}_{\Omega}|N) = \frac{N!}{(N-r)!} \prod_{i=1}^s \frac{\Gamma(N - n_i + \beta)}{\Gamma(N + \alpha + \beta)}, \quad (13.5)$$

and the unnormalized posterior for N is

$$f(N|\mathbf{a}_{\Omega}) \propto \frac{N!}{(N-r)!} \prod_{i=1}^s \frac{\Gamma(N - n_i + \beta)}{\Gamma(N + \alpha + \beta)} f(N), \quad (13.6)$$

with $f(N)$ denoting the prior distribution of N . Castledine (1981) identified that use of the prior $f(p_i) \propto p_i^{-1}$ results in the same posterior distribution for N irrespective of whether Darroch's (1958) model A or B is used as the data model. This is the extended version of that discussed for the two-sample model in Sect. 13.3.1. The equivalence of the posterior distributions does not necessarily translate to classical approaches. The maximum likelihood estimate for N differs between Darroch's models A and B (Schofield & Barker 2016).

Castledine (1981) used quadrature to evaluate posterior summaries for $\log(N)$. A normal approximation was required to find credible intervals. Since N is one-dimensional and discrete, a grid search could also be used between r and a maximum value N_{max} chosen so that $\Pr(N > N_{max}|\mathbf{a}_{\Omega})$ is adequately small. Smith (1991) noted that the posterior values can be efficiently calculated using the recursion

$$\frac{f(N+1|\mathbf{a}_{\Omega})}{f(N|\mathbf{a}_{\Omega})} = \frac{N+1}{N-r+1} \prod_{i=1}^s \left\{ \frac{N - n_i + \beta}{N + \alpha + \beta} \right\} \frac{f(N+1)}{f(N)}.$$

13.4.1 Poisson Model

Jewell (1983) applied closed population models to the problem of error detection, e.g., assessing defects in a production lot, errors in a manuscript, or bugs in computer code. There are s reviewers who independently assess and report the errors they have found. Jewell considered model M_t, where N , the total number of errors, is a realized random variable from a Poisson distribution with parameter λ and assumed

a gamma prior distribution for λ and independent beta prior distributions for p_i , $i = 1, \dots, s$. The model for the data becomes

$$f(\mathbf{a}_\Omega | \mathbf{p}_s, \lambda) = k \lambda^r e^{-\lambda(1 - \prod_{i=1}^s (1 - p_i))} \prod_{i=1}^s p_i^{n_i} (1 - p_i)^{r - n_i}, \quad (13.7)$$

where k is a normalizing term that depends only on the data, i.e., it does not include parameters λ or \mathbf{p}_s .

The term $e^{-\lambda(1 - \prod_{i=1}^s (1 - p_i))}$ in (13.7) is problematic. The dependence induced in the posterior densities of p_1, \dots, p_s makes it difficult to find the conditional distribution for p_i , $i = 1, \dots, s$. Despite this, the posterior predictive distribution for $U = N - r$ is available in closed form. The variable U is the number of errors that have not been discovered and is the quantity of interest. The posterior predictive distribution is

$$f(U | \mathbf{a}_\Omega) = \int f(U | \lambda, \mathbf{p}_s) f(\lambda, \mathbf{p}_s | \mathbf{a}_\Omega) d\lambda d\mathbf{p}_s.$$

The two terms in the integrand are

$$\begin{aligned} f(U | \lambda, \mathbf{p}_s) &= \frac{e^{-\lambda \prod_{i=1}^s (1 - p_i)} (\lambda \prod_{i=1}^s (1 - p_i))^U}{U!}, \\ f(\lambda, \mathbf{p}_s | \mathbf{a}_\Omega) &\propto \lambda^r e^{-\lambda(1 - \prod_{i=1}^s (1 - p_i))} \prod_{i=1}^s p_i^{n_i} (1 - p_i)^{r - n_i} f(\mathbf{p}_s) f(\lambda). \end{aligned}$$

The exponential terms involving \mathbf{p}_s cancel when multiplying $f(U | \lambda, \mathbf{p}_s)$ and $f(\lambda, \mathbf{p}_s | \mathbf{a}_\Omega)$. Moreover, if we assume independent beta priors for p_i with parameters α_i and β_i , $i = 1, \dots, s$, and a gamma prior for λ with parameters a and b , then the posterior distribution for U is

$$f(U | \mathbf{a}_\Omega) \propto \frac{\Gamma(U + r + a)}{U!} \left(\frac{1}{1 + b} \right)^U \prod_{i=1}^s \frac{\Gamma(U + r - n_i + \beta_i)}{\Gamma(U + r + \alpha_i + \beta_i)}. \quad (13.8)$$

The posterior can be calculated efficiently via the recursion

$$\frac{f(U + 1 | \mathbf{a}_\Omega)}{f(U | \mathbf{a}_\Omega)} = \frac{U + r + a}{U(1 + b)} \prod_{i=1}^s \frac{U + r - n_i + \beta_i}{U + r + \alpha_i + \beta_i}.$$

Jewell (1983) noted that the tails of the posterior distribution in (13.8) are negative-binomial, since

$$\frac{\Gamma(U + r - n_i + \beta_i)}{\Gamma(U + r + \alpha_i + \beta_i)} \rightarrow 1 \text{ as } U \rightarrow \infty.$$

13.4.2 Sequential Updating of Posterior for N

The capture occasions in a typical capture-recapture study occur sequentially in time. Gazey and Staley (1986) and Zucchini and Channing (1986) leveraged this in finding the posterior distribution for N using Bayesian updating (see Appendix A.9.1 for details on Bayesian updating). The idea is that the posterior distribution for N is updated after each successive sampling occasion. Gazey and Staley (1986) considered three sampling models: hypergeometric, binomial, and negative-binomial. The sequential approach is particularly useful if estimation is required in real time while data are being collected. It can also be useful when planning fieldwork; we can use simulated data and assess how knowledge about N changes with sampling duration. However, this approach has a computational cost. Each time N is updated requires evaluation of the posterior mass function. In most cases, it is preferable to find the posterior distribution once when all data are available.

13.5 Bayesian Development of Species Distribution Models

In parallel with the early Bayesian development of capture-recapture models were Bayesian treatments for species distribution models. There are close connections between these two fields of study. As stated by Hill (1979), the general problem is trying to infer the “number of types in a population, given as data a sample from that population.” More concretely, we can think of a population of N units containing M different species, with M the quantity of interest. The observed data consist of the values n_1, \dots, n_m , where m is the total number of species observed, n_j is the number of species j observed, and $n = \sum_j n_j$ is the total sample size.

Hill (1968) evaluated the posterior model of M when sampling is without replacement from a finite population. A scale prior $f(M) \propto M^{-1}$ was used, truncated above by N . Hill (1979) started with the model of Hill (1968) and let $N \rightarrow \infty$ to consider the case when either the population is infinite, or sampling is with replacement. A negative-binomial prior was used for M . After much algebra, various moments of the posterior distribution of M were obtained, including the mean and variance.

Lewins and Joanes (1984) motivated their model under the assumption of an infinite population. They used a multinomial model for the observations in terms of “species relative abundances” given by the probabilities p_1, \dots, p_M . They modelled these as a Dirichlet distribution with homogeneous parameter k , using a negative-binomial prior for M . When $k = 1$, the posterior distribution they obtained is the same as that of Hill (1979). However, they found posterior dependence on the choice of k . They suggested that a prior for k could be included, albeit with increased algebraic complexity. A possible compromise suggested by Good (1967) is to use an empirical Bayes approach, using the data to estimate k . A similar multinomial model was used by Boender and Rinnooy Kan (1987).

13.6 Model Fitting with Markov Chain Monte Carlo

The first Markov chain Monte Carlo (MCMC) algorithm widely introduced to the statistical community was the Gibbs sampler (Gelfand & Smith 1990). The availability of MCMC algorithms led to an increased interest in Bayesian methods across all areas of statistical research. The capture-recapture community was no exception.

The Gibbs sampler generates values from Markov chain, whose stationary distribution is the posterior distribution. The algorithm is straightforward. Suppose that we wish to sample from the density $f(x, y)$. The Gibbs sampler proceeds by alternatively sampling from the conditional distributions $f(x|y)$ and $f(y|x)$. It can be useful in applications where the full conditional posterior distributions are known, but the joint or marginal posterior is not available. See Seber and Schofield (2019) for additional details.

George (1992) used the Gibbs sampler to generate samples from the posterior distribution for model M_t . The full conditional distributions for both N and p_i are

$$\begin{aligned} f(N|\mathbf{a}_\Omega, \mathbf{p}_s) &\propto \frac{N!}{(N-r)!} \left\{ \prod_{i=1}^s (1-p_i) \right\}^N f(N) \\ f(\mathbf{p}_s|\mathbf{a}_\Omega, N) &\propto \left\{ \prod_{i=1}^s p_i^{n_i} (1-p_i)^{N-n_i} \right\} f(\mathbf{p}_s). \end{aligned} \quad (13.9)$$

The full conditional distribution of N is of known form for several choices of prior $f(N)$. If $f(N)$ is Poisson with parameter λ , then the full conditional distribution for $U = N - r$ is also Poisson, namely,

$$f(U|\mathbf{a}_\Omega, \mathbf{p}_s) = \frac{\Lambda^U e^{-\Lambda}}{U!},$$

where $\Lambda = \lambda \prod_{i=1}^s (1-p_i)$. If $f(N)$ is negative-binomial with parameters δ and π ,

$$f(N) = \frac{(N+\delta-1)!}{N! \Gamma(\delta)} \pi^\delta (1-\pi)^N,$$

then the full conditional distribution for $U = N - r$ is also negative-binomial with parameters $\Delta = r + \delta$ and $\Pi = 1 - (1-\pi) \prod_{i=1}^s (1-p_i)$, namely

$$f(U|\mathbf{a}_\Omega, \mathbf{p}_s) = \frac{(U+\Delta-1)!}{U! \Gamma(\Delta)} \Pi^\Delta (1-\Pi)^U.$$

A special case of the latter is the scale prior $f(N) = N^{-1}$ where the full conditional distribution for U is negative-binomial with parameters r and $1 - \prod_{i=1}^s (1 - p_i)$, namely,

$$f(U|\mathbf{a}_\Omega, \mathbf{p}_s) = \frac{(U+r-1)!}{U! (r-1)!} \left(1 - \prod_{i=1}^s (1 - p_i)\right)^r \left\{ \prod_{i=1}^s (1 - p_i) \right\}^U. \quad (13.10)$$

The full conditional distribution for N is somewhat unusual in our experience for two reasons. The first is that the prior is specified for one random variable (N) and conjugacy is expressed in terms of another random variable $U = N - r$. This is a form of censoring, in particular, after collecting the data the likelihood includes the information that $N \geq r$. The second is that there are two families of prior distribution that are conjugate for N .

When the priors for p_i ($i = 1, \dots, s$) are independent beta distributions with parameters α_i and β_i , then the full conditional distribution for \mathbf{p}_s is a product of independent beta distributions with parameters $a_i = \alpha_i + n_i$ and $b_i = \beta_i + N - n_i$, namely,

$$f(\mathbf{p}_s | \mathbf{a}_\Omega, N) = \prod_{i=1}^s \frac{\Gamma(a_i + b_i)}{\Gamma(a_i)\Gamma(b_i)} p_i^{a_i-1} (1 - p_i)^{b_i-1}.$$

George (1992) also considered hierarchical models for p_1, \dots, p_s . Two models were considered, a beta model for p_i , as in Smith (1991), and a logit normal model for p_i , as in Castledine (1981) (these models are discussed in Sect. 13.11.5). In both instances, at least one variable is difficult to directly sample from but has a log-concave full conditional distribution. George (1992) used the adaptive rejection sampling algorithm of Gilks and Wild (1992) to sample from these distributions. In the years following 1992, other MCMC algorithms, such as the Metropolis-Hastings algorithm, the slice sampler, and Hamiltonian Monte Carlo, have become routinely used and are implemented in Bayesian software (cf. Lunn et al. (2000), Plummer (2003), Carpenter et al. (2017)). This has increased the applicability of Bayesian approaches for capture-recapture (and other) models as it relaxes the requirement of a log-concave full conditional distribution. We provide details of various MCMC algorithms in Seber and Schofield (2019).

13.7 List Dependence

If we consider an epidemiological population, data arise from multiple incomplete lists, e.g., inpatient, outpatient, doctor, and pharmaceutical records. In many situations, the assumption that lists are independent is unrealistic.

Madigan and York (1997) used decomposable Bayesian graphical models to allow for dependence among lists. A Bayesian graphical model specifies the probability structure of a statistical model using graph theory (see Lauritzen 1996, for details). A decomposable graph is defined by cliques and separators. A clique is a collection of variables that are related to one another. A separator variable partitions the graph so that the cliques are conditionally independent. Madigan and York (1997) applied this to list dependence, where each clique is a group of dependent lists. They specified a “hyper-Dirichlet” prior distribution that is defined over the cliques (Dawid & Lauritzen 1993) and has the benefit of allowing the posterior distribution for abundance to be specified in closed form.

Madigan and York (1997) gave the posterior estimates and model probabilities for each of the possible dependence structures. They found a model-averaged posterior for N that accounted for model uncertainty (see Appendix A.9.2 for further details). It is more variable than the model-specific posteriors as it includes the variation due to not knowing the true dependence structure. Such model uncertainty is an important feature of our knowledge about abundance, at least for the data considered by the authors.

The sensitivity of the posterior distribution was assessed for various choices of prior for the parameters. The default prior used for the probabilities was a Dirichlet distribution with parameter 1 and led to the most favored model having posterior probability of 0.37. If the Dirichlet prior parameter was changed to 0.5, the corresponding posterior probability increased to 0.48. If the Dirichlet prior parameter was changed to 2, the posterior probability dropped to 0.29, and the corresponding model was no longer the most favored model. Despite this, there was minimal change in the model averaged posterior distribution. For the three values of the Dirichlet parameter (1, 0.5, 2), the corresponding posterior mean was $\hat{N} = 731, 732, 730$. Madigan and York (1997) also discuss how covariates and hierarchical models could be accommodated.

King and Brooks (2001) specified list dependence through interaction terms in a log-linear modeling framework (Fienberg 1972). This allowed for a greater range of dependence relationships compared to restricting attention to decomposable models. They cited a contingency table example from Dellaportas and Forster (1999) where the best log-linear model was approximately 700 times better than the best decomposable graph.

The downside to the log-linear modeling framework is that the posterior distribution can no longer be obtained in closed form. To overcome this, King and Brooks (2001) used MCMC to generate samples from the posterior density. For a given dependence model, they used the Metropolis and Gibbs algorithms to generate posterior draws. The need to use MCMC also creates difficulties in accounting for model uncertainty. King and Brooks (2001) followed Madigan and York (1997) using Bayesian model averaging and implemented a reversible jump MCMC (RJMCMC) algorithm. The use of log-linear models greatly increased the size of the model set. There were 168 possible models that could have been explored by the RJMCMC algorithm for the example of King and Brooks (2001).

Fienberg et al. (1999) explored list dependency within the context of individual heterogeneity. They used a Rasch model, namely,

$$\log \left(\frac{p_{ij}}{1 - p_{ij}} \right) = \theta_i + \beta_j,$$

where p_{ij} is the probability that object i appears on list j , θ_i is a term that describes object i 's catchability, and β_j describes the effect of list j . If θ_i does not vary with i , the model reduces to model M_t with independent lists. If θ_i varies with i , then the θ_i are typically modeled as random effects, i.e., the θ_i are assumed to be realizations from a common distribution. This heterogeneity induces a specific form of list dependence. After marginalizing over θ , the model can be expressed as a log-linear model with an interaction term that depends on the number of lists on which an individual was observed, but not on the lists themselves. Fienberg et al. (1999) referred to this as a quasi-symmetry model of dependence.

To fit this model, Fienberg et al. used the Gibbs sampler. They alternated between updating β_j , sampling the parameters in the random effect distribution of θ , and a joint update for N and $\theta_1, \dots, \theta_N$. The joint update was necessary to account for one of the variables (N) controlling the dimension of the other variable (θ). They also mentioned that reversible jump MCMC (Green 1995) can be used to update N but preferred the joint update, as they found it to be faster and produce more stable chains. We provide details of different approaches for updating N in Sect. 13.10.

The distinction between heterogeneity and dependence models is blurred. Individual heterogeneity in the Rasch model implies a specific form of list dependence. An individual with high θ_i will be more likely to be seen on many lists, whereas an individual with low θ_i will be seen on few lists. As the value of θ_i is unknown, this will manifest as list dependence. An individual seen on list j will be more likely to be observed on other lists. An individual not observed on list j will be less likely to be seen on other lists.

13.8 Behavior

Model M_{tb} of Otis et al. (1978) allowed for the probability of capture to change as a consequence of earlier capture. In particular, the probability of capture of individual i , $i = 1, \dots, N$, in sampling occasion j , $j = 1, \dots, s$ can be written as

$$p_{ij} = \phi_j^{z_{ij}} p_j, \quad i = 1, \dots, N, \quad j = 1, \dots, s, \tag{13.11}$$

where $\phi_1 \equiv 1$ and $z_{ij} = 1$ if individual i has been captured before occasion j and $z_{ij} = 0$ otherwise, $i = 1, \dots, N$, $j = 1, \dots, s$. The parameter $\phi_j > 0$ modifies the capture probability. If $\phi_j < 1$, earlier captures lower the recapture probability, and we refer to such animals as “trap-shy.” When $\phi_j > 1$, earlier captures increase recapture probability, and we refer to animals as “trap-happy.” We may expect a

trap-shy response if the capture process is unpleasant or painful for the animals. We may expect a trap-happy response if baiting is used to help capture animals.

As described by Otis et al. (1978), the full model is overparameterized. Lee et al. (2003) considered a special case of (13.11) by setting $\phi_j = \phi$, $j = 1, \dots, s$, and a Gibbs sampler was used to fit the model. They assumed a scale prior for N and showed that the full conditional distribution for U is the negative-binomial distribution specified in (13.10), the same as for model M_t. This equivalence is not surprising as the probability that an individual is never observed is $\prod_{j=1}^s (1 - p_j)$ in both models. We note that it is only the conditional distributions that are the same; the marginal posterior distribution for U (and N) differ between the two models.

Lee et al. (2003) followed George (1992) in using the adaptive rejection sampler for sampling from the full conditional distributions for p_j , $j = 1, \dots, s$ and ϕ . In both instances, the support of the parameters is restricted to ensure that $\phi p_j < 1$, $j = 2, \dots, s$.

An earlier approach by Lee and Chen (1998) considered a re-parameterized version of the overparameterized model in (13.11). In this model, the recaptures do not inform estimation of N . A paucity of information about N is reflected in unstable inference and prior dependence (Lee et al. 2003).

13.9 Heterogeneity

It is reasonable to expect that there may be differences in the capture probabilities for individuals in the population. We summarize various approaches that have been considered for including capture heterogeneity using Bayesian methods. In outlining these models, we highlight the fundamental difficulty associated with specifying and fitting models that include individual heterogeneity. As shown by Link (2003), it is possible for two models to fit the observed data equally as well, but have very different estimates of N .

13.9.1 Rasch Models

As noted above, Fienberg et al. (1999) allowed for individual heterogeneity within the context of list dependence. Their model assumed that the probability of capture of individual i in occasion j was

$$\log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = \theta_i + \beta_j, \quad i = 1, \dots, N, \quad j = 1, \dots, s.$$

As individuals $i = r + 1, \dots, N$ are unobserved, further assumptions about θ are necessary. They assumed that θ_i are independent normal realizations with mean 0

and variance σ^2 , $i = 1, \dots, N$. This model assumed that differences in individual catchability, given by the value of θ_i , are realizations of a normal model.

King and Brooks (2008) considered a Rasch-type additive model with effects for behavior, time, and heterogeneity. The probability of capture for individual i on occasion j is assumed to follow

$$\text{logit}(p_{ij}) = \mu + \alpha_j + \beta z_{ij} + \gamma_i, \quad i = 1, \dots, N, \quad j = 1, \dots, s,$$

where $z_{ij} = 0$ if individual i has not been caught previous to occasion j , and $z_{ij} = 1$ otherwise. The parameter μ represents the overall mean capture probability, and β describes the behavioral effect. The values α_j , $j = 1, \dots, s$, are the time effects. These are assumed to be independent realizations from a normal distribution with mean 0 and unknown variance σ_α^2 . The values γ_i , $i = 1, \dots, N$, represent the individual effects, which are assumed to be independent realizations from a normal distribution with mean 0 and unknown variance σ_γ^2 .

The general model corresponds to model M_{thb} of Otis et al. (1978). The other seven models outlined by them are obtained by restricting various parameters and are also described in detail in Chap. 12:

| Constraint | Model |
|-------------------------------------------------------------------------|-----------------|
| $\beta = 0$ | M_{th} |
| $\gamma_i = 0, \forall i$ | M_{tb} |
| $\alpha_j = 0, \forall j$ | M_{bh} |
| $\beta = 0$ and $\gamma_i = 0, \forall i$ | M_t |
| $\beta = 0$ and $\alpha_j = 0, \forall j$ | M_h |
| $\gamma_i = 0, \forall i$ and $\alpha_j = 0, \forall j$ | M_b |
| $\beta = 0$ and $\gamma_i = 0, \forall i$ and $\alpha_j = 0, \forall j$ | M_0 |

The authors used a scale prior for N so that $f(N) \propto N^{-1}$. The prior for μ varies between models. For most models, the prior is normal with mean 0 and variance 10. The exception is model M_0 for which the prior for μ is a logistic distribution with location 0 and scale 1. This corresponds to an induced uniform distribution for the transformed variable $\text{logit}^{-1}(\mu)$. The prior for σ_α^2 is an inverse gamma distribution with shape 4 and scale 3, so that the prior mean and variance are 1 and 0.5, respectively. The prior for σ_γ^2 is also an inverse gamma distribution with shape 4 and scale 3. The prior for β is normal with mean 0 and variance σ_β^2 , where σ_β^2 is unknown with an inverse gamma prior with shape 4 and scale 3. This corresponds to a three-parameter Student's- t prior for β with 8 degrees of freedom, location 0, and scale 0.75.

A feature of King and Brooks (2008) was the use of RJMCMC, mentioned above, to account for model uncertainty. They fitted the model to three datasets to demonstrate the approach. In all three examples, the same model (M_{th}) has the largest posterior model probability. They reported that there is little sensitivity to

the priors chosen, either with regard to posterior model probabilities or estimation of population size N .

13.9.2 Mixture Model

Basu and Ebrahimi (2001) explored models for error detection. The setup is that there are N errors (say in computer code) and s reviewers, and they considered the case where errors are non-homogeneous. They developed a mixture model that allows for errors to be easy to detect (E) or hard to detect (H), namely,

$$p_i = \begin{cases} \pi_H, & \text{with probability } \alpha, \\ \pi_E, & \text{with probability } 1 - \alpha. \end{cases}$$

A beta prior was assumed for α with parameter $\kappa\gamma$ and $\kappa(1 - \gamma)$. The prior mean is $E[\alpha] = \gamma$, which can be chosen to represent the prior belief about the proportion of hard errors. The value of κ can be chosen to represent the strength of our prior belief. This can be aided by noting that the prior variance is

$$\frac{\gamma(1 - \gamma)}{1 + \kappa}.$$

The authors assumed uniform priors for π_H and π_E with support $[h_1, h_2]$ and $[e_1, e_2]$ respectively, but constrained $\pi_H \leq \pi_E$, $h_1 \leq e_1$ and $h_2 \leq e_2$. They presented a result that stated the posterior distribution for N is proper when either a uniform or scale prior for N is assumed, provided h_1 and e_1 are positive.

The Gibbs sampler is used to obtain samples from the posterior distribution. For appropriate prior choices, the full conditional distributions can be sampled from easily. The full conditional distribution for α is a beta distribution with parameters $a = \kappa\gamma + m_H$ and $b = \kappa(1 - \gamma) + m_E$, where m_H is the number of individuals that are in mixture H , i.e.,

$$m_H = \sum_{i=1}^N I(p_i = \pi_H),$$

where $I(\cdot)$ is the indicator function. The variable m_E is the number of individuals in mixture E , with $m_E = N - m_H$.

The joint full conditional distribution for π_H, π_E is proportional to

$$\begin{aligned} \pi_H^{l_H} (1 - \pi_H)^{s m_H - l_H} \pi_E^{l_E} (1 - \pi_E)^{s m_E - l_E} \times \\ I(h_1 < \pi_H < h_2) I(e_1 < \pi_E < e_2) I(\pi_H < \pi_E), \end{aligned} \tag{13.12}$$

where l_H are the number of captures for individuals in mixture H ,

$$l_H = \sum_{i=1}^N y_i I(p_i = \pi_H),$$

where y_i is the number of captures for individual i . The quantity l_E is equivalently defined for the E mixture. Basu and Ebrahimi (2001) suggested that this joint full conditional distribution can be sampled by drawing two truncated beta random variables. The first is for π_H ; it is truncated between h_1 and h_2 and has parameters $a_H = 1 + l_H$ and $b_H = 1 + s m_H - l_H$. The second is for π_E ; it is truncated between $\max(e_1, \pi_H)$ and e_2 and has parameters $a_E = 1 + l_E$ and $b_E = 1 + s m_E - l_E$.

Unfortunately their the sampling scheme does not generate samples from the correct joint full conditional distribution. We would obtain a different distribution if we were to reverse the order of updating, i.e., sample π_E from (e_1, e_2) before sampling π_H from $(h_1, \min(h_2, \pi_E))$. This can be confirmed numerically.

The amount of error depends on the true marginal posteriors of π_H and π_E . If there is minimal overlap in their density, the scheme of Basu and Ebrahimi (2001) will have little error. The error increases with the overlap in density between π_H and π_E . The problem can be fixed in one of two ways. The first is to correctly sample from the distribution in (13.12). This can be done by generating π_H and π_E from independent beta distributions with parameters a_H, b_H and a_E, b_E , respectively. If $\pi_H < \pi_E$, the draw is kept. If $\pi_H > \pi_E$, the sample is discarded, and we repeat until $\pi_H < \pi_E$. The second approach is to update π_H and π_E separately within the Gibbs sampler. The full conditional distribution for π_H is a beta distribution with parameter a_H and b_H truncated between h_1 and $\min(h_2, \pi_E)$. The full conditional distribution for π_E is beta with parameter a_E and b_E , truncated between $\max(e_1, \pi_H)$ and e_2 .

The final step is the update of N . As with the Rasch model, this is made difficult with the dimension of the capture probabilities given by the value of N . Basu and Ebrahimi (2001) followed Fienberg et al. (1999) in updating N and p jointly, and averaging over the mixture, they showed that N can be updated from the distribution proportional to

$$\frac{N!}{(N-n)!} [\alpha(1-\pi_H)^s + (1-\alpha)(1-\pi_E)^s]^N f(N).$$

This is of the same form as (13.9), and for a variety of choices of $f(N)$, the full conditional distribution is of known form. Basu and Ebrahimi (2001) sampled N and then determined the class of each observation $i = 1, \dots, N$. These are independent Bernoulli distributions with probability

$$\frac{\alpha\pi_H^{y_i}(1-\pi_H)^{s-y_i}}{\alpha\pi_H^{y_i}(1-\pi_H)^{s-y_i} + (1-\alpha)\pi_E^{y_i}(1-\pi_E)^{s-y_i}}.$$

The authors noted that this model can be generalized to $k > 2$ types of error. In this case, we have π_1, \dots, π_k with $\pi_1 \leq \dots \leq \pi_k$. The beta prior can be generalized to a Dirichlet distribution.

Basu (1998) considered a specific form of model M_{th} with

$$p_{ij} = e_i r_j,$$

where e_i is the effect of error i and r_j is the effect of reviewer j . Mixture models were assumed for both e_i and r_j . Two approaches for updating N were considered, and we describe these in Sect. 13.10.

A similar approach was considered by Arnold et al. (2010) who fitted a special case of model M_{th} in terms of two-way mixtures. The model is a Rasch-type model, where $\text{logit}(p_{ij})$ comprises an error effect, a reviewer effect, and an interaction term. The error effects were considered a finite mixture with A classes, the reviewer effects were specified as a finite mixture with B classes, and the interaction term was defined across the $A \times B$ mixture levels. There were constraints placed on the main effects and the interaction term to ensure identifiability. The interaction term was multiplied by a binary parameter ζ so that when $\zeta = 0$, the interaction term is not included in the model. The authors used RJMCMC to account for model uncertainty.

Ghosh and Norris (2005) extended the mixture model of Basu and Ebrahimi (2001) to account for behavior and heterogeneity. They assumed that the probability of capture for individual i , $i = 1, \dots, N$ is p_{1i} if individual i has not been caught previously and p_{2i} if it has been captured previously. Let \mathbf{p}_i be the collection (p_{1i}, p_{2i}) . It was assumed that \mathbf{p}_i , $i = 1, \dots, N$ takes the value $\boldsymbol{\theta}_h = (\theta_{h1}, \theta_{h2})$, $h = 1, \dots, k$ with probability α_h such that $\sum_{h=1}^k \alpha_h = 1$. The authors noted that for large k , this approximates a Dirichlet process prior. An interesting feature of this model is that we can have mixtures of trap-happy and trap-shy individuals in the population. This is an example of model M_{bh} (Otis et al. 1978), and setting $k = 1$ results in a behavior model (Model M_b). Enforcing $\theta_{h1} = \theta_{h2}$, $h = 1, \dots, k$ gives a heterogeneity model (Model M_h).

To ensure identifiability, Ghosh and Norris (2005) assumed that $0 < \theta_{11} < \dots < \theta_{1k} < 1$. This avoided issues around label switching and is common in mixture models; we saw a similar assumption in the model of Basu and Ebrahimi (2001). We note that such assumptions are not innocent and do change the model. To see this, consider the case where the (unconstrained) posteriors for $\theta_{21} - \theta_{11}$ overlaps 0. The identifiability constraint restricts the parameter space so that the posterior for $\theta_{21} - \theta_{11}$ is entirely positive. An alternative is to not constrain the parameters and account for label switching via post-processing, e.g., using the approach of Stephens (2000).

Ghosh and Norris (2005) mentioned that many model selection processes are difficult to adapt to mixture model settings. They outlined a model selection process based on the posterior predictive distribution. Also, they calculated GGC, defined as the expected mean square predicted error on the log scale, where the expectation is calculated across the posterior predictive distribution.

13.9.3 Latent Class Model

Latent class models were suggested by Bartolucci et al. (2004) to account for individual heterogeneity and can account for complex dependence between lists. Latent class models are extensions of mixture models. Each individual can belong to one of k classes. An individual is assigned to class h with probability α_h and has probability π_{jh} of being caught on occasion (or list) j . The authors used a constraint $\alpha_1 < \alpha_2 < \dots < \alpha_k$. If $\pi_{jh} = \pi_h$, $\forall h, j$, the model collapses to the mixture model of Basu and Ebrahimi (2001).

The model of Bartolucci et al. (2004) had two other notable features. The first is that the number of classes k is included in the model as an unknown parameter. The prior distribution for k is discrete uniform between 1 and k_{max} . The second is that abundance N is not included in the model. The authors conditioned on an individual being observed, and did not model the 0...0 capture history. Instead, they estimated N in a second stage as $r/(1 - \hat{p}_{0...0})$, where $p_{0...0}$ is the probability an individual in the population is not observed during the study given by

$$p_{0...0} = \sum_{h=1}^k \alpha_h \prod_{j=1}^s (1 - \pi_{jh}).$$

Model fitting was difficult due to the variable k defining the dimensionality of other variables in the model (specifically α and π). The authors used the RJMCMC algorithm of Green (1995) and updated the variables α and π along with k . They defined a “combine proposal” that is used when the proposed value of k (denoted k') is less than the current value of k (in the description below, we assume $k' = k - 1$). A pair of adjacent classes (h_1, h_2) were randomly chosen and merged into a new class h^* . The proposed values of π and α have

$$\alpha'_{h^*} = \alpha_{h_1} + \alpha_{h_2} \quad \text{and} \quad \pi'_{jh^*} = \pi_{jh_2}, \quad j = 1, \dots, s.$$

The other values of α' and π' are the same as those for α and π , other than to allow for the reduction in dimension to k' . This scheme did not combine the classes h_1 and h_2 but removed class h_1 and reallocated all individuals into class h_2 .

Likewise, Bartolucci et al. (2004) defined a “split proposal” that is used when k' is greater than the current value of k (in the description below we assume $k' = k + 1$). A random class h^* is chosen and split into two new classes h_1 and h_2 . The proposed values of π and α have

$$\begin{aligned} \alpha'_{h_1} &= \pi_{h^*} u & \alpha'_{h_2} &= \pi_{h^*}(1 - u) \\ \pi'_{jh_1} &= v_j & \pi'_{jh_2} &= \pi_{jh^*}, \quad j = 1, \dots, s, \end{aligned}$$

where u is a realization from density $f(u)$ and v_j , $j = 1, \dots, s$ are independent realizations from density $g(v)$, with both densities having support in $(0, 1)$. The

authors used a Beta(2, 4) distribution for $f(\cdot)$ and beta(1, 1) distribution for $g(\cdot)$. The other values of α' and π' were the same as those for α and π , other than to allow for the increase in dimension to k' . This scheme introduced a new class (h_1) such that $100u\%$ of the individuals in class h^* are allocated to class h_1 .

Implementing the “combine proposal” and “split proposal” led to a specific implementation of RJMCMC, and the acceptance probability was found accordingly. We note that different schemes for updating k are possible. For example, when combining, we could choose π'_{jh^*} as $(\pi'_{jh_1} + \pi'_{jh_2})/2$. However, in devising such modifications, we need to ensure that (i) the algorithm remains reversible and (ii) that we re-derive the acceptance probability to accommodate the specific scheme that we have adopted. Details of the RJMCMC algorithm are given in Sect. 13.10.9.

Bartolucci et al. (2004) implemented a delayed rejection algorithm, giving a second chance for acceptance, if the initial proposal was rejected. The latent class model was fitted to the snowshoe hare data set used by Cormack (1992). They found considerable posterior uncertainty in k . The most probable value was $k = 3$, and this had posterior probability of approximately 0.2. Non-negligible posterior probabilities were evident for all values of k from 1 to 10.

13.9.4 Non-parametric Approaches

Various Bayesian non-parametric approaches have been proposed. Consider individual heterogeneity, where p_i represents the capture probability for individual i . Tardella (2002) assumed that p_1, \dots, p_N come from some distribution G , but do not specify its functional form, and instead rewrote the model in terms of the first S moments of the distribution G , treating these as unknown quantities. In practice, the author parameterized in terms of the canonical moments rather than raw moments.

Farcomeni and Tardella (2010) took a similar approach and included heterogeneity, behavior, and time effects. The probability of capture for individual i , $i = 1, \dots, N$, in capture occasion j , $j = 1, \dots, s$, is

$$p_{ij} = \begin{cases} \beta + (1 - \beta)\alpha[\gamma_j + (1 - \gamma_j)\delta_j\theta_i] & \text{until first capture} \\ \gamma_j + (1 - \gamma_j)\delta_j\theta_i & \text{after first capture.} \end{cases}$$

The parameters α and β described behavioral effects. The remaining term

$$\gamma_j + (1 - \gamma_j)\delta_j\theta_i$$

can be considered a latent class process. An individual in capture occasion j belongs to the first latent class with probability γ_j , and an individual in this class is encountered with certainty. Otherwise, an individual falls into the second latent class with probability $1 - \gamma_j$. An individual in this class is captured with probability $\delta_j\theta_i$. The variable θ_i is assumed to be a realization from an unspecified distribution G . As

with Tardella (2002), this distribution is expressed in terms of its first s canonical moments.

Manrique-Vallier (2016) explored a non-parametric latent class model. This is where a latent class model potentially has an infinite number of states. The class probabilities are given by $\pi = (\pi_1, \pi_2, \dots)$. The prior for π is a Dirichlet process, also referred to as a stick-breaking process. The author commented that this prior concentrates the bulk of the probability mass into its first few coordinates, inducing sparsity and avoiding overfitting. For computational convenience, a finite-dimensional approximation was used with an upper bound on the number of states, and a Gibbs sampling scheme was used to obtain samples from the posterior distribution.

We find that non-parametric approaches for describing heterogeneity cannot overcome fundamental problems with identifiability in capture-recapture models with heterogeneity. As stated by the author:

Thus, having shown the effectiveness and computational convenience of our [nonparametric] method, we may be tempted to see it as a sort of silver bullet that allows assumption-free multiple-recapture estimation. Unfortunately, such belief would be unfounded. As Fienberg (1972) warned, multiple-recapture estimation – as any other extrapolation technique – relies on the untestable assumption that the model that describes the observed counts also applies to the unobserved ones. This makes the problem essentially nonidentifiable in the sense that we can produce infinite models that assign different probabilities for the unobserved cell **0** and the same set of probabilities [for the observed capture histories]...We note that this is an intrinsic limitation of multiple-recapture estimation, regardless of the particular implementation. This was also observed by Link (2003) in the simpler context of binomial estimation with heterogeneity.

13.10 Estimating Abundance

In this section, we explore the various approaches in the to find the posterior distribution for the abundance N . This is difficult when there is heterogeneity in detection because N defines the dimensionality of detection parameters, e.g., in model M_h , the detection parameters are $\mathbf{p}_N = (p_1, \dots, p_N)$. We are unable to use “standard” MCMC algorithms, such as the Gibbs sampler. It makes no sense to update N while holding \mathbf{p}_N fixed; any change to the value of N necessitates a change in dimension of \mathbf{p}_N .

Let y_i be the number of detections of individual i across the s sampling occasions. We define the collection $\mathbf{y} = (y_1, \dots, y_N)'$, where the observations are arranged so that $y_{r+1} = \dots = y_N = 0$. Model M_h is

$$\Pr(\mathbf{y}|\mathbf{p}_N, N) = \frac{N!}{(N-r)!f_1!\dots f_s!} \prod_{i=1}^N \binom{s}{y_i} p_i^{y_i} (1-p_i)^{s-y_i}, \quad (13.13)$$

where f_j is the number of individuals with $y_i = j$,

$$f_j = \sum_{i=1}^r I(y_i = j), \quad j = 1, \dots, s.$$

The combinatorial term gives the number of ways to obtain the data \mathbf{y} from a population of size N .

The model in (13.13) is overspecified as a likelihood for N . Further assumptions are necessary as there is no direct information about p_{r+1}, \dots, p_N . The standard approach is to treat \mathbf{p}_N as latent random variables and include a hierarchical model $g(p|\theta)$,

$$p_i \stackrel{iid}{\sim} g(p|\theta), \quad i = 1, \dots, N,$$

where θ are unknown parameter(s) for the unspecified distribution $g(\cdot)$. Some of the approaches we outline below rely on specific choices of $g(\cdot)$.

The joint model for \mathbf{y} and \mathbf{p}_N is

$$\Pr(\mathbf{y}, \mathbf{p}_N | \theta, N) = \frac{N!}{(N-r)! f_1! \dots f_s!} \prod_{i=1}^N \binom{s}{y_i} p_i^{y_i} (1-p_i)^{s-y_i} g(p_i | \theta). \quad (13.14)$$

This has been referred to as the complete data likelihood for model M_h , where the data have been “completed” by including the latent variable \mathbf{p}_N , and is an example of data augmentation.

Data augmentation was introduced by Tanner and Wong (1987) for the situation where a posterior density $f(\theta|y)$ is difficult to evaluate or sample from. It may be possible to augment the observed data y with latent data z to give a “posterior” $f(\theta|y, z)$ that is more convenient to work with. One such example is in the case of missing data, where z represents the missing observations that “complete” the data. The idea of Tanner and Wong (1987) was to include z in the model to enable easier model evaluation while making use of Monte Carlo approaches (which were limited in 1987) to marginalize over z and ensure that we obtain the correct posterior $f(\theta|y)$. The complete data likelihood $f(y, z|\theta)$ is the likelihood found when using the complete data y and z . The likelihood $f(y|\theta)$ can be referred to as the observed data likelihood.

The observed data likelihood for model M_h is found by marginalizing \mathbf{p}_N out of (13.14). Burnham and Overton (1978) defined it in terms of sufficient statistics f_1, \dots, f_s as

$$\Pr(f_1, \dots, f_s | N, \theta) = \frac{N!}{(N-r)! f_1! \dots f_s!} \pi_0^{N-r} \prod_{h=1}^s q_h^{f_h}, \quad (13.15)$$

where $r = \sum_{h=1}^s f_h$ is the number of individuals observed throughout the study and

$$q_h = \int_0^1 \binom{s}{h} p^h (1-p)^{s-h} g(p|\theta) dp, \quad h = 0, 1, \dots, s. \quad (13.16)$$

We outline below the various Bayesian approaches that have been used for estimating abundance. Where possible, we describe these methods in terms of model M_h .

13.10.1 Marginalize over p

If we assume that $g(p|\theta)$ is a beta distribution with parameters α and β , the integrals in (13.16) can be found in closed form (Burnham & Overton 1978)

$$q_h = \binom{s}{h} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(h + \alpha)\Gamma(s - h + \beta)}{\Gamma(s + \alpha + \beta)}, \quad h = 0, \dots, s.$$

This is known as the beta-binomial model (Dorazio & Royle 2003).

Standard MCMC algorithms can be implemented if the beta-binomial model is used as an observed data likelihood. The parameters in the model are N , α , and β . The abundance N no longer defines the dimension of another variable. If we assume a scale prior for N , $f(N) \propto N^{-1}$, then the full conditional distribution for N is negative-binomial.

The beta model may be re-parameterized in terms of its mean and variance,

$$\begin{aligned} \mu &= \frac{\alpha}{\alpha + \beta}, \\ \sigma^2 &= \frac{\mu(1 - \mu)}{\alpha + \beta + 1}. \end{aligned}$$

Our experience is that this parameterization often has lower posterior correlation than that in terms of α and β .

It is also possible to analytically marginalize over \mathbf{p}_N when $g(p|\theta)$ is a finite mixture. In a K -part mixture, there are K possible detection probabilities π_1, \dots, π_K . The detection probability π_j is assigned to an individual with probability w_j , where w_1, \dots, w_K are referred to as the mixture weights with $\sum_j w_j = 1$. This leads to

$$q_h = \binom{s}{h} \sum_{j=1}^K w_j \pi_j^h (1 - \pi_j)^{s-h}, \quad h = 0, \dots, s.$$

If $K = 2$ we have

$$q_h = \binom{s}{h} \left(w_1 \pi_1^h (1 - \pi_1)^{s-h} + (1 - w_1) \pi_2^h (1 - \pi_2)^{s-h} \right).$$

Again, standard MCMC algorithms can be used if we adopt a discrete mixture and make use of the closed form of the marginal likelihood in (13.15). The model has unknown parameters N, π_1, \dots, π_K , and w_1, \dots, w_{K-1} .

Label switching can be a problem when fitting mixture models. It manifests as multiple modes in the likelihood and subsequently in the posterior density. It may be possible to correct label switching using post-processing, e.g., Stephens (2000).

13.10.2 Joint Update

Basu (1998) and Fienberg et al. (1999) used the idea of marginalization within the context of a Gibbs sampler. A standard Gibbs sampler proceeds by drawing a value from the full conditional distribution of each parameter in turn. Consider model M_h with a $K = 2$ finite mixture prior. The unknowns (parameters) in the model are $N, \mathbf{p}_N, \pi_1, \pi_2$, and w_1 . As N defines the dimensionality of \mathbf{p}_N , we cannot apply the Gibbs sampler without modification. It does not make sense to consider N conditional on \mathbf{p}_N since the dimension of \mathbf{p}_N changes with N .

Basu proposed to implement a joint update of \mathbf{p}_N and N . Letting θ represent the parameters π_1, π_2, w_1 , the joint conditional distribution of N and \mathbf{p}_N is

$$f(N, \mathbf{p}_N | \mathbf{y}, \theta) = f(N | \mathbf{y}, \theta) f(\mathbf{p}_N | N, \mathbf{y}, \theta). \quad (13.17)$$

The abundance N is generated from $f(N | \mathbf{y}, \theta)$, a distribution that does not depend on \mathbf{p}_N . We can then use the realized N to generate the detection parameters p_1, \dots, p_N . The distribution $f(N | \mathbf{y}, \theta)$ is

$$\begin{aligned} f(N | \mathbf{y}, \theta) &= \int f(N, \mathbf{p}_N | \mathbf{y}, \theta) d\mathbf{p}_N, \\ &\propto f(N) \int f(\mathbf{y} | N, \mathbf{p}_N) f(\mathbf{p}_N | \theta, N) d\mathbf{p}_N. \end{aligned}$$

The integrand in the second line is related to those we consider in Sect. 13.10.1.

Basu implemented a joint update for a special case of model M_h , written in terms of a finite mixture model. Fienberg et al. (1999) also jointly updated N and \mathbf{p}_N , although they were unable to analytically marginalize over \mathbf{p}_N for some of the models they considered for $g(p | \theta)$. They used numerical integration, with no further details as to how this was done. Numerical integration is discussed in more depth in the next section.

13.10.3 Numerical Integration

There does not appear to have been a wide adoption of numerical marginalization within an MCMC sampler for many years following Fienberg et al. (1999). Bayes approaches that used numerical integration for estimating abundance were considered again by Bonner and Schofield (2014) who explored a model where an individual covariate x informs the capture probability, namely,

$$\begin{aligned} \text{logit}(p_i) &= \beta_0 + \beta_1 x_i \\ x_i &\sim \text{normal}(\mu, \sigma^2), \quad i = 1, \dots, N. \end{aligned} \quad (13.18)$$

A subtle difference from model M_h is that the capture probability is defined for all observed individuals (conditional on β_0, β_1) as we observe $\mathbf{x} = (x_1, \dots, x_r)$. The values x_{r+1}, \dots, x_N are unknown. They gave the likelihood

$$\mathcal{L}(\theta, N; \mathbf{x}, \mathbf{y}) = \binom{N}{r} \pi_0^{N-r} \prod_{i=1}^r p_i^{y_i} (1 - p_i)^{s-y_i} \phi\left(\frac{x_i - \mu}{\sigma}\right),$$

where $\phi(\cdot)$ is the standard normal density function, p_i is the inverse logit of (13.18), i.e.,

$$p_i = \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)},$$

and π_0 is the probability of being unseen throughout the experiment,

$$\pi_0 = \int_{-\infty}^{\infty} \left(\frac{1}{1 + \exp(\beta_0 + \beta_1 x)} \right)^s \phi\left(\frac{x - \mu}{\sigma}\right) dx. \quad (13.19)$$

For simplicity, the authors fixed $\beta_0 = 0$, $\beta_1 = 1$, and $\sigma = 1$, so that the only unknown parameters are μ and N .

Consider updating μ with the Metropolis-Hastings algorithm. If we are in iteration t , with current values given by $\mu^{(t)}$, $N^{(t)}$, we

1. Propose a candidate value μ^* from the density $q(\mu^* | \mu^{(t)})$.
2. Accept the candidate with probability $\min(1, \alpha)$, where

$$\alpha = \frac{f(\mu^* | N^{(t)}, \mathbf{x}, \mathbf{y})}{f(\mu^{(t)} | N^{(t)}, \mathbf{x}, \mathbf{y})} \frac{q(\mu^{(t)} | \mu^*)}{q(\mu^* | \mu^{(t)})} \quad (13.20)$$

3. If candidate is accepted, we set $\mu^{(t+1)} = \mu^*$; otherwise, we set $\mu^{(t+1)} = \mu^{(t)}$.

The problem is in step 2 of the algorithm above. The full conditional distribution includes π_0 , and we must evaluate π_0 at the candidate value μ^* and the current value² $\mu^{(t)}$.

Bonner and Schofield (2014) proposed to use Monte Carlo (MC) integration to estimate π_0 . To obtain an MC estimate of the integral in (13.19), K values z_1, \dots, z_K are generated from a normal distribution with mean μ and variance σ^2 , and then evaluate

$$\hat{\pi}_0 = \frac{1}{K} \sum_{i=1}^K \left(\frac{1}{1 + \exp(\beta_0 + \beta_1 z_i)} \right)^s. \quad (13.21)$$

The estimator is stochastic in the sense that each sample z_1, \dots, z_K leads to a different estimate of π_0 .

The use of MC integration means that the numerator and denominator of α in (13.20) involve MC estimates of π_0 . The authors referred to such an algorithm as MCWM (Monte Carlo within Markov chain Monte Carlo sampling). They relied on theoretical results from Andrieu and Roberts (2009), who showed that provided the MC estimator is unbiased and the size of the MC sample (K) is large, the stationary distribution found using MCWM approximates the true posterior density.

The term that was estimated with MC integration is $Q(\mu) = \pi_0^{N-r}$. The MC estimator used by Bonner and Schofield (2014) was

$$\tilde{Q}(\mu) = \hat{\pi}_0^{N-r},$$

where $\hat{\pi}_0$ is given in (13.21). This estimator is a biased but consistent estimator of $Q(\mu)$. The conjecture of the authors was that provided K was large enough, this approach would generate draws that approximate the true posterior distribution. They dismissed an unbiased MC estimator of $Q(\mu)$, as the complexity of the estimator depended on N , the very problem they were trying to avoid. Full details of the algorithm are given by the authors, including two extensions to improve its performance. The algorithm is completed with a negative-binomial update for the parameter N . As we have numerically marginalized over x_{r+1}, \dots, x_N , the parameter N no longer determines the dimension of any other variables in the model.

The authors used simulation to assess their methodology, comparing it to RJMCMC and data augmentation (these approaches are described in Sects. 13.10.9 and 13.10.7). They found that the MCWM algorithm performed well. When compared to the other approaches, it had less error and was much more efficient when a small proportion of the population was marked.

² In practice, the value of π_0 at $\mu^{(t)}$ is required in the previous iteration of the algorithm. It is more efficient to “carry forward” this quantity than reevaluate it.

13.10.4 Semi-complete Data Likelihood

King et al. (2016) generalized the approach of Bonner and Schofield (2014). If we consider model M_h , their idea is to split \mathbf{p}_N into $\mathbf{p}_{obs} = (p_1, \dots, p_r)$ and $\mathbf{p}_{mis} = (p_{r+1}, \dots, p_n)$. The first collection, \mathbf{p}_{obs} , gives the capture probabilities for the observed individuals, whereas \mathbf{p}_{mis} are the values relating to individuals in the population never seen. King et al. numerically marginalized over \mathbf{p}_{mis} while including \mathbf{p}_{obs} in the model as latent variables.

The rationale for numerically marginalizing \mathbf{p}_{mis} is twofold. It is computationally efficient as only a single integral is required. Secondly, the abundance N no longer defines the dimension of any variables. The authors used Gauss-Hermite quadrature to marginalize, which is a deterministic approximation to the integral, unlike the MC integral in Bonner and Schofield (2014). A consequence is that King et al. (2016) implemented the approach into Bayesian software JAGS (Plummer 2003). A potential disadvantage is that quadrature becomes increasingly difficult for multidimensional integrals, should they be required.

King et al. (2016) did not numerically marginalize \mathbf{p}_{obs} . This would make the algorithm more complex and is unlikely to greatly improve efficiency in many cases. For model M_h , we would need to introduce an additional s quadrature calculations to approximate the integrals in (13.16). Including p_1, \dots, p_r as latent variables is comparatively simple. Standard MCMC algorithms such as the Gibbs sampler or the Metropolis-Hastings algorithm can be used. The authors used the semi-complete approach to fit model M_h and a spatial capture-recapture model. In both of these applications, the semi-complete approach has improved efficiency when compared to other standard approaches.

13.10.5 Laplace Approximation

The approaches in Sects. 13.10.3 and 13.10.4 can be thought of as numerical methods to approximate an integral. In contrast, Herliansyah et al. (2022) investigated the use of an analytical approximation to solve the necessary integrals. The marginal (or observed data) likelihood for Model M_h is given in (13.15) and includes $s + 1$ integrals given in (13.16) and repeated here

$$q_h = \int_0^1 \binom{s}{h} p^h (1-p)^{s-h} g(p|\theta) dp, \quad h = 0, 1, \dots, s.$$

Their approach was to use a Laplace approximation to these integrals, where each integrand in (13.16) is approximated by a Gaussian. They used maximum likelihood to estimate the parameters of the model, but Bayesian methods could also be adopted. As described in Sect. 13.11.5, Smith (1991) implemented a Laplace approximation when including a hierarchical model for \mathbf{p}_s .

Herliansyah et al. (2022) considered two extensions. The first is to allow for a more accurate approximation. The standard Laplace approximation is found by taking the first two terms in a Taylor series expansion of the integrand. They also considered using four terms in the Taylor series to make the approximation more accurate. The second extension is to allow for multidimensional integrals, which is applied when fitting an open population capture-recapture model with individual-specific time-varying covariates. In this case, they also used automatic differentiation due to the difficulty of obtaining closed-form expressions for the derivatives needed for the Laplace approximation.

Simulations were used to assess the performance for model M_h . The standard Laplace approximation exhibited some negative bias with coverage below nominal rates. The fourth-order approximation had improved bias and coverage, with properties similar to those found with Gauss-Hermite quadrature. Fitting the model with Laplace approximation was faster than alternative methods, particularly for complex models, and more details are given by Herliansyah et al. (2022).

13.10.6 Bounding Abundance

A solution proposed by Basu (1998) was to specify an upper bound M for the abundance. The approach was outlined for model M_{th} ; here we describe it for model M_h .

The upper bound enforces a finite set of values that N can take; $N \in \{0, 1, \dots, M\}$. After observation of the data, the possible values of N are reduced; $N \in \{r, r + 1, \dots, M\}$. The key step in the approach is to include the detection probabilities for all M possible individuals $\mathbf{p}_M = (p_1, \dots, p_M)$ in the model using data augmentation (see Sect. 13.10.7). The abundance N does not determine the dimensionality of \mathbf{p}_M , and standard MCMC algorithms can be used. Basu (1998) noted that updating \mathbf{p}_M using Gibbs sampling is done in two steps. The first step updates p_1, \dots, p_N according to their full conditional distribution. The second step updates p_{N+1}, \dots, p_M from the model $g(p|\theta)$.

A potential shortcoming of bounding abundance is that the quantity M needs to be chosen by the analyst. If M is chosen too small, it can become influential, truncating the posterior distribution of N . As M increases, the computational burden increases. Moreover, if M is too large, the values p_{N+1}, \dots, p_M can “overwhelm” the full conditional distribution of θ leading to slow mixing of the Markov chain (Basu 1998). However, a simulation study conducted in Bonner and Schofield (2014) found that there were situations where increasing M increased efficiency of the algorithm (as measured by the effective sample size of N per second). A possible explanation is that a large superpopulation can generate more populations of size N , allowing the composition of the population to change more freely. It is unclear how general this result is.

Specifying an upper bound allows models that include heterogeneity to be implemented in general-purpose Bayesian software such as BUGS (Lunn et al.

2009, 2013) and JAGS (Plummer 2003). Such software packages were influential in making Bayesian modelling accessible to a wider group of applied statisticians and quantitative scientists. The capture-recapture community was no exception, e.g., Durban and Elston (2005), Royle et al. (2007b), Gimenez et al. (2009), and Schofield et al. (2009).

The first use of an upper bound for abundance in BUGS was Durban and Elston (2005), following Fienberg et al. (1999) in fitting Rasch-type models for detection. They allowed the model for p_i to differ depending on whether $i \leq N$ or $i > N$. When $i \leq N$ so that i is a member of the population, the model for p_i is $g(p|\theta)$. When $i > N$, the model for p_i is arbitrary and can be chosen by the analyst. Durban and Elston (2005) refer to this distribution as a pseudo-prior. The name reflects that the choice of this model does not influence the posterior distribution of the parameters of interest, aside from the efficiency of the MCMC algorithm. They emphasized that a poor choice of the pseudo-prior will lead to poor mixing in the MCMC algorithm and suggested that a pilot run be used to determine a good choice. One possible default choice is to use $g(p|\theta)$ as suggested by Basu (1998). While more efficient choices certainly exist, this appears to work adequately well in practice (Schofield & Barker 2014).

13.10.7 Data Augmentation

Another approach that relies on an upper bound for abundance is that of Royle et al. (2007b). We follow their derivation and motivate the approach when there is no heterogeneity in detection between individuals or by sampling occasion. This allows the focus to be on highlighting connections between models for site occupancy data and capture-recapture data.

Consider an occupancy study where the goal is to detect a certain species at M sites. If each of these sites is visited s times, the data is given by $\mathbf{y} = (y_1, \dots, y_M)$, where element $y_i = k$ denotes that the species of interest was detected in $k \in \{0, 1, \dots, s\}$ of the s visits. The occupancy model allows for false-negative observations, i.e., visits where the animal was truly present but not detected. In the simplest occupancy model, there are two parameters: ψ , the probability that a site is occupied, and p , the probability of detection given an occupied site.

To complete model specification, we introduce a partially observed collection $\mathbf{z} = (z_1, \dots, z_M)$. The value $z_i = 1$ denotes that site i was occupied, while $z_i = 0$ otherwise. The variable is partially observed because we know that $z_i = 1$ if site i had a recorded detection ($y_i > 0$), whereas the value z_i is unknown for site i if $y_i = 0$. The model is

$$\begin{aligned} z_i &\sim \text{Bern}(\psi), \quad i = 1, \dots, M, \\ y_i | z_i &\sim \text{Binomial}(s, p z_i), \quad i = 1, \dots, M, \quad j = 1, \dots, s, \end{aligned} \tag{13.22}$$

where $\text{Bern}(\pi)$ denotes a Bernoulli random variable with success probability π , and $\text{Binomial}(m, \pi)$ represents a binomial random variable with index m and success probability π . The probability $p z_i$ reflects that if the site is occupied ($z_i = 1$), the probability of detection is p ; otherwise, the probability of detection is 0. Marginalizing over z_i allows us to write the likelihood as in (2.2) of Royle et al. (2007b), namely,

$$\begin{aligned} \mathcal{L}(N, \psi, p; \mathbf{y}, M, r) = \\ \left[\frac{N!}{(N-r)!} p^{\sum_i y_i} (1-p)^{sN - \sum_i y_i} \right] \left[\frac{M!}{N!(M-N)!} \psi^N (1-\psi)^{M-N} \right], \end{aligned} \quad (13.23)$$

where r is the number of sites with at least one detection ($y_i > 0$).

The factorization in (13.23) motivates the data augmentation algorithm of the authors. The term in the first square bracket of (13.23) is the likelihood for closed population model M_0 . The implication is that we can use the model in (13.22) to fit capture-recapture model M_0 .

For capture-recapture data, the authors referred to M as the size of a superpopulation. This is an abstract construct that provides an upper bound on the abundance N made up of individuals who could have been included in the population of interest N . That is, we are augmenting the r observed individuals, with an additional $M - r$ individuals that may be in the population but were not observed in the study ($y_i = 0$). For capture-recapture, the quantity M is set by the analyst as discussed above in Sect. 13.10.6.

The second square bracket in (13.23) is the prior distribution for N implied through use of the model in (13.22), namely,

$$N \sim \text{Binomial}(M, \psi).$$

The parameter ψ also has a prior $f(\psi)$ so that the marginal prior for N is

$$f(N|M) = \int f(N|\psi, M) f(\psi) d\psi. \quad (13.24)$$

The first thing to note is that the prior for N depends on superpopulation M . Different choices for $f(\psi)$ have been suggested. Royle et al. (2007b) used a uniform prior for ψ that implies a discrete uniform prior distribution for N truncated above by M . Link (2013) pointed out that the improper prior $f(\psi) \propto \psi^{-1}$ leads to a discrete scale prior $f(N|M) \propto N^{-1}$, again truncated above by M .

Royle et al. (2007b) extended the model to allow for heterogeneity in detection. The model in (13.22) becomes

$$\begin{aligned} z_i &\sim \text{Bern}(\psi), \quad i = 1, \dots, M, \\ y_i | z_i &\sim \text{Binomial}(s, p_i z_i), \quad i = 1, \dots, M, \quad j = 1, \dots, s, \\ p_i &\sim g(p|\theta). \end{aligned} \quad (13.25)$$

This model has two changes from (13.22). The first is that the model for y_i now depends on p_i , the capture probability for individual i . The second is that we include $g(p|\theta)$ as a model for p_i .

The authors argued that the models in (13.22) and (13.25) should be referred to as parameter-expanded data augmentation (PX-DA). They gave two reasons. The first is to distinguish their approach from other uses of the term data augmentation in capture-recapture modeling, particularly the work of Schofield and Barker (2008) and Wright et al. (2009).³ The second is the recognition that the model in (13.22) includes an additional parameter (ψ) that is not present in the likelihood for model M_0 .

13.10.8 Comparing Approaches That Bound Abundance

A benefit of bounding abundance by M is that we can use general software such as BUGS and JAGS for model fitting. This is the case both for the approach of Durban and Elston (2005) and the PX-DA specification of Royle et al. (2007b). Durban and Elston (2005) referred to their approach as model selection. This is because their formulation, which bounds abundance by M , is a special case of the general model selection approach of Carlin and Chib (1995). Royle et al. (2007b) claimed that their PX-DA scheme overcomes the problem of moving from one model to another. Link et al. (2010, 210) disagreed, asserting that PX-DA is equivalent to an RJMCMC approach. In response, Royle et al. (2012) emphasized that their approach is not an MCMC algorithm (as RJMCMC is) but rather a re-parameterization of the model that allows for simplified MCMC sampling. They argued that the value of their approach is that the re-parameterization invokes a naturally simple MCMC fitting approach, not because that fitting approach corresponds to a particular RJMCMC implementation.

Royle et al. (2007b) were critical of Durban and Elston (2005) and stated that their use of pseudo-priors is “as much an art as a science” and claimed that “the trans-dimensional Gibbs sampler lacks flexibility and generality.” Royle et al.

³ Schofield and Barker (2008) considered open population models where the times of entry and exit to the population are included as latent variables. Wright et al. (2009) allowed for uncertain identification and included the true identities as a latent variable.

(2012) further claimed that “To our knowledge, RJMCMC algorithms suitable for capture-recapture models have not been implemented in popular software programs for Bayesian analysis.”

Schofield and Barker (2014) considered the various approaches for modeling closed population data with heterogeneity in detection, focusing on the PX-DA algorithm of Royle et al. (2007b), and the approach of Durban and Elston (2005). They argued that both approaches can be viewed from a model selection lens; for Royle et al. (2007b), the models are indexed by the variable \mathbf{z} , whereas for Durban and Elston (2005), they are indexed by the value of N . Schofield and Barker (2014) studied three differences between the two approaches: (i) specification of combinatorial terms, (ii) MCMC algorithms for model fitting, and (iii) the parameterization of the models. We consider each of these in turn.

The marginal model for M_h is given in (13.15) and includes the combinatorial term

$$\frac{N!}{(N - r)! f_1! \dots f_s!}. \quad (13.26)$$

This gives the number of ways in which we could have observed f_1 individuals with one capture, f_2 individuals with two captures, \dots , f_s individuals with s captures, out of a population of N individuals. The PX-DA model in (13.25) does not include this combinatorial term. The apparent difference is reconciled because the counting is done implicitly while sampling from the posterior distribution with MCMC. Schofield and Barker (2014) showed that the implied combinatorial is proportional to (13.26). Neither the approach of Durban and Elston (2005) or Royle et al. (2007b) involved any change of dimension in their model, due to the upper bound M . Standard MCMC algorithms can be used to fit both models.

The most obvious difference between the approaches is the parameterization of the model. The approach of Royle et al. (2007b) included the latent variable \mathbf{z} and has abundance N as a derived quantity that does not appear as a parameter in the model. The PX-DA model assumes that when an individual is not in the population ($z_i = 0$), then the detection model is $p_i \sim g(p|\theta)$. Durban and Elston (2005) suggested that a pseudo-prior be used when an individual is not in the population ($i > N$). Our experience is that the approach of Basu (1998) that sets the detection to be $p_i \sim g(p|\theta)$, $i > N$ in a matter similar to that of Royle et al. (2007b), appears to work well in practice and avoids the need for pilot runs of the algorithm.

Schofield and Barker (2014) argued that there is no inherent advantage to modeling in terms of \mathbf{z} as in Royle et al. (2007b), particularly if we wish to model the abundance N . Their motivating example involved closed population capture-recapture experiments replicated in time. At time $t \in \{1, \dots, T\}$, the capture-recapture data can be used to estimate abundance N_t . Interest is not in the values N_1, \dots, N_T themselves but in a model that describes the change in abundance through time. Such a model is difficult to construct using the PX-DA approach. The values N_1, \dots, N_T are not included in their model, making it difficult to specify a hierarchical model in this context. They showed that extending the

approach of Durban and Elston (2005) to include a hierarchical model for N_t is straightforward.

13.10.9 Reversible Jump Markov Chain Monte Carlo

The remaining approach we consider includes \mathbf{p}_N as a latent variable in the model. An algorithm is implemented that is designed to work when one variable (N) defines the dimensionality of another variable (\mathbf{p}_N).

The reversible jump Markov chain Monte Carlo (RJMCMC) algorithm is an extension to the Metropolis-Hastings algorithm that allows one variable to define the dimensionality of another variable (Green 1995). It is a general algorithm, most commonly used for model selection, e.g., King and Brooks (2001). As we explain below, the key to the RJMCMC algorithm is the introduction of augmenting variables to “match the dimension” when updating the variable that defines the dimension of other variables in the model. We outline how to implement the RJMCMC algorithm in the special case of the capture-recapture model M_h . For a general outline of the RJMCMC algorithm, we refer the reader to Green (1995) and Seber and Schofield (2019).

Model M_h is given by a hierarchical model. The model for the capture histories conditional on heterogeneous detection probabilities is

$$g(\mathbf{y}|\mathbf{p}_N, N) = \frac{N!}{(N-r)!f_1!\dots f_s!} \prod_{i=1}^N p_i^{y_i} (1-p_i)^{k-y_i},$$

where $\mathbf{y} = (y_1, \dots, y_N)'$. We leave the model for the detection probabilities, $g(p|\theta)$, as unspecified, other than requiring that p_i is independent of p_j , $j \neq i$. As our focus is on the updating of N , we treat θ as if it were known. To complete the model specification, a prior for N , denoted $f(N)$, is required.

In iteration t of an MCMC algorithm, we have the values

$$N^{(t)}, p_1^{(t)}, \dots, p_N^{(t)}.$$

If we were to update N using a Metropolis-Hastings algorithm, we would first draw a candidate value N^* from a proposal distribution $p(N^*|N^{(t)})$. If $N^* > N^{(t)}$, there are two related problems:

1. The “proposed model” (where $N = N^*$) has more variables (p_1, \dots, p_{N^*}) than the “current model” (where $N = N^{(t)}$) with ($p_1, \dots, p_{N^{(t)}}$).
2. The variables $p_{N^{(t)}+1}, \dots, p_{N^*}$ are not defined in the “current model.”

The opposite problems occur if $N^* < N^{(t)}$:

1. The proposed model (where $N = N^*$) has fewer variables (p_1, \dots, p_{N^*}) than the current model (where $N = N^{(t)}$) with ($p_1, \dots, p_{N^{(t)}}$).
2. The variables $p_{N^*+1}, \dots, p_{N^{(t)}}$ do not feature in the proposed model.

The key to the RJMCMC algorithm is that augmenting variables are defined so that the proposed model has the same number of variables as the current model.

When $N^* > N^{(t)}$ we generate $N^* - N^{(t)}$ augmenting variables that function as the “extra” detection variables p_{N+1}, \dots, p_{N^*} . A convenient choice for the distribution we use to generate the augmenting variables is $g(p|\theta)$. This is convenient, not only because it should generate reasonable values but also because it leads to the cancellation of terms in the acceptance probability. When $N^* < N^{(t)}$ we take as augmenting variables those “leftover,” i.e., $p_{N^*+1}, \dots, p_{N^{(t)}}$.

The proposal N^* is accepted with probability $\min(1, \alpha)$, with

$$\alpha = \frac{N^*!}{(N^* - r)!} \frac{(N^{(t)} - r)!}{N^{(t)}!} \frac{K_1}{K_2} \frac{f(N^*)}{f(N^{(t)})},$$

where

$$K_1 = \begin{cases} \prod_{i=N^{(t)}+1}^{N^*} (1 - p_i)^s, & \text{when } N^* > N^{(t)} \\ 1, & \text{when } N^* < N^{(t)} \end{cases}$$

$$K_2 = \begin{cases} 1, & \text{when } N^* > N^{(t)} \\ \prod_{i=N^*+1}^{N^{(t)}} (1 - p_i)^s, & \text{when } N^* < N^{(t)}. \end{cases}$$

In the description above, we have used $g(p|\theta)$ as the distribution we used to generate the augmenting variable. This assumption can be relaxed and other generating distributions considered (the full generality of RJMCMC allows for other extensions that we do not consider). One possible choice is to use $g(p_i|y_i = 0, \theta)$ as the augmenting distribution. While this choice may lead to a more efficient algorithm, it adds complexity as the conditional distribution is unlikely to be algebraically tractable.

RJMCMC algorithms such as that outlined above have been used to update abundance in closed populations with heterogeneity (Fienberg et al. 1999; King & Brooks 2008; Arnold et al. 2010).

13.11 Priors

The sections above focused on methodological developments in the specification and fitting of Bayesian models for abundance estimation. Here we focus on the choice of prior distribution. We first review three classes of prior distribution.

The first is an informative prior distribution. This is a probabilistic description of the knowledge about θ before an experiment is undertaken. The resulting posterior distribution updates our prior knowledge with information about θ obtained from the data. An informative prior is particularly useful in situations where the posterior distribution is used to inform decision-making. In Sect. 13.11.3, we present an example where informative priors were used.

The second class is that of non-informative priors, also referred to as objective or vague priors. This is where we desire the prior distribution to represent ignorance about the value of θ . This can be difficult. As noted by Link (2014):

Difficulties arise when Bayesian methods are extended to situations where little or nothing is known about parameters. Unfortunately, there is no universally accepted procedure for assigning noninformative priors. There is no perfectly motivated procedure. Some Bayesian statisticians reject the entire notion of objective Bayesian inference on these grounds, with its vague and sometimes improper priors.

Consider an example. Suppose that we have an observation y that we assume is a realization of a binomial experiment with known index N and unknown parameter π ,

$$y \sim \text{Binomial}(N, \pi).$$

It seems reasonable that a uniform prior distribution would represent ignorance about π , as it has the same density for every possible parameter value $\pi \in (0, 1)$. A perfectly legitimate alternative is to model on the logit scale, so that

$$y \sim \text{Binomial}(N, \text{logit}^{-1}(\eta)).$$

Using the same logic from above, we reason that a uniform prior on η represents ignorance. Two difficulties arise. The first is that a uniform prior on η is improper, as the parameter space for η is unbounded. The second is that a uniform prior for η and a uniform prior for π are not equivalent. Adopting uniform priors to reflect ignorance can lead to priors that are not invariant to transformation.

Several approaches have been developed for the specification of non-informative prior distributions, including approaches that ensure resulting priors are invariant to transformation. An overview of non-informative prior specification is given by Kass and Wasserman (1996).

The final class we consider are weakly informative priors. As described by Gelman (2006), a weakly informative prior is proper but defined so that it provides less information than whatever actual prior knowledge is available. This definition is vague, with weakly informative priors often used to help regularize model fitting. For example, Gelman et al. (2008) suggested a weakly informative prior for logistic regression. They built and justified their prior choice based on the reasoning that it is unlikely that the difference in logit probability will exceed 5 when comparing a predictor value one standard deviation above the mean to a value one standard

deviation below the mean.⁴ An important feature of this weakly informative prior is that it belongs to the t family of distributions, a deliberate choice due to heavy tails that allow for robust inference.

13.11.1 Non-informative Priors for Capture-Recapture

Bolfarine et al. (1992) considered the estimation of N in model M_t . They showed that when there are no recaptures, maximum likelihood estimation gives an infinite estimate of abundance. Similarly, other likelihood-based approaches (profile likelihood and adjusted orthogonal profile likelihood) also give infinite estimates of abundance if there are no recaptures and other conditions are satisfied.

They re-parameterized in terms of N and μ_1, \dots, μ_s , where $\mu_j = p_j N$, $j = 1, \dots, s$, and specify a joint prior

$$f(\mu_1, \dots, \mu_s, N) = \left(\frac{1}{N}\right)^s \frac{1}{(N+1)^2}$$

The authors referred to this as providing minimal information, although they offered no justification for its choice. They summarized the posterior in terms of the mode of the marginal posterior distribution of N , a Bayes estimator that is always finite. We note that in the case where there are no (or very few) recaptures, the prior is the dominant source of information about abundance and its choice is important.

Madigan and York (1997) suggested a non-informative prior proposed by Rissanen (1983), namely,

$$f(N) \propto 2^{-\log^* N},$$

where $\log^* N$ is the sum of the positive terms in $\{\log_2(N), \log_2(\log_2(N)), \dots\}$. This prior is complex and to our knowledge has not been widely implemented.

Wang et al. (2007) explored the choice of prior distribution for model M_t when the sample size is small to moderate. For capture probabilities, they explored the beta class of priors

$$f(p_j) \propto p_j^{\alpha_j-1} (1-p_j)^{\beta_j-1}, \quad j = 1, \dots, s.$$

This class includes the uniform prior ($\alpha_j = \beta_j = 1$) and the Jeffreys prior ($\alpha_j = \beta_j = 0.5$), both of which had previously been used in capture-recapture models. For abundance, they explored a class of priors

$$f(N) \propto \frac{1}{N^r}.$$

⁴ Noted that Gelman et al. (2008) defined their prior with the assumption that the input variables have mean 0 and standard deviation 0.5.

This includes the uniform prior ($r = 0$), and the Jeffreys prior ($r = 1$). The authors established the conditions under which the posterior distribution is proper, and they looked at more detail at four prior combinations:

1. Jeffreys prior for p_1, \dots, p_s , uniform prior for N ;
2. Jeffreys prior for p_1, \dots, p_s , Jeffreys prior for N .
3. uniform prior for p_1, \dots, p_s , uniform prior for N ;
4. uniform prior for p_1, \dots, p_s , Jeffreys prior for N ;

Theoretical arguments were used to show that among these four choices, $E[N|\text{data}]$, the posterior expectation of abundance is ordered as listed above (from largest to smallest).

A simulation study was conducted to explore the properties of these priors compared to MLE. The simulation settings varied the capture probabilities (low, medium, high), the number of occasions (from $s = 3$ to $s = 10$), and the abundance (from $N = 50$ to $N = 800$). The simulation results were difficult to summarize; the best-performing prior depended on the criterion (bias, MSE, coverage), and the simulation settings, with the number of sampling occasions being particularly important. See Wang et al. (2007) for more details.

Xu et al. (2014) constructed priors for model M_t using the reference prior approach outlined in Berger et al. (2012). They assumed that the prior for p_1, \dots, p_s followed a beta distribution with parameters a and b . Provided $s \geq 2$ and $0 < a \leq 1$, the reference prior for N was given by

$$f(N) \propto \frac{(\log N)^{\frac{s+\lfloor b \rfloor - 1}{2}}}{(N+1)^{\frac{b}{2}}},$$

where $\lfloor b \rfloor$ is the largest integer less than or equal to b . Properity of the posterior distribution of N was established, and the prior was assessed with simulated and real data. See Xu et al. for details.

Bord et al. (2018) suggested caution when choosing prior distributions for abundance estimation in removal models. They considered the class of prior

$$f(N, p) \propto \frac{1}{N^c} p^{a-1} (1-p)^{b-1},$$

and claimed that the posterior distribution is proper if and only if $a + c > 1$ and that the posterior mean of N is finite if and only if $a + c > 2$.

13.11.2 Bayes Ancillarity

Bayes ancillarity was defined by Severini (1995), which we now describe. Suppose there are two data sources x and y and two parameters θ and ψ , where the model

can be expressed as

$$f(x|\theta, \psi) f(y|\theta).$$

The data x are said to be Bayes ancillary for θ given the prior $f(\psi|\theta)$ if

$$\int f(x|\theta, \psi) f(\psi|\theta) d\psi$$

does not involve θ for any possible value of x . Intuitively, the data x provide no information about θ after we have accounted for ψ . There have been several instances in the capture-recapture literature where priors have been chosen because of this property.

Garthwaite et al. (1995) considered prior choices for model M_t and in particular for the capture probabilities p_1, \dots, p_s . Recall that model M_t is

$$f(\mathbf{a}_\Omega | \mathbf{p}, N) = \underbrace{\frac{N!}{\prod_\omega a_\omega!(N-r)!} \prod_{i=1}^s \binom{N}{n_i}^{-1}}_{f(\mathbf{a}_\Omega | \mathbf{n}, N)} \underbrace{\binom{N}{n_i} p_i^{n_i} q_i^{N-n_i}}_{f(\mathbf{n} | p, N)}.$$

They argued that the choice of prior should be such that the model for sample size n_i should have little effect on the marginal prior distribution for N . In other words, the majority of the information about N should come through Darroch's model B, given by $f(\mathbf{a}_\Omega | \mathbf{n}, N)$ in the first term above. The authors justified this approach based on the Fisher information matrix showing that n_1, \dots, n_s provides little information about N . It also makes intuitive sense as the model for sample size has s observations (n_1, \dots, n_s) and s parameters (p_1, \dots, p_s) that are not informed from other data.

The prior

$$f(p_1, \dots, p_s) \propto \prod_{i=1}^s p_i^{-1} \tag{13.27}$$

ensures that the sample sizes do not influence the posterior distribution for N . This is an example of Bayes ancillarity. That is, the observations n_1, \dots, n_s are Bayes ancillary for N given the prior in (13.27). This prior is an extension of that discussed by Castledine (1981) for two-sample problems.

The authors derived properties of the marginal posterior distribution for N when assuming the prior in (13.27). They examined how the posterior mean compared to two classical estimators (the MLE and a bias-corrected estimator) in four examples and showed that there can be substantial differences between the various estimators. As the authors noted, the differences are most notable when there is little information and the posterior mass function is skewed. This is best seen visually in Fig. 13.2. As with the example in Roberts (1967) in Sect. 13.3.1, the

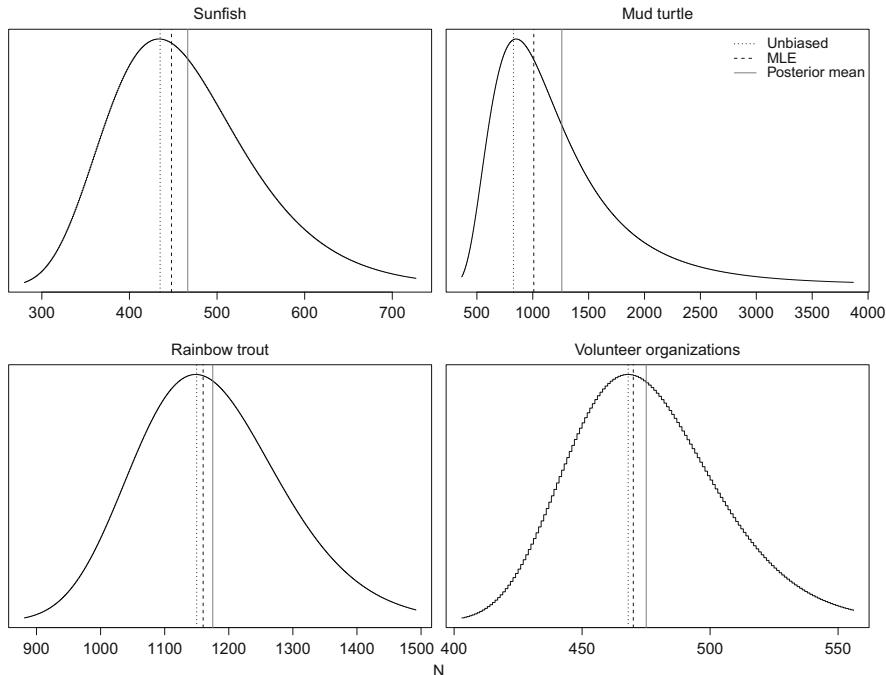


Fig. 13.2 The posterior mass function of N for the four examples examined by Garthwaite et al. (1995) with a scale prior for N ($f(N) \propto N^{-1}$). The three lines represent the estimators compared by them

difference between the Bayesian and classical estimators is largely due to the choice of posterior summary (determined by the choice of loss function). If we were to summarize the posterior using the mode or median, the Bayesian estimates would have been closer to the classical estimates.

Another example of a Bayes ancillary prior was given by Barker et al. (2014) and Schofield et al. (2018) in the context of continuous-time capture-recapture. Data summaries included c_i , the number of times individual i is caught between time 0 and the endpoint τ ; and T , the total number of captures across all individuals. Denoting time by t , Barker et al. (2014) used a Poisson process model with intensity $\lambda_i(t)$ and total intensity

$$v_i = \int_0^\tau \lambda_i(u) du.$$

Building on simpler models, Barker et al. (2014) allowed v_i , $i = 1, \dots, N$ to be modelled as i.i.d. gamma random variables with parameters α and β . The resulting model can be partitioned as

$$f(\mathbf{c}|T, N, \alpha) f(T|N, \alpha, \beta),$$

with two components: a model for the counts c_1, \dots, c_N conditional on the total number of observations and a model for the total number of observations. The conditional model for T has a negative-binomial distribution with parameters $N\alpha$ and β . The authors suggested the prior $f(\beta) \propto \beta^{-1}$. This ensures that

$$\int f(T|N, \alpha, \beta) f(\beta) d\beta$$

does not depend on N , so that the data T do not influence the posterior distribution for N . See Barker et al. (2014) for further details.

Related to Bayes ancillarity is the work of Link (2013) investigating prior specification for N , exploring how uniform priors can have undesirable properties when used for abundance parameters in capture-recapture models. This was motivated by heavy use of uniform prior distributions for N with data augmentation (see Sect. 13.10.7 for more details). A simplification of capture-recapture was used to illustrate the problem. Suppose that we have data

$$x \sim \text{Binomial}(n, p), \quad (13.28)$$

where the index n is observed, with unknown p . In addition we have observations y_1, \dots, y_s

$$y_j \sim \text{Binomial}(N_j, p), \quad j = 1, \dots, s, \quad (13.29)$$

where N_1, \dots, N_s are unknown. Link (2013) asserted that it is common sense that all knowledge about p should come from the data x , as we learn virtually nothing about p from the observations y_1, \dots, y_s that have unknown indices N_1, \dots, N_s .

Sequential updating simplifies the problem further (see Appendix A.9.1). If we assume a beta prior for p with parameters α and β , we need to only consider observations y_1, \dots, y_s where p has a beta $a = x + \alpha$, $b = n - x + \beta$ prior. If N_1, \dots, N_s have independent uniform prior distributions, Link (2013) showed that the posterior distribution of p is

$$f(p|x, y_1, \dots, y_s) \propto p^{a-s-1} (1-p)^{b-1}.$$

Of interest:

- This is a beta distribution provided $a > s$.
- The observations y_1, \dots, y_s influence estimation of detection probability through the index s .
- The information content of a beta random variable is often given by the sum of its parameters $\alpha + \beta$. The implication is that including the data y_1, \dots, y_s decreases the information about p .

Link (2013) showed that if we use the scale prior

$$f(N_j) \propto N_j^{-1}, \quad j = 1, \dots, s,$$

then the marginal posterior of p is

$$f(p|x, y_1, \dots, y_s) \propto p^{a-1}(1-p)^{b-1},$$

so that $f(p|x, y_1, \dots, y_s) = f(p|x)$. That is, the observations y_1, \dots, y_s do not influence the estimation of p , as we would expect.

He showed that undesirable properties are also evident with models of the form (13.28) and (13.29) when using maximum likelihood estimation. If $x = 6, n = 20$, then estimating p from x alone leads to an estimate of $\hat{p} = 0.3$. If $s = 1$ and we observe $y_1 = 30$, the intuitive estimate of N_1 is $30/0.3 = 100$, but this is not the MLE. The MLE of p is $\hat{p} = 0.3077$ and the MLE of N_1 is 97. If $s = 2$ with $y_1 = y_2 = 30$, then the MLE for p changes to 0.3173, and the MLE for N_1 is now 94. If $s = 3$ with $y_1 = y_2 = y_3 = 30$, the MLEs are $\hat{p} = 0.3243$ and $\hat{N}_1 = 92$.

Schofield et al. (2023) considered a similar setup to that of Link (2013). They explored the properties of various estimators of abundance, motivated by a mark-recapture distance sampling study where data were collected across multiple transects. Treating each transect as a distinct subpopulation, they highlighted similar problems with maximum likelihood estimation to that showed by Link above. Specifically, if there is a large number of subpopulations, there can be considerable negative bias and large MSE. They also showed that these problems do not occur in Bayesian methods provided the scale prior is adopted. They also considered other estimators and modelling strategies, and further details are given in their article.

Another prior for N was suggested by Villa and Walker (2014) in a response to Link (2013). They objected to the selection of prior based on performance alone, arguing that any prior used also needs solid motivation. They reasoned that motivation for a Jeffreys (scale) prior on N was “obsolete” as invariance to transformation has no meaning on a discrete parameter space. Also, they argued that an objective prior needs to have an applied meaning. They subsequently derived a prior by assigning worth to each possible value N in the parameter space, as measured by what is lost if that parameter value is removed. This leads to the prior

$$f(N) \propto \frac{1}{(N+1)(1-p)} \exp \left(\sum_{x=0}^N \log(N+1-x) \binom{N}{x} p^x (1-p)^{N-x} \right) - 1.$$

They considered the snowshoe hares dataset and compared the posterior found with their prior, to the posterior when a scale prior is used. The distributions were similar, and their prior led to a posterior median⁵ of $N = 74.6$ and 95% credible interval

⁵ The posterior summaries are those given by Link (2014).

(69.3, 82.8). The scale prior leads to a posterior median of $N = 74.8$ and 95% credible interval (69.4, 83.2).

In reply, Link (2014) noted that there is debate as to how best to choose non-informative prior distributions, arguing that prior choice based on operational performance is reasonable. The author advocated for simplicity, noting that the prior of Villa and Walker (2014) was complex and closely approximated the scale prior, provided there is good information about p .

13.11.3 *Informative Priors*

Our experience is that non-informative priors remain the standard approach in applied Bayesian modelling. If well specified, a non-informative prior will “let the data speak” so that the posterior distribution is dominated by the likelihood.

Informative priors tend to be less common in practice. They quantify our belief about parameters in the model based on expertise and knowledge about the system under study. Many of the early papers we describe above discuss the use of informative priors, for example, Roberts (1967), Castledine (1981), and Smith (1991), although the focus of their work was on the methodological development of the model. While estimates were often compared for a range of prior choices, there was typically no strong justification for any one choice.

King et al. (2005) provided an example where a feature of their work is the use and justification of an informative prior distribution. Their interest was in estimating the abundance of injecting drug users (IDUs) in Scotland from four medical lists. The data included three covariates: region, sex, and age group; each is categorical with two levels. With four sources of data and three dichotomous covariates, the data can be represented as a 2^7 contingency table (with missing cells for the number of individuals on none of the four lists within each covariate combination). Log-linear models with main effects and one-way interactions were considered. The goal was to estimate the abundance taking into account model uncertainty.

They considered informative prior distributions. External information included the drug-related death rate of Europe’s IDUs and the European male-to-female IDU ratio. Prior distributions were chosen to reflect the uncertainty of each information source. They fitted the model with both non-informative and informative priors, comparing the results. When informative priors were used, the three models with the largest posterior model probability all had abundance estimates between 25,000 and 28,000. In contrast, when vague priors were used, the second most probable model had an estimate of less than 17,000 individuals. Further details are given in their article.

13.11.4 Weakly Informative Priors

An example of weakly informative priors was given by Schofield et al. (2018) in the context of modeling continuous-time capture-recapture data. Considering the same model discussed in Sect. 13.11.2, they assumed that v_i , $i = 1, \dots, N$ are i.i.d. variables with a gamma distribution having parameters ρ^{-1} and $\pi/(1 - \pi)$. They marginalized over the latent variables v_1, \dots, v_N .

Their choice of parameterization in the model for v was deliberate. Replacing β with $\pi/(1 - \pi)$ ensured that the full conditional distribution of π is in known form. This has computational benefits when model fitting. The parameter ρ is the squared coefficient of variation for the model of v_1, \dots, v_N . The quantity $\rho = 0$ represents no variation in v_1, \dots, v_N , and the model is equivalent to model M_t. Parameter values $\rho > 1$ resulted in the gamma distribution for v having mode 0, implying a highly heterogeneous population; the majority of individuals were essentially uncatchable. An exponential prior with parameter $\theta = \log(100) = 4.6$ was proposed as a default weakly informative prior. This was chosen so that the prior probability of $\rho > 1$ is 0.01. This prior is compared to alternatives $\theta = \log(10000) = 9.2$ and $\theta = 50$. Full details are given in their article.

13.11.5 Hierarchical Models/Priors

A hierarchical model is one where the parameters themselves are assumed to be random variables and modelled as realizations of a probability distribution. An example in statistics is the linear mixed model. Hierarchical modeling is a natural extension when using a Bayesian approach as all unknowns are modelled as random variables. To assist in our description, we outline a generic hierarchical model where we have data y that are modelled in terms of parameters θ that are themselves modelled in terms of parameters ξ . This model can be represented as

$$f(y|\theta)f(\theta|\xi),$$

where θ are latent random variables. The specification is completed with a prior for ξ , denoted $f(\xi)$.

To fit the model, one option is to integrate over θ so that we are working with the marginal likelihood,

$$f(y|\xi) = \int f(y|\theta)f(\theta|\xi) d\theta.$$

Evaluating this integral can be difficult. Another option is to make use of a model-fitting strategy that allows for latent variables. One such approach is the EM algorithm that can be used when fitting using maximum likelihood; see Seber

and Schofield (2019) for further information. When using Bayesian methods, data augmentation and MCMC can be used, as discussed for abundance estimation in Sect. 13.10. This approach keeps θ in the model, and the MCMC sampler generates samples from the posterior distribution

$$f(\theta, \xi | y) \propto f(y|\theta) f(\theta|\xi) f(\xi).$$

The first examples of a hierarchical model for capture-recapture were Castledine (1981) and Smith (1991), who both considered hierarchical specifications for p in model M_t . Neither author had the benefit of modern computational algorithms, and due to the difficulty in marginalizing over p , neither was able to find the posterior distribution for N . Castledine (1981) assumed that $\text{logit}(p_1), \dots, \text{logit}(p_s)$ have a common normal distribution with unknown mean. They abandoned their approach as they were only able to specify the mode of the joint posterior $f(\mathbf{p}_s, N | \mathbf{a}_\Omega)$ and were unable to approximate the posterior $f(N | \mathbf{a}_\Omega)$.

Among other models, Smith (1991) assumed a beta model for p_1, \dots, p_s with unknown parameters α and β and found a posterior distribution for N using empirical Bayes. Here estimates $\hat{\alpha}$ and $\hat{\beta}$ were found by maximizing

$$f(\mathbf{a}_\Omega | \alpha, \beta) = \sum_{N=r}^{\infty} f(\mathbf{a}_\Omega | N, \alpha, \beta) f(N),$$

where $f(\mathbf{a}_\Omega | N, \alpha, \beta)$ is the marginal likelihood in (13.5), now indexed by α and β to reflect that they are considered unknown parameters in this setting. The empirical Bayes posterior is $f(N | \mathbf{a}_\Omega)$ in (13.6) with $\alpha = \hat{\alpha}$ and $\beta = \hat{\beta}$. Smith (1991) approximated the true posterior by accounting for the variation in the estimates $\hat{\alpha}$ and $\hat{\beta}$ and used two approximations. The first evaluated the variance of $(\hat{\alpha}, \hat{\beta})$ conditional on N , while the second used a Laplace approximation to marginalize over the parameters.

Royle and Link (2002) provided an early overview of hierarchical models for an ecological audience. While not fitting a closed population model, they used MCMC to fit a band recovery model, where both survival and recovery rates are modelled as a logit normal hierarchical model. An excellent early example of how hierarchical models can incorporate knowledge of the study system was given by Johnson and Hoeting (2003). While also not a closed population model, they modelled survival parameters as an autoregressive process to account for the dependence among demographic parameters from one period to another.

With increased availability of modern software for Bayesian computing, the use of hierarchical modelling is now commonplace in statistical ecology and applied statistics in general. As such, we do not attempt to review its use but refer the interested reader to Cressie et al. (2009), Link and Barker (2010), King et al. (2010), and Kéry and Royle (2020).

13.12 Bayes in the Twenty-First Century

Throughout the twentieth century, the Bayesian community was small, and the methodological contributions we have highlighted tended to lag the equivalent frequentist development by some years. The first two decades of the twenty-first century have seen the popularity of Bayesian inference rapidly increase.

The catalyst for increased adoption of Bayesian methods was primarily due to computational advances, particularly the development of MCMC. The introduction of general-purpose Bayesian software made model fitting accessible to applied scientists. An early example was the BUGS family of software (Lunn et al. 2000). More recent packages include JAGS (Plummer 2003), PyMC (Salvatier et al. 2016), Stan (Carpenter et al. 2017), and NIMBLE (de Valpine et al. 2017). We expect the number and variety of general-purpose Bayesian software options to continue to grow in the future.

The increased popularity of Bayesian inference has resulted in many emerging areas of research having either seen (near) simultaneous development of frequentist and Bayesian methodology or having been led by those using a Bayesian approach. We highlight some of these below. Our descriptions are brief, as the methodological developments are listed elsewhere in this book.

The seminal work of MacKenzie et al. (2002) was the first to include an occupancy parameter directly into the likelihood (MacKenzie et al. 2004). It was motivated by data for a single species with multiple binary observations (detected/not detected) across several sites. Initial model development was conducted in a maximum likelihood framework (MacKenzie et al. 2002), with Bayesian models discussed shortly thereafter by MacKenzie et al. (2004). Within three years of the initial contribution, Dorazio and Royle (2005a) not only provided a Bayesian analogue of the model of MacKenzie et al. (2002) but extended it to allow for multiple species. Letting ψ_{ij} denote the occupancy probability of species i at site j and θ_{ij} be the detection probability for species i at site j , Dorazio and Royle (2005a) included Rasch-type models

$$\begin{aligned}\psi_{ij} &= u_i + \alpha_j \\ \theta_{ij} &= v_i + \beta_j,\end{aligned}$$

where the i subscript denoted a species effect and a j subscript denoted site effects. An important feature was a hierarchical model for (u_i, v_i) that included a correlation between the species-level occupancy and detection probabilities. As we note in Sect. 13.10.7, the relationship between the occupancy and capture-recapture models was the inspiration behind the PA-DX algorithm of Royle et al. (2007b). Bayesian analysis has continued to feature in the development of occupancy models. A review of multi-species occupancy models found that around 65% of studies used Bayesian approaches (Devarajan et al. 2020). Other Bayesian methodological developments included the use of variational approximations for model fitting (Clark et al.

2016) and the development of Gibbs sampling approaches using Pólya-gamma data augmentation (Clark & Altweig 2019).

An important development in the modelling of capture-recapture data was the ability to include covariate information (Pollock et al. 1984). It allowed one to explain variation in parameters (say, according to weather conditions), account for individual differences in the capture process, and potentially test various scientific hypotheses. The emergence of spatial capture-recapture (SCR) models went one step further; they allowed for an individual covariate, location, which is unobserved for all individuals in the population. These locations (often referred to as activity centers) are assumed to be generated according to a Poisson process. The rate parameter of the Poisson process is the density (expected abundance per unit area) and is the parameter of interest. The model was first considered by Efford (2004), who used inverse prediction to estimate density. There was near-simultaneous development of likelihood and Bayesian models, the former led by Borchers and Efford (2008) and the latter by Royle and Young (2008).

The two estimation approaches typically differ in terms of how the unobserved individual variable is handled. Model fitting with maximum likelihood requires multiple (numerical) integrals to marginalize the activity centers, while in Bayesian model fitting, they are usually included using data augmentation. This makes model specification easier and helps explain why extensions to the model, such as allowing for open populations, were developed earlier for Bayesian models (Gardner et al. 2010) than for maximum likelihood approaches (Glennie et al. 2019; Efford & Schofield 2020). That said, recent work explores the advantages of marginalization when fitting using MCMC (Turek et al. 2021).

A key assumption in the modelling of capture-recapture data is the correct identification of tagged individuals. This may not be a reasonable assumption in many contexts, for example, with tag loss (Conn et al. 2004). Technological advances have only amplified the problem, e.g., identification errors can occur with genetic tagging (Lukacs and Burnham, 2005; Wright et al. 2019) and camera traps (McClintock et al. 2013; Bonner and Holmberg 2013). Bayesian approaches have been at the forefront of models to tackle this.

One approach is to consider the observed capture histories as an error-prone version of the “true” capture histories (histories that record when and how an error was made). Early papers used quasi-likelihood for estimation, e.g., Lee (2002) and Sutherland and Schwarz (2005). Bayesian models were developed that include the true records as latent variables (Link & Barker 2010; McClintock et al. 2013; Bonner & Holmberg 2013; Schofield & Bonner 2015; and Bonner et al. 2016). A maximum likelihood approach that used a saddlepoint approximation has recently been developed (Zhang et al. 2019, 2021, 2022).

An alternative approach is to model the misidentification process at the observation level. This is a natural approach in many situations. Often, there is information at the observation level that informs the probability of whether an individual is misidentified or not. For example, Wright et al. (2009) used information from multiple genotyping sequences to inform the probability of being misidentified due to allelic dropout. They fitted the model using a Bayesian approach, with the true

identity of each observation included in the model as a latent variable using data augmentation. Augustine et al. (2018, 2020) considered various error processes and extended the model of Wright et al. (2009) to allow for spatial information to be used; if two observations are located close, they are more likely to belong to the same individual than two observations a long way apart. To our knowledge, there are currently no maximum likelihood approaches that account for misidentification at the observation level.

The common theme in the description of Bayesian approaches above is the use of data augmentation to include latent variables in the model. The ease with which complex models can be specified when using data augmentation leads to rapid model development, provided any computational inefficiencies introduced are not too extreme. Once models are developed, attention then turns to efficiency, scalability, and generalizability.

13.13 Conclusion

This chapter outlines the development of Bayesian methods for estimating abundance in closed population capture-recapture models. We describe model development prior to the use of Markov chain Monte Carlo (MCMC), which relied on mathematical derivation of the posterior distribution for relatively simple models. The rapid development of Bayesian models after the introduction of MCMC is reviewed and includes models that allow for behavioral effects due to previous capture, individual heterogeneity, and dependence between capture occasions (or list dependence).

We consider in detail Bayesian estimation of abundance in the presence of individual heterogeneity. It is a difficult problem due to one parameter (abundance) defining the dimensionality of another parameter (capture probability). Prior choice for closed population capture-recapture models is discussed. This includes priors that attempt to include little to no prior information (non-informative), priors that attempt to include external information (informative), and priors that incorporate a weaker form of the information available (weakly informative).

Chapter 14

Spatial and Camera Methods



14.1 Spatial Recapture Models

14.1.1 Introduction

In the past, various arrays of traps have been used to study animal populations that have been largely closed. This method of capture-recapture has a problem with some populations, namely, the movement of animals on the boundary of the trapping area as animals may move in and out of the area. Those animals with territories overlapping the array may have less exposure to traps than those in the interior of the grid and therefore have a lower probability of capture, thus introducing heterogeneity into the capture process.

One major problem is how to define the area of catchability, related to the edge effect, previously discussed in Sect. 2.4. This causes problems in determining the population density with nonspatial capture-recapture, and there is also the possibility of heterogeneity. This has led to an extensive development of spatial (explicit) capture-recapture (SCR/SECR) models for both closed and open populations and particularly for multiple species. SCR models are distinguished from non-spatial CR in that they include individuals' locations as latent or hidden variables. This chapter overlaps with the end of Chap. 12 on capture-recapture models, where spatial CR models with continuous captures are used.

Parsons et al. (2015) noted that non-spatial models that account for inherent individual heterogeneity may perform nearly as well as spatial models, especially where immigration and emigration are limited. Also they are computationally less demanding and do not make implicit assumptions related to the isotropy of home ranges. Borchers and Marques (2017) discussed how spatial models can be combined with capture-recapture or distance sampling (cf. the Appendix of his paper). Royle et al. (2014,a) gave a basic introduction to spatial capture-recapture, and Borchers and Fewster (2016) gave a helpful description of the various models

that were currently used. General reviews are given by Royle et al. (2018) and Tourani (2022).

Crum et al. (2021) compared the SCR method with distance sampling from a line transect. They noted that SCR was developed to address two shortcomings of traditional mark-recapture. First, abundance estimates from mark-recapture are not associated with a well-defined study area, so that population density cannot be formally estimated. Second, mark-recapture does not account for individual heterogeneity in detection probability due to individuals' locations relative to detectors. The authors found that spatial capture-recapture abundance estimates had lower root mean squared error than distance sampling estimates. Also, spatial capture-recapture 95% credible intervals for abundance had nominal coverage, i.e., contained the simulation value for abundance in 95% of simulations, whereas distance sampling credible intervals had below nominal coverage. The two methods are combined below in Sect. 14.4.4.

14.1.2 Description

SCR models generally consist of two processes: (1) an explicit process model that describes the spatial organization of home-range centers (or instead activity centers or both) and the movement of individuals over time and (2) an observation model that is conditional on the underlying state process that describes the probability of encounter as a function of an individual's location at the time of the sample together with a probability of detection parameter. Using the spatial information inherent in individual detections, SCR methods eliminate the need for ad hoc estimation of the effective sampling area, hence allowing for the estimation of density. Also unmodeled heterogeneity in the capture probability can result from the distance between animal "activity centers" and survey locations, and this will also have an affect on the animals in the area along with the possibility of trap addiction or trap shyness, introducing further heterogeneity. To minimize any invasiveness by researchers, camera methods that use spatial capture-recapture models have been increasingly used, and these are discussed below in Sect. 14.6.

The SCR methods have been widely applied to field data to estimate mammalian (carnivore) density, especially more recently for those with large home ranges and ranges of movement. To indicate the wide range of applications, they have been for birds (Mollet et al. 2015), sharks (Bradley et al. 2017), amphibians (Muñoz et al. 2016), and insects (Torres-Vila et al. 2012). The majority of SCR models are applied to camera trap survey data of naturally marked individuals (cf. Sect. 14.6 below) and secondly to data generated from genetic sampling methods (Sect. 14.8 below). However, many species are not uniquely identifiable from camera trap images (but see below), and other means of individual identification may be prohibitively costly or invasive. The methods have been applied to other data types such as acoustic recordings. Efford et al. (2009) and Dawson and Efford (2009) apparently first proposed the application of SECR methods to detection data collected without

physically capturing the animals themselves, but from an acoustic survey using an array of microphones (see Sect. 14.2.3).

A nontechnical introduction to SCR models was given by Borchers (2012) who listed further applications (with references) to the cage-trapping of possums, mist-netting birds, acoustic “trapping” of cetaceans from their vocalizations and birds from their song, DNA methods, and camera-trapping. He noted the relationship of SCR models and mark-recapture distance sampling, where “recaptures” occur with more than one observer. Spatially explicit capture-recapture models can be fitted using the SECR package version secr 4.0 (Efford 2011). See also R package version 2.1.0. (<http://cran.r-project.org/>, Efford 2013). Further background to the SCR model was given by Efford and Fewster (2013). Dey et al. (2019) described Bayesian model selection for SCR models.

Chandler and Royle (2013) stated that two advantages of SCR models are that they yield explicit estimates of animal density instead of population size within an unknown area, and they account for heterogeneity in capture probability arising from the juxtaposition of animal activity centers and sample locations. They noted that although the utility of SCR methods is gaining recognition, the requirement that all individuals can be uniquely identified excludes their use in many contexts. A modification that allows for this problem is considered in Sect. 14.5.2. However, SRC models can sometimes be prohibitively difficult for computing. A recent development in the SCR model is the use of combined data sources into one model, called integrated modeling, which can be helpful with sparse data sets. Examples of this are spatial resight models described in Sect. 14.5 and combined DNA and camera trapping (Sect. 14.8.2). Tourani et al. (2020) discussed this question of combining data sets that originate from different survey methods, or different data types collected using the same survey method, and developed a straightforward model. They investigated what improvement in estimation can be obtained using sparse data on brown bear (*Ursus arctos*) from the western Himalayas in Pakistan.

14.1.3 A Basic Model

When we have repeat visits to a site, we effectively have “recaptures.” We now consider a general introduction to SCR models from Linden et al. (2017). Let y_{ijk} be the encounter history for individual i at site j on survey k ($k = 1, \dots, K$) modeled as a Bernoulli variable depending on the location of the individual’s latent activity center $\mathbf{s}_i = (s_{i1}, s_{i2})'$ such that $\Pr(y_{ijk} = 1 \mid \mathbf{s}_i) = p_{ijk}$. If d_{ij} is the distance between the activity center, \mathbf{s}_i , and the location for trap j , they assumed that

$$p_{ijk} = p_{0,ijk} \exp[-d_{ij}^2/(2\sigma_i^2)].$$

Here $p_{0,ijk}$ is the encounter probability when $d_{ij} = 0$. Then $\text{logit}(p_{0,ijk})$ and $\log(\sigma_i)$ can be expressed in terms of covariates. The sites, described as grid cells at locations

g (where g describes the grid number and its location), have expected densities $E[D_g] = \mu(g, \beta)$, which are assumed to satisfy the regression

$$\log(E[D_g]) = \mathbf{w}'_g \boldsymbol{\beta}.$$

We now need to model the distribution of activity centers, namely,

$$\Pr(\mathbf{s}_i = g \mid \boldsymbol{\beta}) = \frac{\mu(g, \boldsymbol{\beta})}{\sum_g \mu(g, \boldsymbol{\beta})}.$$

The authors performed all model fitting in **R** (R Development Core Team 2016) with the package *unmarked* by Fiske and Chandler (2011), which maximizes the joint likelihood of model parameters and facilitates model comparisons and assessment. We now consider a number of variations to the above theory and some applications.

14.2 Some Examples

Before giving a number of different examples, we recall that a “trap” can also include any kind of recording device such as a camera or a human observer conducting a point-count survey. Also, we note that SCR models have some technical limitations (Romairone et al. 2018). For example, animal mortality in catch traps can be very high in small mammals, trap saturation can occur when using single-catch devices, and there is a need for a specific spatial organization of traps in the study area that depends on individual movements and home-range sizes. The authors focussed on small mammals and developed a trap attached to a nest to reduce animal deaths in traps. A Bayesian approach was used with covariates to estimate the parameters.

14.2.1 Multi-catch Traps

Single-catch traps have been previously considered, and we now consider multi-catch traps. Borchers and Efford (2008), building on the original introduction of Efford (2004), introduced new, flexible capture-recapture models that used the capture locations to estimate animal locations and spatially referenced capture probability, which we now consider. Suppose animals are counted at J traps on K occasions where the sample occasion can be an arbitrary time period, such as a single day in a camera trap study, or a 10 min survey interval. Trap locations are characterized by the matrix of spatial coordinates, \mathbf{X} with rows \mathbf{x}'_j ($j = 1, \dots, J$), where $\mathbf{x}_j = (x_{j1}, x_{j2})'$. The data resulting from this design gives the $J \times K$ matrix of

counts, $\mathbf{n} = \{n_{jk}\}$ ($k = 1, \dots, K$), where n_{jk} is the number observed at trap j and time k .

The authors focussed on traps that do not fill up, but which stop an animal from advancing to another trap within the same occasion, so-called “multi-catch” traps that hold an individual animal until it is released, but are able to simultaneously hold multiple individuals (e.g., mist nets for birds and pitfall traps for lizards). Contrary to other count-based methods, the authors induced correlation in the neighboring counts by organizing the traps sufficiently close together so that individual animals might be encountered at multiple locations. They did not make the customary assumptions that counts can be viewed as independently and identically distributed outcomes and that no movement occurs between sampling occasions. They assumed that home range centers occur independently in a plane according to an inhomogeneous Poisson process, and observed covariates were also included. Maximum likelihood estimates were obtained.

14.2.2 Passive Detector Array

Efford et al. (2009) extended the above model to estimate densities using a passive detector array, with “proximity” detectors such as automatic cameras and devices that passively collect DNA samples from animals without limiting their movement. Microphone arrays can also be used to assess the density of songbirds or whales from their vocalizations, which we focus more on in Sect. 14.2.3. A proximity detector records the presence of an individual at or near a point, but leaves the individual free to visit other detectors on the same occasion. In contrast to previous methods, density estimates may be obtained with this method from spatial observations accumulated during just one time interval (i.e., a single sample), since with proximity detectors each animal is potentially detected at multiple detectors.

The scope of this approach was enlarged by Efford et al. (2009) to allow for other types of trapping process under the same umbrella. In particular, they distinguished three kinds of detector: multi-catch trap previously mentioned, single-catch trap, (cf. a single-catch continuous spatial model in Sect. 12.15.7) and “proximity detectors,” which we now consider with a change in their notation ($k \rightarrow j$, $\mathbf{X} \rightarrow \mathbf{s}$, $S \rightarrow k$, and $\mathbf{Y} \rightarrow \mathbf{x}$).

The authors assumed that the probability of a particular individual i with home range center located at $\mathbf{s}_i = (s_{i1}, s_{i2})$ and recorded at detector j located at $\mathbf{x}_j = (x_{j1}, x_{j2})$ is a function of the distance $d_j(\mathbf{s}_i) = \| \mathbf{s}_i - \mathbf{x}_j \|$ and possibly of other covariates. It is assumed that there is independence between visits to different detectors. Each occasion-specific entry in an encounter history from an array of J proximity detectors gives a vector of length J , whose elements take the value 1 for detectors at which the individual was recorded at least once, and 0 otherwise.

Let $p_i(\mathbf{s}_i; \boldsymbol{\theta})$ be the probability that individual i is caught at least once, where $\boldsymbol{\theta}$ is a vector of detection parameters. Assuming an underlying Poisson process, we have that n , the number of capture histories, is $\text{Poisson}(Da)$, where D is the density,

$a = \int_{\mathcal{R}} p_{\cdot}(\mathbf{s}; \boldsymbol{\theta}) d\mathbf{s}$, and \mathcal{R} is the survey region. For $d_j(\mathbf{s}) = \|\mathbf{s} - \mathbf{x}_j\|$, we have

$$p_{\cdot}(\mathbf{s}; \boldsymbol{\theta}) = 1 - \prod_{j=1}^J \prod_{k=1}^K [1 - p_k(d_j(\mathbf{s}); \boldsymbol{\theta})],$$

where $p_k(d_j(\mathbf{s}); \boldsymbol{\theta})$ is analogous to the detection function in distance sampling. For example, using a half-normal function, we have

$$p_k(d_j(\mathbf{s}); \boldsymbol{\theta}) = g_0 \exp[-d_j(\mathbf{s})^2/(2\sigma^2)].$$

The authors also mentioned the hazard and negative exponential detection functions. The probability of the capture history ω_i for a given home range location \mathbf{s}_i and model parameters $\boldsymbol{\theta}$ is

$$\Pr(\omega_i | \mathbf{s}_i, \boldsymbol{\theta}) = \prod_{j=1}^J \sum_{k=1}^K p_{jk}^{\delta_{ijk}} (1 - p_{\cdot k})^{1 - \delta_{i k}},$$

where p_{jk} is the probability of detection at detector j on occasion k , $\delta_{ijk} = 1$ if individual i was detected at j on occasion k , and $\delta_{i \cdot k} = 1$ if $\sum_j \delta_{ijk} > 0$, and 0 otherwise. The likelihood function is then

$$L(\boldsymbol{\theta}, D) = \frac{(DA)^n \exp[-DA]}{n!} \prod_{i=1}^n \frac{\int \Pr(\omega_i | \mathbf{s}; \boldsymbol{\theta}) d\mathbf{s}}{a}.$$

The above general formulation applies to all three types of detection previously mentioned.

For an array of proximity detectors, we use

$$p_{jk} = p_k(d_j(\mathbf{s}); \boldsymbol{\theta}).$$

For the multi-catch trap method where the traps “compete” for animals, a competing risks hazard-rate form is appropriate, namely,

$$p_{jk} = \frac{h(d_j(\mathbf{s}))}{h_{\cdot}(\mathbf{s})} (1 - \exp[-h_{\cdot}(\mathbf{s})]),$$

where $h(d_j(\mathbf{s})) = -\log[1 - p_k(d_j(\mathbf{s}; \boldsymbol{\theta}))]$ and $h_{\cdot}(\mathbf{s}) = \sum_{j=1}^J h(d_j(\mathbf{s}))$. Maximum likelihood estimates of D and $\boldsymbol{\theta}$ can be obtained iteratively from $L(D, \boldsymbol{\theta})$, and details are given by the authors as well as an alternative conditional procedure and some extensions.

Efford (2011) modified an earlier model of Borchers and Efford (2008) as follows. Population density is defined as the intensity of a Poisson spatial point process with mean $D(\mathbf{x}; \boldsymbol{\phi})$, where \mathbf{x} is the location and $\boldsymbol{\phi}$ is a vector of parameters

relating density to location $\mathbf{x} = (x_1, x_2)$. For an observed individual i ($i = 1, \dots, n$), let ω_i be a set of detection histories on K successive occasions at a set of J detectors whose locations are known, along with ancillary data specific to each detection (e.g., sound intensity on a microphone array). The probability of observing a particular ω_i depends on a vector of detection parameters θ and on \mathbf{s}_i , the unknown home range center of individual i . This can be integrated out of the likelihood, namely,

$$L(\theta, \phi) \propto \Pr(n | \theta, \phi) \prod_{i=1}^n \int \Pr(\omega_i | \omega_i > \mathbf{0}, \mathbf{s}, \theta) f(\mathbf{s} | \omega_i > 0; \theta, \phi) d\mathbf{s},$$

where $\omega_i > \mathbf{0}$ indicates a detection history for animal i detected at least once, and f is the probability density function of centers for those detected.

The search area is assumed to be some sort of polygon (e.g., a square) and each search a sampling “occasion.” Let p_{jk} be the probability of detecting a particular animal in polygon j on occasion k , and model the instantaneous location \mathbf{u} with a bivariate probability density function such as a circular bivariate normal. Depending on how the various functions are defined, Efford provided several different models for estimating population density using the **R** package “SECR” (Efford 2013). Efford and Mowat (2014) modeled sex-based and other differences in home range size as they can potentially cause heterogeneity in individual detection and bias in estimates of density.

More recently Efford (2019) focussed on the shape of activity areas. He noted that the standard detection model in SECR is a monotonic decline with radial distance from the activity center, which follows intuitively if activity areas (or home ranges) are circular, though this is rarely the case. However, he used simulation to provide evidence that SECR estimates of density are robust to elliptical elongation of these areas when sampling is not directionally biased.

14.2.3 Spatial Passive Acoustical Surveys

This is a continuation of the previous section but with a focus on acoustical arrays. Acoustical methods have been referred to in several chapters in this book. Here we utilize additional spatial information. Stevenson et al. (2015) developed a general framework for animal density estimation from acoustic detections across a fixed microphone array. They noted that the method of Efford et al. (2009) has some problems. Firstly, it is only appropriate if each individual is only detectable on a single occasion (e.g., by virtue of making exactly one call). The likelihood presented assumes independent detections between calls, thus independence between call locations. This is unlikely to hold when individuals emit more than a single call, as locations of calls made by the same individual are almost certainly related.

Secondly, some individuals may not be recognizable from their calls. They included information on signal strength and arrival times and estimated call density. A detection was defined to be a received acoustic signal of a call that has a strength above a particular threshold, c , so that it is easily identifiable above any background noise. Detections with strengths below this threshold were discarded. They used the method to estimate calling male density of the Cape Peninsula Moss Frog *Arthroleptella lightfooti*.

A variation on the previous model was given by Stevenson et al. (2021) to estimate call rate and population density from passive acoustic surveys. They mentioned two further disadvantages of the SCR methods previously mentioned. The first is that they estimate call density (calls produced per unit area per unit time), which confounds the two parameters that are usually of interest: animal density (individuals per unit area) and call rate (calls produced per individual per unit time). The second is that their likelihoods are constructed under an assumption that call locations are a realization of a Poisson point process, so that once the location of one call is independent of all others. However, two calls produced by the same individual will have the same location if animals do not move, or similar locations if they do. For notation we begin with a study region \mathcal{R} with m detectors, and \mathbf{x}_k is the location of the k th detector. Latent (unobserved) data are listed as follows with ω , an indicator taking the value 0 or 1.

n = Unknown number of calling animals in \mathcal{R} ,

c_i = Number of calls produced by the i th animal during the survey,

ω_{ijk} = Detection indicator of j th call by i th animal at k th detector,

$\boldsymbol{\omega}_{ij} = (\omega_{ij1}, \dots, \omega_{ijc_i})'$, the capture history of j th call by i th animal,

$\boldsymbol{\Omega}_i = (\boldsymbol{\omega}_{i1}, \dots, \boldsymbol{\omega}_{im})$, the capture history matrix of the i th animal,

\mathbf{t}_{ij} = Vector of detections of arrival times associated with detections $\boldsymbol{\omega}_{ij}$,

$\mathbf{T}_i = \{t_{ijk}\}$ for all $i, j, k : \omega_{ijk} = 1$, signal arrival times for i th animal with detections in $\boldsymbol{\Omega}_i$.

We use a similar notation for the observed data, a truncated version of the above, but with superscript *, “call” replaced by “detected call,” and “animal” replaced by “detected animal,” apart from the following:

n^* = Number of animals with at least one detected call,

c_i^* = Number of calls by the i th animal detected by at least one detector.

We now develop a likelihood function with the observed data, namely, n^* , $\mathbf{c}^* = (c_1^*, \dots, c_{n^*}^*)'$, $\boldsymbol{\Omega}^* = (\boldsymbol{\Omega}_1^*, \dots, \boldsymbol{\Omega}_{n^*}^*)$, and $\mathbf{T}^* = (\mathbf{T}_1^*, \dots, \mathbf{T}_{n^*}^*)$. Let \mathbf{s}_i be the physical location of the i th animal ($\mathbf{s}_i \in \mathcal{S}$); then it is assumed that animals’ locations are a realization of an inhomogeneous Poisson process over the survey region, with

intensity (i.e., animal density) $D(\mathbf{s})$ at location \mathbf{s} . We can then model

$$\log(D(\mathbf{s})) = \beta_0 + \boldsymbol{\beta}' \mathbf{x}(\mathbf{s})$$

with covariates $\mathbf{x}(\mathbf{s})$. Assuming independence of individuals and building in dependence on call locations as latent variables, we end up with a likelihood of the form

$$\begin{aligned} L(\theta) &= f(n^*, \mathbf{c}^*, \boldsymbol{\Omega}^*, \mathbf{T}^*) \\ &= f(n^*) \prod_{i=1}^{n^*} \int_{\mathcal{R}} f(\mathbf{T}_i^* | \boldsymbol{\omega}_i^*, c_i^*, \mathbf{s}_i^*) f(\boldsymbol{\Omega}_i^* | \mathbf{c}_i^*, \mathbf{s}_i^*) f(c_i^* | \mathbf{s}_i^*) f(\mathbf{s}_i^*) d\mathbf{s}_i^*, \end{aligned}$$

where the various probability distributions $f(\cdot)$ were given by the authors. Distances $d(\mathbf{x}_k, \mathbf{s}_i)$ were built into the call detection function, where a hazard half-normal function was used. Maximum likelihood estimation was carried out.

Oedekoven et al. (2021) gave a comprehensive paper which compared the performance of the SECR/SCR model of Sect. 14.1.1 with plot sampling and distance sampling using passive acoustic monitoring (PAM) data. Generally the data are used to estimate either absolute animal density or some index of relative animal density such as call density (number of calls per unit area per unit time) or call counts (number of calls per unit time detected on a sensor). As population density estimation requires information that may not be available, the authors restricted their comparison to estimating call density. Each method, which is given a helpful description, requires different assumptions and also demands different capabilities from the PAM system in terms of the ability to localize detected calls. They used simulation and data from migrating bowhead whales to investigate the effect of violating the assumptions required by the methods. A major problem with SECR was lack of independence of the observations.

14.2.4 Some Miscellaneous Applications

Sollmann et al. (2012) investigated how the spatial arrangement and size of the trapping array influenced parameter estimates for SCR models. Their simulated study using black bear data showed that SCR models performed well as long as the extent of the trap array was similar to or larger than the extent of individual movement during the study period, and movement was at least half the distance between traps. SCR models performed well across a range of spatial trap setups and animal movements. Contrary to non-spatial capture-recapture models, they do not require the trapping grid to cover an area several times the average home range of the studied species.

Royle et al. (2013a) introduced a new concept of “ecological distance” instead of the usual Euclidean distance. This new metric endeavors to relate distance to

landscape usage as a function of cost-weighted distance to move from one location to another, along any path. Technically it is the least cost path.

Manning and Goldberg (2010) developed a method to construct spatially explicit capture-recapture encounter histories from locations of unmarked animals for estimating population size with conventional capture-recapture models. Animal locations were recorded as point coordinates during survey occasions, and the parameter of interest was the abundance of individual activity centers. The method was applied to burrowing owls using their nest burrow as an activity center. It can be used as a noninvasive method for conspicuous species where there are high detection probabilities and low territorial overlap.

It is generally assumed in the models we have considered that the home range is spatially stationary during the investigation. However, this is not always the case, and Royle et al. (2016) proposed a model to allow for home range shifts using Markovian transience or dispersal based on a Gaussian random walk in two dimensions. They introduced three models, partial transience, complete transience, and dispersal.

Sirén et al. (2016) developed a camera-trap to identify the unique ventral patches of American martens (*Martes americana*). Their method was designed to (1) determine the optimal trap configuration to photograph ventral patches; (2) evaluate the use of temporally clustered photographs to determine independence and improve identification; and (3) determine factors that influence identification probability. Their method was tested by comparing camera- and live-trap density estimates using SCR models. They documented camera-trap visits by radio-collared martens with unique artificial markings and uncollared martens over two winters in a remote forest in northern New Hampshire.

14.2.5 Using Telemetry

Ivan et al. (2013) used telemetry in a spatial capture-recapture survey. Telemetry tagging can occur prior to or during the mark-recapture process. Locations are then obtained from tagged animals during and/or immediately following the mark-recapture sampling session in order to produce an estimate of the proportion of time that each tagged animal spent within the study site. Let p_i be the probability that animal i is detected on any given occasion, so that $P_i = 1 - (1 - p_i)^K$ is the probability that animal i is captured one or more times on K occasions. The estimator of the superpopulation of animals that could have used the site during the trapping session is a Horvitz-Thompson type estimator

$$\widehat{N}_{SP} = \sum_{i=1}^r \frac{1}{\widehat{P}_i},$$

where r is the total number of different animals detected. Here \widehat{p}_i can be modeled as a logistic regression with covariates such as the distance-to-edge, the minimum distance from the mean capture location for each individual to the edge of the site. Let $\widehat{\theta}_i$ be the estimated proportion of time animal i spent on the study site (estimated via telemetry). Then

$$\widehat{N}_{SS} = \sum_{i=1}^r \frac{\widehat{\theta}_i}{\widehat{P}_i},$$

is an estimate of the number of animals within the boundary of the site. Variance estimation was also considered, as well as the advantages the method has over traditional methods.

Royle et al. (2013b), using telemetry, jointly modeled space usage, resource selection, and population density by integrating SCR data, such as from camera traps, mist-nets, or conventional catch traps, with resource selection (RSF) data from telemetered individuals. They provided a framework for estimation based on marginal likelihood to estimate simultaneously the parameters of the SCR and RSF models, an integrated SCR-RSF model (see also Royle et al. 2014). Linden et al. (2018) extended their model to accommodate overlap of individuals between data sources and add the functionality to an open-source R package (**oSCR**, Sutherland et al. 2019).

14.2.6 Modeling Interactions

A Poisson process is frequently used to model the spatial process for independent animals. However, animals do not always act independently due to avoidance or territorial behavior. Reich and Gardner (2014) introduced a model to allow for some independence using an approximate Strauss process to model the home range locations. The process includes a parameter that determines the strength of repulsion between home ranges, but it has some computational challenges. McLaughlin and Bar (2021) introduced an SCR model that includes dependency among capture probabilities through attractions, which could be considered as an individual stalking or investigating the territory of a nearby individual, or communal behaviors such as breeding or kill sharing. For a motivating example, we can consider tigers that, despite being mostly reclusive, will travel outside of their immediate territories and are known to be aware of nearby individual's movement and activities. We now consider their SCR model with attractions and notational change.

The attractions are modeled through the animal's movement by allowing for an occasion-level activity center, \mathbf{c}_{ik} for individual i at day k . This can change between sampling occasions by varying between individual i 's own home range center \mathbf{s}_i (representing the center of an individual's territory where it usually stays, sleeps, and

marks territory) and home range centers of nearby individuals. These home range centers are assumed to be uniformly distributed over the trapping area containing a trapping array of J traps located at \mathbf{x}_j ($j = 1, \dots, J$), with an outside buffer area to ensure each individual and its movement are contained within the entire area being considered. Here, $i = 1, \dots, N$ individuals are exposed to the trapping array during a period of K days. It is assumed that each individual i has a fixed but unknown home range center \mathbf{s}_i and an occasion-level activity center \mathbf{c}_{ik} for day k .

Let $\pi_{ii'k}$ be the probability that i 's activity center is located at the home range center of individual i' on day k (i.e., individual i is attracted to individual i' 's home range). Since $i' = 1, \dots, N$,

$$\mathbf{c}_{ik} \sim \text{Categorical}(\pi_{i1k}, \pi_{i2k}, \dots, \pi_{iNk}),$$

where a categorical distribution is a multinomial distribution with index 1. Here $\pi_{ii'k}$ is modeled as

$$\pi_{ii'k} = \begin{cases} 1, & \alpha_{ik} = 1, i = i' \\ \frac{\exp[-d(\mathbf{s}_i, \mathbf{s}_{i'})^2/(2\sigma_\pi^2)]}{\sum_{\ell:\ell\neq i} \exp[-d(\mathbf{s}_i, \mathbf{s}_\ell)^2/(2\sigma_\pi^2)]}, & \alpha_{ik} = 0, i' \neq i \\ 0, & \text{otherwise.} \end{cases}$$

It is assumed that $\alpha_{ik} \sim \text{Bernoulli}(\psi_\alpha)$ with ψ_α the probability that an individual remains at its own home range center on a given day, $d(\mathbf{s}_i, \mathbf{s}_{i'})$ is the Euclidean distance between the home range centers of individuals i and i' , and σ_π is a scale parameter dictating the range of attractions.

Let $y_{ijk} = 1$ if individual i is observed at the j th trap on the k th day with probability P_{ijk} , where

$$P_{ijk} = \Pr(y_{ijk} = 1) = 1 - \exp[-\lambda_0 \exp\{-d(\mathbf{c}_{ik}, \mathbf{x}_j)^2/(2\sigma^2)\}].$$

Using the complementary log-log link, we have

$$\text{cloglog}(P_{ijk}) = \log[-\log(1 - P_{ijk})] = \log(\lambda_0) - d(\mathbf{c}_{ik}, \mathbf{x}_j)^2/(2\sigma^2).$$

Estimation was carried using data augmentation (cf, A.5). The authors advised against using the model when considering large populations with high spatial density, as “with higher density and more overlapping of individual’ home ranges it would also become increasingly less feasible to discern attractions even from simulated CR data”

14.2.7 Occupancy and Spatial Combined

Clare et al. (2017) considered issues arising in combining two sets of data, especially that of dependence between the sets as independence can be difficult to obtain. They provided a method for combining occupancy and spatial capture-recapture models that permit covariance between the detections produced by different methods. But first, under a standard single method design, we have individuals i , sites j , and occasions k with an observation y_{jk} (or y_{ijk} depending on whether species or individual-level detections are considered). Let $z_j = 1$ if an individual is present at site j with probability ψ_j , and 0 otherwise, and let p_{jk} be the probability of detecting a species present at site j during interval k . The observations y_{jk} are then Bernoulli($\psi_j p_{jk}$), and ψ_j and p_{jk} can be expressed as logistic linear regressions.

With two sets of data, however, y_{jk} is redefined as a new variable taking four values, namely, $y_{jr} \sim \text{Categorical}(\pi)$. This has four possible detection outcomes π_r : detection by just method one (1), detection by just method two (2), or by both (3), or by neither (4). The authors gave a table of the four probabilities, for the values of y_{jr} depending on the models used. Let p_m be the probability of observing an individual given present by method m ($m = 1, 2$). For example, with independence of the two data sets, these four probabilities are, for the detection outcomes, $\pi_1 = z_j p_1(1 - p_2)$, $\pi_2 = z_j p_2(1 - p_1)$, $\pi_3 = z_j p_1 p_2$, and $\pi_4 = z_j(1 - p_1)(1 - p_2) + 1 - z_j$. If the two data sets are correlated, these probabilities are now $z_j \phi_1$, $z_j \phi_2$, $z_j \phi_3$, and $z_j \phi_4 + (1 - z_j)$, where the ϕ_r have a prior Dirichlet distribution with parameters $\alpha_1, \dots, \alpha_4$.

Bringing in the spatial capture-recapture, let individual i have a latent activity center s_i , where its detection probability during a specific sampling occasion is equal to p_0 . We now use y_{ijk} instead of y_{jk} with corresponding p_{ijk} to allow for individuals. The four probabilities for the four categorical variables are now $z_i p_{ijk1}(1 - p_{ijk2})$, $z_i p_{ijk2}(1 - p_{ijk1})$, $z_i p_{ijk1} p_{ijk2}$, and $z_i(1 - p_{ijk1})(1 - p_{ijk2}) + (1 - z_i)$ for the independence case, and $z_i \phi_1 p_{ijk}$, $z_i \phi_2 p_{ijk}$, $z_i \phi_3 p_{ijk}$, and $z_i(1 - p_{ijk}) + (1 - z_i)$ for the dependent case. Let d_{ij} be the distance of individual i from location j , so that we can assume p_{ijk} has some distribution such as $p_{ijk} = p_0 \exp[-d_{ij}^2/(2\sigma^2)]$. We can express $\text{logit}(p_0)$ and $\log(\sigma)$ as linear in individual, trap, or temporal covariates. The authors provided covariance expressions, and Markov chain Monte Carlo simulations were carried out.

14.3 Bayesian Models

As Bayesian methods are extensively used in this chapter, we give some additional examples of them, which use camera trapping information. Camera methods and their advantages, etc. are considered in more detail in Sect. 14.6.2.

14.3.1 Binomial Point Process

A Bayesian hierarchical model with data augmentation was developed by Royle et al. (2009). It consisted of a binomial point process model describing the distribution of home range centers of individuals and a model describing the observation of individuals trapped. Let \mathbf{s}_i ($i = 1, \dots, N$) be the location of the i th center, and assume that the centers are independently and identically distributed with a uniform distribution over some investigated region, say \mathcal{S} . A camera trapping array was used with \mathbf{x}_j the location of the j th trap ($j = 1, \dots, J$). If individual i is encountered y_{ijk} times in trap j , during interval k , we can use the model $y_{ijk} \sim \text{Poisson}(\lambda_0 g_{ij})$, where g_{ij} is some function of the distance $d_{ij} = \|\mathbf{s}_i - \mathbf{x}_j\|$, and λ_0 is the baseline encounter intensity. For g we can use, for example, the half-normal model $g_{ij} = \exp[-d_{ij}^2/\sigma]$ or the exponential form $g_{ij} = \exp[-d_{ij}/\sigma]$.

For the case where individuals can be captured at most once per trap, the observations are binary with $y_{ijk} = 1$ if individual i is captured in trap j during sample occasion k , and $y_{ijk} = 0$ otherwise. The authors viewed the binary observations conceptually as reductions of the counts that one could have observed in the more general case. For example, in DNA-based sampling, an individual might be encountered a number of times during any period, and the biological material (hair and so on) accumulates but cannot be partitioned into distinct visits after it is collected. More generally, while camera traps may yield multiple captures during each occasion, it is difficult to imagine that such multiple captures are independent. Instead of either applying an arbitrary and subjective rule to determine how to partition events into independent recaptures or devising a model of within-trap dependence, it is natural to reduce such data into binary encounter events. We then have in both situations

$$y_{ijk} \sim \text{Bernoulli}(\pi_{ij}),$$

where

$$\pi_{ij} = \Pr(y_{ijk} = 1) = 1 - \exp[\lambda_0 g_{ij}],$$

giving us Bernoulli and Poisson encounter models. For the Bernoulli model with half-normal g , the authors suggested the complementary log-log transform

$$\log(-\log(1 - \pi_{ij})) = \log(\lambda_0) - (1/\sigma)d_{ij}^2,$$

For Poisson observations we have $y_{ijk} \sim \text{Poisson}(\lambda_{ij})$, where $\lambda_{ij} = \lambda_0 g_{ij}$ so that we can use

$$\log(g_{ij}) = \log(\lambda_0) + (1/\sigma)d_{ij}^2.$$

The authors included other covariates in the above log regression models as well.

With data augmentation, one augments the n observed encounter histories with some large number of “all-zero” histories, say M such histories, which include the actual N individuals in the population as a subset. Bayes formulation can be introduced by assuming a discrete uniform prior for N on the integers $0, 1, \dots, M$. Let ψ be the probability that an individual on the list of M is a member of the population of size N that was exposed to sampling by the trap array. We then assume $N \sim \text{Binomial}(M, \psi)$ and that ψ has a uniform distribution on $[0,1]$. Integration of the binomial prior for N over the uniform prior for ψ yields the discrete uniform prior for N . One now introduces $M - n$ observations $y_i = 0$ for $i = n + 1, \dots, M$, for which there is no trap information so that $y_{ijk} = 0$ for all j and k as well. Let $w_i = 1$ if the i th element of the augmented list ($i = 1, \dots, M$) is a member of the population of size N and 0 otherwise. Imposing $w_i \sim \text{Bernoulli}(\psi)$, the induced prior distribution on $N = \sum_i w_i$ is uniform on the integers $0, 1, \dots, M$. Derived parameters can then be estimated such as the population size of region S of area $A(S)$, namely, $N(S) = \sum_{i=1}^M w_i$ and density $D(S) = N(S)/A(S)$. Markov chain Monte Carlo can be used for the analysis of the augmented data, and further details and methods for model testing were given by the authors Royle et al. (2009).

Method of Royle et al. (2009)

Following, Royle et al. (2009), suppose we have a grid of J traps at locations \mathbf{x}_j ($j = 1, \dots, J$) and, for animal i in the population of N animals, its center of activity is $\mathbf{s}_i = (s_{1i}, s_{2i})'$. These activity centers are assumed to be fixed for the sampling period of the population, and the ensuing model is conditional on the unknown activity centers, which are treated as random effects. It is assumed that these centers are uniformly distributed over some region of area A that contain the traps. The camera observations are denoted by y_{ik} , taking possible values $0, 1, \dots, K$, where $y_{ik} = 0$ indicates no capture of individual i at sampling occasion k ($k = 1, \dots, K$), and nonzero values indicate the trap location at which the animal is caught at occasion k . For example, if $K = 5$ and $J = 10$ traps, then the encounter history $\omega_i = (0, 8, 9, 0, 8)'$ indicates an animal not captured in occasions 1 and 4 and captured in trap number 8 (occasions 2 and 5) and number 9 (occasion 3).

It is assumed that the y_{ik} are independent observations of a categorical random variable with probabilities π_i , a vector of length $J + 1$ including the zero case of not being caught. Here, the π_i depend on the latent variables \mathbf{s}_i and \mathbf{x}_j . The model can be thought of as describing the result of rolling an $J + 1$ -sided die so that the observation model can be described as

$$y_{ik} | \mathbf{s}_i, \boldsymbol{\theta} \sim \text{Categorical}(\pi_i(\mathbf{s}_i, \boldsymbol{\theta})),$$

depending on a parameter $\boldsymbol{\theta}$. There is an equivalence between a categorical random variable and a multinomial trial. If we create a vector of 0s, of length $J + 1$, and insert a single 1 into that vector in position $y_{ij,k+1}$, then the resulting vector is a multinomial trial.

It is assumed that the y_{ij} are independent among survey occasions for each individual and among individuals. This means that multiple individuals may be

captured in the same trap during any occasion, which is possible with cameras. However, the model implies that capture in traps for each individual is mutually exclusive events. This means that (i) individuals cannot be captured more than one time in the same trap for each period; and (ii) individuals cannot be captured in more than one trap. While the first is technically incorrect, since cameras (usually) operate continuously between checks, as an operational matter when the sampling interval is short (e.g., daily), data can be processed into single trap-specific encounters of individuals for each sampling occasion. The first implication is effectively satisfied under the current manner in which the data are processed.

While (i) can be achieved if the sampling interval is short (e.g., on one day), (ii) can be modelled by assuming multiple captures represent independently and identical multinomial trials, though it may not occur if data are sparse. The encounter process will depend on the movement process, and, following Royle and Young (2008), the authors modeled this process by assuming that individuals move around their activity center \mathbf{s}_i according to a bivariate normal distribution with mean \mathbf{s}_i and standard deviation σ , so that locations on consecutive days (\mathbf{u}_{ik} , say) are independent draws from this distribution. Assuming a trap has physical area, say χ_j , over the region around trap j within which an animal may be exposed. The probability of exposure of individual i to trap j is

$$\phi_{ij} = \int_{\mathbf{x} \in \chi_j} f(\mathbf{x}; \mathbf{s}_i, \sigma) d\mathbf{x},$$

which for small χ_i gives us approximately

$$\phi_{ij} = \alpha f(\mathbf{x}_j; \mathbf{s}_i, \sigma)$$

where $\alpha = \text{area}(\chi_j)$. If exposure to traps is independent and mutually exclusive, then the total exposure is

$$\phi_i = \sum_{j=1}^J \phi_{ij} = \alpha \sum_{j=1}^J f_{ij} = \alpha E_i,$$

where $f_{ij} = f(\mathbf{x}_j; \mathbf{s}_i, \sigma)$. Let

$$\begin{aligned} p_i &= \Pr(\text{capture individual } i) \\ &= \Pr(\text{capture} \mid \text{exposed}) \Pr(\text{exposed}) \\ &= r \phi_i \text{ say} \\ &= r \alpha E_i. \end{aligned}$$

To get round an identifiability issue for r and α , the cell probabilities π_i are modeled in terms of other parameters.

To estimate the effective sampled area A_e , say, Royle et al. (2009) suggested dividing up the population area A into G area pixels (where a pixel can be a large unit of area), g_1, \dots, g_G . They imagined sampling A by flipping a coin for every pixel and then sampling that pixel if the coin comes up “heads.” In that case the pixel area is surveyed. If the coin has probability P_i of coming up heads for pixel i , then we would expect to sample an area of size $A_e = \sum_{i=1}^G P_i$, where $P_i = 1 - (1 - \phi_i)^K$. A Bayesian method of estimation using data augmentation (Sect. A.5) was carried out (see also Royle et al. 2009). For further details see Royle et al. (2009). This model was developed further by Gardner et al. (2009), Gardner et al. (2010), and Royle et al. (2011) and also used for analyzing DNA data.

14.3.2 Connecting Bayes and Frequency Methods of Estimation

Dorazio (2013) developed Bayes and empirical Bayes estimators of abundance and density from spatial capture-recapture data and used a model similar to the previous one (but with j and k interchanged). He established a formal connection between frequency and Bayesian methods, as well as between the two Bayesian methods. Once again we have the region of interest to be \mathcal{S} of area A , a bounded subset of \mathbb{R}^2 with population size $N(\mathcal{S})$. For the number of activity centers where an individual moves randomly around its activity center, we can assume a Poisson point process to specify the spatial distribution of $N(\mathcal{S}) = N$. Alternatively, we can use a binomial point process to specify the spatial distribution, where N is now a constant, and a model can be fitted using Bayesian methods and data augmentation as above. As a good deal of mathematical modeling is used, we provide just an outline of some of the key ideas.

For the Poisson model, let $\lambda(\mathbf{s})$ be the first-order intensity function of an inhomogeneous Poisson process for $N(\mathcal{S})$, where $\lambda(\mathbf{s})$ is an abbreviation for $\lambda(\boldsymbol{\nu}(\mathbf{s})) = \exp[\boldsymbol{\beta}' \boldsymbol{\nu}(\mathbf{s})]$ at (latent) location \mathbf{s} . Then

$$N(\mathcal{S}) \sim \text{Poisson}(\mu), \text{ with } \mu = \int_{\mathcal{S}} \lambda(\mathbf{s}) d\mathbf{s}.$$

Given a realized population of size $N(\mathcal{S}) = N$, the N activity centers are assumed to be independent with probability density function $\lambda(\mathbf{s})/\mu$, and the conditional density of the N locations \mathbf{s}_i is

$$f(\mathbf{s}_1, \dots, \mathbf{s}_N) = \prod_{i=1}^N [\lambda(\mathbf{s}_i)/\mu].$$

(Note that if $\lambda(\mathbf{c}) = \lambda$, then $\mu = \lambda A(\mathcal{S})$, and $\lambda/\mu = 1/A(\mathcal{S})$, the uniform distribution on \mathcal{S}).

It is assumed that y_{ij} is the number of times an individual i may be detected at the j th trap ($j = 1, \dots, J$) during each of K_j trapping periods of equal duration, and recaptures of an individual in different traps occur independently. Although multiple detections of an individual at the same trap are readily modeled, the author considered a model in which individuals are assumed to be detected only once per trapping period at any trap. Each detection is assumed to be an independent Bernoulli outcome with probability $p(\mathbf{s}, \mathbf{x}_j)$, a function of the Euclidean distance between an individual's activity center \mathbf{s} and the location \mathbf{x}_j of trap j , for example,

$$p(\mathbf{s}, \mathbf{x}_j) = p_0 \exp[-\|\mathbf{x}_j - \mathbf{s}\|^2 / (2\sigma^2)],$$

where p_0 denotes the maximum probability of capture (applicable when $\mathbf{s} = \mathbf{x}_j$).

Now $y_{ij} \sim \text{Binomial}(K_j, p(\mathbf{s}, \mathbf{x}_j))$, or abbreviated as $\text{Bin}(y_{ij} | K_j, p(\mathbf{s}, \mathbf{x}_j))$. Let $\boldsymbol{\theta} = (\boldsymbol{\beta}, p_0, \sigma)$; then the overall likelihood function turns out to be

$$L(\{y_{ij}\}, N | \boldsymbol{\theta}) = \frac{\exp[-\mu]}{N!} \prod_{i=1}^N \left[\int_{\mathcal{S}} \lambda(\mathbf{s}_i) \prod_{j=1}^J \text{Bin}(y_{ij} | K_j, p(\mathbf{s}_i, \mathbf{x}_j)) d\mathbf{s}_i \right],$$

However, this model assumes that N is known and all the y_{ij} are known. Let \mathbf{Y}_{obs} be the matrix of observed values of y_{ij} . If n is the number observed and $n_0 = N - n$, then summing out n_0 , the author showed that the appropriate likelihood is

$$L(\mathbf{Y}_{obs}, n, | \boldsymbol{\theta}) = \frac{\exp[-\mu(1 - \pi_0)]}{n!} \prod_{i=1}^n \left[\int_{\mathcal{S}} \lambda(\mathbf{s}_i) \prod_{j=1}^J \text{Bin}(y_{ij} | K_j, p(\mathbf{s}_i, \mathbf{x}_j)) d\mathbf{s}_i \right],$$

where π_0 , the probability that an individual in the population was unobserved, is

$$\pi_0 = \int_{\mathcal{S}} \frac{\lambda(\mathbf{s})}{\mu} \prod_{j=1}^J [1 - p(\mathbf{s}, \mathbf{x}_j)]^{K_j} d\mathbf{s}.$$

The maximum likelihood estimate of $\boldsymbol{\theta}$ can then be found numerically.

We now find that $n_0 | \mathbf{Y}_{obs}, n, \boldsymbol{\theta} \sim \text{Poisson}(\mu\pi_0)$, and an empirical Bayes estimator of N is

$$\widehat{N} = n + E[n_0 | \mathbf{Y}_{obs}, n, \widehat{\boldsymbol{\theta}}] = n + \widehat{\mu}\widehat{\pi}_0,$$

where $\widehat{\mu}$ and $\widehat{\pi}_0$ correspond to the values of μ and π_0 evaluated at the estimate $\widehat{\boldsymbol{\theta}}$. To find a Bayes estimator of abundance, we form the (unnormalized) posterior density

$$f(\boldsymbol{\theta}, n_0 | Y_{obs}, n) \propto L(\mathbf{Y}_{obs}, n, n_0 | \boldsymbol{\theta}) f(\boldsymbol{\theta}).$$

Since $N = n + n_0$, Bayesian inferences about N are based entirely on the posterior distribution of n_0 , which has probability mass function

$$f(n_0 | \mathbf{Y}_{obs}, n) = \int f(n_0 | \mathbf{Y}_{obs}, n, \boldsymbol{\theta}) f(\boldsymbol{\theta} | \mathbf{Y}_{obs}, n) d\boldsymbol{\theta}.$$

Summaries of the posterior distribution of N (e.g., mean, mode, quantiles, etc.) may be estimated directly using the above equation.

Considering a Bayesian approach to the above, we need a prior distribution for $\boldsymbol{\theta}$. It is assumed that each i th individual moves randomly around its activity center \mathbf{s}_i , and its actual location during the k th survey, \mathbf{u}_{ik} , has a bivariate normal distribution $N_2(\mathbf{s}_i, \sigma^2 \mathbf{I}_2)$, where \mathbf{I}_2 is a 2×2 identity matrix. Further details are given by the author.

Marques et al. (2012) used spatially explicit capture-recapture methods (SECR) to estimate minke whale density of sounds from data collected at 16 bottom mounted hydrophones. They described the Bayesian and likelihood-based approaches to SECR and used some simulation and real data to evaluate the performance of the two approaches using the same model for sound detection. The main differences between the two approaches are that, firstly, the Bayesian model is parameterized in terms of abundance N within an area A , rather than density D , and this N is assumed to be a fixed quantity. This leads to a binomial likelihood for n the sample size given N , rather than, say, Poisson likelihood for n given D . Secondly, data augmentation can be used to deal with the unobserved capture histories and sound source locations for the Bayesian method. Thirdly, prior distributions are used on all unknown parameters, although since uniform priors with widely spaced limits are commonly used, the posterior and likelihood surfaces will have the same shape within the truncation bounds. Both likelihood and Bayesian methods gave very similar estimates on average for the study. However, the likelihood-based methods were easier to implement, thanks to the secr R package, and appeared to be more convenient for simpler applications.

14.4 Further Extensions

14.4.1 Stratified Populations

Royle and Converse (2014) developed a hierarchical SCR model for stratified/grouped populations that are stratified by space, time, or other factors. Let $y_{ijk} = 1$ if individual i is “captured” in trap j ($j = 1, \dots, J$) on occasion k ($k = 1, \dots, K$) with probability p_{ijk} , and 0 otherwise. The vector of observations for individual i during occasion k is given by $\mathbf{y}_{ik} = (y_{i1k}, y_{i2k}, \dots, y_{iJk}, y_{i,J+1,k})'$ where, for an unobserved individual, the last element $y_{i,J+1,k}$ corresponds to “not captured” and contains a single 1 with the remaining values being 0. The sampling was carried out in $g = 1, \dots, G$ groups or strata. Let \mathbf{x}_j describe the location of

trap j and \mathbf{s}_i the activity center of individual i . With $d_{ij} = \|\mathbf{s}_i - \mathbf{x}_j\|$, the authors considered the model,

$$p_{ijk} = p_0 \exp[-d_{ij}^2/(2\sigma^2)],$$

which can also be expressed in a loglinear form $\log(p_{ijk}) = \alpha_0 + \alpha_1 d_{ij}^2$.

Focussing now on standard live traps, also called “single-catch” traps, an individual can be captured in at most one trap. Let N_g be the size of the g th strata and $N = \sum_g N_g$, the total population size. For each captured individual i ($i = 1, 2, \dots, n$), let g_i indicate the stratum to which individual i belongs, with N_g implying a specific prior distribution for the group membership variables, g_i . Also $n = \sum_{g=1}^G n_g$ is the sum of the group-specific sample sizes.

For the hierarchical model, we have two main components: (1) the observations \mathbf{y}_{ik} conditional on the latent population size of each group, which describes the probability of encounter of individuals in traps; and (ii) a model for the latent abundance variables N_g that describes variation in abundance among groups. Now \mathbf{y}_{ik} is a multinomial trial with corresponding probabilities π_{ik} , a vector where each element represents the probability of being encountered in a trap (for elements $1, \dots, J$) or not captured at all (element $J+1$). Here $\pi_{i,J+1} = 1 - \sum_{j=1}^J \pi_{ij}$. The authors proposed using a multinomial logistic model

$$\pi_{ij} = \exp[\eta_{ij}]/(1 + \sum_{j=1}^J \exp[\eta_{ij}]),$$

so that

$$\text{mlogit}(\pi_{ij}) = \eta_{ij} = \alpha_0 + \alpha_1 d_{ij}^2.$$

Additional covariates can also be added by adding a term such as a behavioral term $\alpha_2 c_{ik}$ where $c_{ik} = 0$ before the occasion of the first capture, and $c_{ik} = 1$ thereafter.

Looking now at the group structure, let \mathcal{S}_{g_i} denote the two-dimensional state-space of the random variable \mathbf{s} for an individual belonging to group g , such as a polygon around trap array g . Conditional on knowing which group individual i belongs to, that is, the value of g_i , it is assumed that the distribution of \mathbf{s}_i given g_i is the uniform distribution on \mathcal{S}_{g_i} . It is also assumed that the N_g are mutually independent Poisson (λ_g) variables and $\log(\lambda_g) = \beta_0 + \beta_1 u_g$ where u_g is some group-specific covariate. In practice, this would suppose that the trapping grids are spaced sufficiently far apart and sampling is sufficiently short in duration so as to preclude movement of individuals among populations. Alternatively, for computational reasons, one can use the multinomial joint distribution model for the N_g , conditional on the total N (the multinomial index) with cell probabilities

$$\theta_g = \lambda_g / \sum_g \lambda_g.$$

Prior distributions were placed on the various parameters and data augmentation was used. Further development on this topic of a stratified population was given by Royle et al. (2015).

14.4.2 Presence-Absence Data Only

Ramsey et al. (2015) extended a model of Chandler and Royle (2013) to where only presence-absence data can be collected and assumed that individuals can be detected at multiple sampling devices (e.g., camera traps, tracking plots, or bait stations), thus producing spatially correlated detections. The sampling design consists of an array of J devices with locations $\mathbf{x}_j = (x_{j1}, x_{j2})$ ($j = 1, \dots, J$) and set for K occasions. The array of devices are laid out in such a way that individuals could potentially be detected at multiple devices (e.g., a grid pattern). Individuals may be detected at any of the device locations (but not captured).

Let $h_{jk} = 1$ if there is at least one detection on device j and occasion k ($k = 1, \dots, K$), giving the $J \times K$ matrix \mathbf{H} . For example, the first row of \mathbf{H} could be (01001) indicating detections on occasions 2 and 5 for device number 1. For a population of N individuals in area A , and assuming individuals can be detected only at most once, individual i has an encounter history $y_{ijk} = 1$ if encountered at site j and occasion k , and 0 otherwise. Each individual i is also defined by a center of activity $\mathbf{s}_i = (s_{i1}, s_{i2})$, its nominal home range center. The locations of home-range centers are unknown, but are considered to be fixed for the duration of sampling. Individuals move about their home-range centers according to some probability distribution (e.g., bivariate normal) and in the process can potentially be exposed to detection. It is also assumed that the \mathbf{s}_i have a bivariate uniform distribution. Assuming detection declines with distance d from a device and movements around the home-range center occur with bivariate normal distribution, then

$$p_{ij} = g_0 \exp[-d_{ij}^2/2\sigma^2], \quad d_{ij} = \| \mathbf{x}_j - \mathbf{s}_i \|,$$

where g_0 is the per occasion probability of detection when the home-range center and the device location coincide (i.e., $d = 0$), and p_{ij} is the probability of detection of individual i at device j . Then

$$y_{ijk} \sim \text{Bernoulli}(p_{ij}) \text{ and } h_{jk} \sim \text{Bernoulli}(P_j).$$

where

$$P_j = 1 - \prod_{i=1}^N (1 - p_{ij}),$$

the probability that at least 1 individual is detected by device j . Assuming data is aggregated at each of the devices, then $n_j \sim \text{Binomial}(P_j, K)$, where $n_j = \sum_{k=1}^K h_{jk}$. The estimation problem now reduces to one of estimating the latent \mathbf{s}_i and N . This is carried out using data augmentation by assuming the existence of M individuals, where M is much greater than N . Here we have M latent indicator variables w_i so that we now have

$$P_j = 1 - \prod_{i=1}^M (1 - p_{ij} w_i), \text{ and } w_i \sim \text{Bernoulli}(\psi),$$

where ψ is equivalent to the proportion of the M individuals that are potentially exposed to detection. This implies that when $w_i = 0$, the probability that individual i is detected in any trap (p_{ij} , say) is also 0, and conversely when $w_i = 1$, individual i contributes their individual detection probability p_{ij} to P_j . An estimate of N , the number of home range centers, is given by

$$\widehat{N} = \sum_{i=1}^M w_i.$$

Here, prior distributions were needed for g_0 , σ , and ψ to carry out MCMC computations and obtain posterior distributions of the parameters, as referred to in the authors' supplementary materials with R code.

14.4.3 Viability for Introduced Species

Chandler et al. (2015) looked at the reintroduction of a species and its viability, which is similar to the idea of colonization. Their spatial occupancy model is described as follows. Let $z_{ik} = 1$ if site i was truly occupied by at least one individual during year k , and $z_{ik} = 0$ if the site was unoccupied. The difference from the usual model is that the true initial occupancy state is determined by the reintroduction design so that the z_{i1} are data, not latent variables. In subsequent years

$$z_{ik} \sim \text{Bernoulli}(\gamma_{i,k-1}(1 - z_{i,k-1}) + \phi_{i,k-1} z_{i,k-1}),$$

such that an unoccupied site at time $k - 1$ is colonized with probability $\gamma_{i,k-1}$, and an occupied site remains occupied with probability $\phi_{i,k-1}$. Here, colonization is a function of the dispersal ability of the species, the occupancy status of all other sites in the metapopulation network, and the spatial arrangement of the sites. A distance-based dispersal model was used as the basis for each pairwise colonization probability, and a model based on the Gaussian kernel was chosen because it is a

standard model used to describe a gradual decrease in dispersal probability with distance (Clobert et al. 2012). We then have that the probability that site i is colonized by at least one individual from site j is

$$\rho_{ijk} = \rho_0 \exp[-d_{ij}^2/(2\sigma^2)]z_{j,k-1},$$

where ρ_0 is the baseline colonization probability for coincident sites, d_{ij} is the Euclidean distance between sites, and σ^2 is the scale parameter determining the rate of decay in colonization probability as a function of distance. This parameterization enforces the condition that site i cannot be colonized by site j if $z_{j,k-1} = 0$. Having defined each pairwise colonization probability, the cumulative probability of a colonization event is given by

$$\gamma_{ik} = 1 - \left\{ \prod_{j=1}^S (1 - \rho_{ijk}) \right\},$$

where S is the total number of sites in the metapopulation network. A local extinction probability ε was introduced, and a pseudo rescue effect was introduced modeled as $\phi_{i,k-1} = 1 - \varepsilon_i(1 - \gamma_{i,k-1})$ i), indicating that the probability that an occupied site remains occupied is one minus the probability it goes extinct and is not recolonized. Let p_{ijk} be the probability of detection, given present, so that $y_{ijk} = 1$ if present, and 0 otherwise, and

$$y_{ijk} \sim \text{Bernoulli}(z_{ik} p_{ijk}).$$

Also $\text{logit}(p_{ijk})$ was expressed in terms of temperature and wind covariates. Bayesian methods were used with vague prior distributions, and Markov chain Monte Carlo sampling was carried out. The method was applied to Chiricahua leopard frogs.

14.4.4 SECR and Distance Sampling (DS)

Borchers et al. (2015) developed a class of models that included hybrids of capture-recapture and distance sampling models. Their class of models accommodates a spectrum of models ranging from nonspatial CR models (with no information on animal locations) through to DS and mark-recapture distance sampling (MRDS) models, in which animal locations are observed without error. Between these lie spatially explicit capture-recapture (SECR) models that include only capture locations and a variety of models with less location data than are typical of DS surveys, but more than are normally used on SECR surveys. They also included means of improving inference from SECR models by adding supplementary location data and a means of incorporating measurement error into DS and MRDS models as follows.

Consider a survey of a region with surface area A in which J detectors are deployed on K occasions. It is assumed that animals are independently distributed in this region according to a nonhomogeneous Poisson process (NHPP) with parameter vector ϕ and intensity $D(\mathbf{s}; \phi)$ at location \mathbf{s} , where the probability that an animal there is detected by at least one detector on the survey is $p(\mathbf{s}; \theta)$. Then the locations of n detected animals, $\mathbf{S} = (\mathbf{s}_1, \dots, \mathbf{s}_n)$ are realizations of a filtered NHPP with intensity $D(\mathbf{s}; \phi)p(\mathbf{s}; \theta)$ at \mathbf{s} . A probability model for the outcomes of a survey is a product of conditional probabilities: (1) the probability of detecting n animals, $f(n; \phi, \theta)$; (2) the probability density function (pdf) of animal locations, \mathbf{S} (the authors used \mathbf{X}), conditional on detection, denoted by $f(\mathbf{S}; \phi, \theta)$; and (3) the probability of observing the capture histories Ω conditional on detections and detected animal locations \mathbf{S} , namely, $f(\Omega | \mathbf{S}, \theta)$. Here $\Omega = (\omega_1, \dots, \omega_n)$, where ω_i is the capture history of the i th animal. Putting the two probability functions and the density function together, we have the joint pdf

$$f(\mathbf{S}, n, \Omega; \phi, \theta) = f(n; \phi, \theta) f(\mathbf{S}; \phi, \theta) f(\Omega | \mathbf{S}, \theta).$$

Each of the three distributions is discussed by the authors and expanded where necessary. For example, the capture histories are expanded so that $\omega_{ijk} = 1$ if animal i is detected by detector j on occasion k , and 0 otherwise. Let $\omega_{ik} = (\omega_{i1k}, \dots, \omega_{iJk})$ with full capture history $\omega_{i1}, \dots, \omega_{iK}$, finally giving Ω for $f(\Omega | \mathbf{S}, \theta)$. In addition to observing ω_{ijk} , $\mathbf{y}_{ijk} = (y_{ijk1}, \dots, y_{ijkM})$ are observed containing M different kinds of data, each of which is a noisy observation of animal location. An example with $M = 2$ is an acoustic survey in which a detection is recorded on a microphone as well as the time of arrival of the sound at a microphone. This leads to multiplying the previous joint distribution by a fourth $f(\mathbf{Y} | \mathbf{S}, \Omega; \gamma)$. Special cases of this model were given by the authors using various conditional distributions. For example, SECR models without location observations \mathbf{Y} are obtained by omitting its density function.

14.5 Spatial Mark-Resight Models

The mark-resight model consists of capturing a sample of animals and either tagging them or identifying them through natural markings during a single marking event that takes place prior to resighting surveys. The marked individuals form an individually identifiable portion of the population, and a noninvasive technique such as camera trapping or visual sighting can be used to obtain “recapture data.” In addition, sightings of unmarked individuals are recorded during resighting surveys, and both data types can be used in the mark-resight model.

Rich et al. (2014) were apparently the first to estimate the density of a population of carnivores using a spatial mark-resight model (SMR) model. They used the zero-truncated Poisson-log-normal model in program **MARK** because marked individuals could not have zero encounter histories (i.e., had to be photographed at

least once to be known). They compared spatial mark-resight and spatially explicit capture-recapture models. McClintock (2012) described how to use **MARK** to analyze mark-resight models.

Kane et al. (2015) applied an SMR model to the endangered African lion (*Panthera leo*). They noted that while the model assumes independent activity centers, the model is known to be robust to this assumption as shown by Russell et al. (2012), who discussed in detail sensitivity to model assumptions including the effect of pooling some dependent observations. By using a hierarchical approach, they were able to combine the data set consisting of marked captures and recaptures with the data set of counts of unmarked individuals at camera trap locations by sharing the scaling parameter (σ) in a joint model. Thus, the process model describing individual home range movements around an activity center is shared between the two component models, and they assumed the distance an individual moves from the hypothetical home range center is similar, whether an animal is marked or unmarked.

Sollmann et al. (2013b) used the SMR model augmented with telemetry data. They noted that the resight models without the spatial component suffered from some of the same shortcomings when it comes to estimating population density. For example, defining the area sampled is problematic, and heterogeneity in capture probability induced by different exposure of individuals to the survey locations cannot be modeled explicitly, though perhaps using, for example, covariates, if available.

Whittington et al. (2017) discussed SMR models. They noted that if a nonspatial-resight model is used, a critical assumption is that the marked and unmarked animals have the same probabilities of being encountered during the resighting occasions. In its simplest form N , the population size is estimated by $\hat{N} = m + u/\hat{p}$, where m is the number marked, u is the number unmarked, and \hat{p} is the estimated probability of encountering a marked individual during a resight survey. If marked animals have a higher probability of encounter than unmarked animals, the model will underestimate the number of unmarked animals in the population and overall abundance. A possible violation of the assumption of the same probability is that the marking process makes marked animals easier to see. If spatial sampling is used, animals are marked on one spatial array of traps and later resighted on the same or different array of detectors. This means the marked and unmarked populations will usually have different encounter probabilities because their spatial distributions differ.

Marked animals will typically be distributed closer to the trap array and will therefore have a higher average encounter rate than unmarked animals when the trap and resighting arrays overlap and a lower average encounter rate if the trapping and resighting arrays do not overlap. To get round this problem, the authors introduced a generalized SMR model that includes a sub-model for the marking processes to relax the assumption of existing SMR models that marked and unmarked animals share the same spatial distribution. It also included telemetry data. The notation is similar to that used by Chandler and Royle (2013) described above with the main difference being that superscripts are used for marked, unmarked, resight,

and telemetry. The model is complex and the reader referred to the paper and its appendix for details.

Alonso et al. (2015) compared mark-recapture and mark-resight methods for bobcats and introduced a hybrid model in which data from natural tags and artificial tags are combined using a Poisson model for resightings of the m artificially marked individuals and a zero-truncated Poisson model for sightings of the n individually identifiable (but not artificially-marked) individuals photographed at least once.

McClintock et al. (2009a) considered estimating abundance using SMR when sampling is with replacement or the number of marked individuals is unknown. They introduced the Poisson log and zero-truncated Poisson log-normal mixed effects models (PNE and ZPNE, respectively). Unlike LNE, the logit-normal mixed effects model for sampling without replacement (McClintock et al. 2009b), PNE may be modified for when the number of marked individuals is not known exactly and sampling is with replacement.

14.5.1 Different Sightings for Marked and Unmarked

Efford and Hunter (2018) considered a mark-resight model where the sighting probabilities of pre-marked individuals can differ from those unmarked because marked animals may not be a random sample. With spatial sampling, animals close to the traps are likely to predominate in the marked sample, so that the marking and sighting processes should be jointly modeled. Data were assumed to be collected at pre-determined points or by searching disjoint polygonal areas. Detectors are of two general types, marking and sighting; it is assumed that detectors are positioned so that some individuals are exposed to both types, except in sighting-only data.

The data consists of n detection histories of marked animals, including any resightings, and counts of the sightings of unmarked animals at each detector in each sampling occasion t ($1 \leq t \leq T$). A detection history $\omega_i = \{w_{ik}\}$ codes the status of marked individual i on each occasion with respect to each detector k ($1 \leq k \leq K$). The population density model used is a two-dimensional inhomogeneous Poisson process with intensity $\lambda(\mathbf{s}, \boldsymbol{\beta})$, where \mathbf{s} is the location of a point in the plane and $\boldsymbol{\beta}$ a vector of spatial covariates. It is assumed that the detection declines with distance $d_{ik} = \| \mathbf{s}_i - \mathbf{x}_k \|$ between the detector location \mathbf{x}_k and \mathbf{s}_i , the home-range center of animal i . Let $h(d_{ik}; \boldsymbol{\theta})$, for some vector parameter $\boldsymbol{\theta}$, be the detection function such as the half-normal $h_0 \exp[-d_{ik}/(\sigma^2)]$. Assuming animal independence, the overall likelihood is the product of the likelihoods for the marking and sightings of marked animals, and for unmarked sightings, with common parameters $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$. Maximum likelihood estimates can then be obtained. Details are given by Efford and Hunter (2018).

14.5.2 Lack of Individual Recognition

Chandler and Royle (2013) developed spatial models for situations in which individual recognition is not always possible in a resight model, e.g., avian point count surveys, which involve counting unmarked individuals from multiple points within a study area, and photographs that are not always sufficient for identification due to similar markings among animals. They were able to apply SCR (spatial capture-recapture) concepts to studies of unmarked or partially marked populations, referring to the combined model as the SMR model. As this term conflicts with the resight models in the previous section, we shall use the term SMRU (Spatial Mark-Recapture with Unmarked). The main difference between SMRU and SCR is that SMRU encounter histories are partially latent because only part of the population is uniquely identifiable. We now give the theory of their method.

Let \mathbf{x}_j be the location of the j th trap where term “trap” is anything capable of recording counts of unmarked individuals such as a camera or a human observer conducting a point count survey, and let \mathbf{X} represent all the trap coordinates. Let z_{ijk} represent the encounter data for individual i ($i = 1, \dots, N$) for trap j ($j = 1, \dots, J$) on occasion k ($k = 1, \dots, K$). If an animal can be detected at most once during a sampling occasion, z_{ijk} will be binary, or if individuals can be detected multiple times during a single occasion, $z_{ijk} > 0$ is an integer. In standard CR studies, the z_{ijk} are observed directly for captured individuals, and n_{jk} , the number caught in trap j on occasion k is

$$n_{jk} = \sum_{i=1}^N z_{ijk}.$$

Although N is unknown and the z_{ijk} are unobserved when animals are unmarked, classical data augmentation can be used to sample the latent data $\mathbf{z}_{jk} = (z_{1jk}, z_{2jk}, \dots, z_{Njk})'$ conditional on n_{jk} . For the time being, it is assumed that N is known so that the focus is on the detection process. For the latent encounter data, it is assumed that

$$z_{ijk} \sim \text{Poisson}(\lambda_{ij}), \quad (14.1)$$

where $E[z_{ijk}] = \lambda_{ij}$ is the expected number of captures or detections of individual i in trap j .

Suppose \mathbf{s}_i is the activity center of individual i (also a latent variable), and $d_{ij} = \|\mathbf{s}_i - \mathbf{x}_j\|$. Let $\lambda_{ij} = \lambda_0 g(d_{ij})$, where λ_0 is the encounter rate at $d = 0$, and $g(d)$ is a positive-valued, typically monotonically decreasing, function of distance. It was assumed that

$$g(d) = \exp[-d^2/(2\sigma^2)].$$

As well as σ determining the rate of decay in encounter probability, it also determines the degree of correlation among counts since animals with large home ranges are more likely to be detected at multiple traps relative to animals with small home ranges. This implies that the full conditional of the latent encounter data is

$$\Pr(z_{1jk}, \dots, z_{Njk}) = \frac{n_{jk}!}{\prod_{i=1}^N z_{ijk}!} \prod_{i=1}^N \pi_{ij}^{z_{ijk}},$$

where $\pi_{ij} = \lambda_{ij} / \sum_i \lambda_{ij}$. Since the sum of independent Poisson processes is also Poisson, we have

$$n_{jk} \sim \text{Poisson}(\Lambda_j), \quad (14.2)$$

where $\Lambda_j = \lambda_0 \sum_{i=1}^N g(d_{ij})$, and

$$n_{j\cdot} = \sum_{k=1}^K n_{jk} \sim \text{Poisson}(K \Lambda_j).$$

The activity centers are assumed to be the outcomes of a spatial point process in a state space $\mathcal{S} \subset \mathbb{R}^2$ large enough so that any edge effect can be ignored. It was assumed that the point process intensity is constant, namely, $\mu(\mathbf{x})_{\mathbf{x} \in \mathcal{S}} = N/v(\mathcal{S})$, where $v(\mathcal{S})$ is the area of the state-space so that the homogeneous model $s_i \sim \text{Uniform}(\mathcal{S})$ was adopted. Under this model, animals can move about their activity centers, but the activity centers themselves do not move. Furthermore, the activity centers are assumed to exhibit no attraction or repulsion.

Up till now we have assumed N known. For the unknown case, Chandler and Royle (2013) used a parameter-expanded augmentation where N is replaced by a much larger number M which is large enough for $\Pr(N = M) \approx 0$, but not too large so as to increase computational time unnecessarily. They defined $w_i = 1$ with probability ψ if the i th animal ($i = 1, \dots, M$) is detected (and satisfies (14.1)), and 0 otherwise when $z_{ijk} = 0$ with probability 1. Then

$$z_{ijk} | w_i \sim \text{Poisson}(\lambda_{ij} w_i),$$

and hence $N = \sum_{i=1}^M w_i$ and the population density is $D = N/v(\mathcal{S})$. Representing a prior distribution by square brackets (e.g., $[x|y]$ is the distribution of x given y), and assuming hyperpriors are independent, we have

$$[\psi, \lambda_0, \sigma] \propto [\psi][\lambda_0][\sigma],$$

and the joint posterior distribution takes the form

$$[\mathbf{z}, \mathbf{w}, \mathbf{s}, \psi, \lambda_0, \sigma | \mathbf{n}, \mathbf{X}] \\ \propto \left\{ \prod_{i=1}^M \left\{ \prod_{j=1}^J \prod_{k=1}^K [n_{ijk} | z_{ijk}] [z_{ijk} | w_i, s_i, \sigma \lambda_0] \right\} [w_i | \psi] [s_i] \right\} [\psi] [\lambda_0] [\sigma],$$

Further details of this Bayesian approach are given by the authors in separately published supplementary material including two MCMC algorithms and the full conditionals, along with R code. Borchers and Marques (2017) commented that “the quality of inferences from this method is poor.” However, he noted that they also “considered situations in which only a fraction of the population is marked, which leads to much more reliable inference. This method is likely to be very useful for many camera trap data sets in which a fraction of a population of individuals that are otherwise not individually identifiable can be marked”.

14.6 Camera Methods

Camera traps have already been inferred above and mentioned in several chapters, especially in Chap. 4 on occupancy with some focus on live trapping. We can also “trap” animals through sighting them with camera(s), so this topic is linked to Sect. 14.5 on resighting models. In recent decades there has been an increase in using various noninvasive methods to sample and monitor animal populations without ever physically capturing or handling animals. Two techniques in particular have been very useful, namely, using DNA samples using some indirect trap device and remote cameras (both triggered and non-triggered). Here we focus particularly on camera traps, where under the photographic mark-recapture framework, researchers noninvasively “capture” animals via photograph and identify individuals by their pelt pattern or other natural markings. This is a mark-resight method where after first capture by a camera trap, animals are considered “marked” based on unique natural characteristics. Encounter histories for marked individuals (i.e., those photographed at least once) can be constructed for a series of recapture occasions from which detection probability and abundance can be estimated.

The data from camera traps has been used to estimate densities of elusive terrestrial mammals, animals with low densities, and those animals difficult to capture or detect (O’Connell et al. 2011). One such example is the study of feral cats (Comer et al. 2018). The use of camera traps to study elusive species was pioneered by Griffiths and van Schaik (1993) while studying the activity patterns of rainforest mammals. In addition to studying such things as nest predation, foraging, and habitat use, camera data can be used to estimate animal abundance. This was done initially in the 1990s to estimate tiger abundance. Animals like large felids such as tigers, ocelots, leopards, snow leopards, bobcats, and jaguars can be

uniquely identified by their stripes or spot pattern. Typically the method involves fitting a conventional type of closed population capture-recapture model to estimate abundance and then using ad hoc methods to determine the effective trapping area, which is related to the home range (refsec:). One of the problems with wide ranging and elusive animals is to get an adequate sample. Sollmann et al. (2013a) found a similar problem with studying endangered Florida panthers (*Puma concolor cory*), which also lack clear visual features for individual identification. They combined camera trapping with some telemetry location data so that only a portion of the population needs to be identified.

Camera methods are increasingly popular because of continued technological improvements and decreasing equipment costs, combined with their demonstrated versatility. Two aspects of camera trapping should be considered. Firstly, the traps should not be too far apart so that animals can live in the study area without visiting a trap, but have a chance of encountering a trap. Clearly this will depend on the home range of the animal. Secondly, there is the question of whether the traps should be baited or not. For example, Du Preez et al. (2014) considered the problem of determining camera spacing for leopards (*Panthera pardus*) that have extensive home ranges and are elusive, solitary, and mostly nocturnal. They compared the quality of data produced by both baited and unbaited camera-trapping surveys in terms of total leopard captures and demographic representation and weighed these against cost and effort. It was found that baited camera-trapping significantly increased leopard capture rates, as well as recording dependent cubs, which the unbaited method failed to detect. In addition, the baited method was more cost-effective. One other problem that does not seem to be considered is that there still might be trap addiction or trap avoidance. For example, Wegge et al. (2004) found evidence of a trap effect for tigers, probably because of the photo flash and because they detected the camera traps from cues from impression pads 50 m from the trap. Is this possibly present in other studies?

14.6.1 Some Reviews

Foster and Harmsen (2012) critically reviewed 47 published studies and considered some problems associated with contemporary population estimates of elusive species from camera-trap data. In particular, they discussed (1) individual identification, (2) sample size and capture probability, (3) camera location and spacing, (4) the size of the study area, and (5) ad hoc density estimation from the calculation of an effective trapping area. These are considered in Sect. 14.6.3 under design considerations.

The authors noted that there are several potential problems with camera trap density studies including misidentification of individuals, low capture probabilities, small sample sizes, camera failure, and small study area size. They also referred to the SECR model discussed above that does not require the intermediate step of estimating an effective trapping area. Some of these problems were considered by

Tobler and Powell (2013) with regard to jaguar populations. We note that with low capture probabilities, there has to be a compromise between keeping the survey time short enough to maintain closure, but long enough to catch enough animals. Also cameras have to be close enough so that all animals have home ranges overlapping a camera; otherwise, some animals will not be catchable.

Burton et al. (2015) reviewed 266 studies published between 2008 and 2013 and “recorded study objectives and methodologies, evaluating the consistency of CT protocols and sampling designs, the extent to which CT surveys considered sampling error, and the linkages between analytical assumptions and species ecology.” They mentioned a number of shortcomings of previous studies and we mention just a few. A majority used response variables that ignored imperfect detection (e.g., presence-absence, relative abundance), and 40.2% of the studies did not use a probability-based sampling design, relying instead on opportunistic or targeted approaches for choosing camera location. A further 21.4% 21.4% provided almost no details on sampling design. About 60% of camera trap studies involved multiple species. Few studies estimated density for unmarked species, focussing instead on occupancy modeling or measures of relative abundance. The authors provided a great deal of practical information, and although they were optimistic about using CT technology, they also called for greater transparency.

Caravaggi et al. (2017) carried out a review of camera methods with an emphasis on animal behavior and discussed some limitations. A review of the history of wildlife camera trapping as a survey tool in Australia was given by Meek et al. (2015). A general overview of camera traps in ecology is given in the book by O’Connell et al. (2011), who focussed primarily on the use of camera traps in the conduct of science and management. The book considers history, equipment, animal behavior, abundance methods, spatial models, occupancy, community dynamics, species methods, and the future of camera traps.

14.6.2 *Camera Advantages*

A major advantage of camera traps is that they can operate for 24 hr/day and they can be left unattended for considerable time. They are particularly useful for collecting quantitative records on elusive, nocturnal, and cryptic species with relatively low labor costs (Burton et al. 2015), and they can also be used by volunteers. For species with individually recognizable markings (i.e., marked populations), a combination of camera trapping with capture-recapture analyses has been a common approach. Camera data has been used to estimate various measures of the overlap of daily activity patterns (Ridout & Linkie 2009) as well as to measure animal travel speed and activity range (Rowcliffe et al. 2014; Palencia et al. 2019; Ait Kaci Azzou et al. 2021). They can also be used for multiple species to assess any interactions, such as competing for prey (Schliep et al. 2018). An R package called **camtrapR** that is available for handling camera-trap data management was given by Niedballa et al. (2016).

Camera trap methods along with resighting are cost-effective methods for elusive animals, and camera methods generally have high-detection rates. Sometimes it is helpful to have more than one camera at a site to capture both sides of an animal for identification (e.g., McGregor et al. 2015: feral cats). Technology has also improved the use of camera traps. For example, McClintock et al. (2013) used bilateral photo-identification records in capture-recapture analysis, which links left- and right-side photos to the same individual by assuming that the true encounter history for each animal is a latent (unobserved) realization from a multinomial distribution.

14.6.3 Design Considerations

There are some potential problems associated with camera methods raised by Burton et al. (2015) with camera trapping, which we now consider.

Individual Identification

When using individual identification, it is essential that unique markers are visible for all members of a sample as some studies use approximate methods. Often two cameras are needed to photograph both sides of an animal for unique identification. The number of photographs that are unidentifiable should always be recorded along with whether or not the animals are included in the study; clear information needs to be provided. A method for modeling such error in a closed population given by Yoshizaki et al. (2009) is described below. As with capture-recapture (CR) models, the new estimators there performed well if the capture probability was greater than 0.2; however, most published studies report smaller capture probabilities. Using two or more observers will help with identification. Resighting methods can also be used provided sampling with replacement is allowed, as a camera may photograph an individual multiple times (e.g., Watts et al. 2008; Jordan et al. 2011).

Sample Size and Capture Probability

For closed populations, sample sizes of 10–20 individuals may be too small for reliable estimates using C-R. However, in 47 published camera-trap studies, the sample size was reported for 122 of 125 data sets (45 studies). The mean sample size tended to be around 10–15 with a wide range of coefficients of variation (CV) of 0.2% to 175%. Such numbers are not adequate for monitoring population trends through time or for between-site comparisons. Also low capture probabilities make it difficult to select an appropriate C-R model, though the heterogeneity model M_h is commonly used. The jackknife estimator for it (Sect. 12.5.5) will underestimate abundance if capture probabilities are less than about 0.05, which happened in some of the studies.

Camera Location and Spacing

Trap spacing and location choice are both important. One recommendation (Dillon and Kelly 2007) is that camera spacing should seek to maximize the capture probability by including at least 2 traps per average home range. Traps which are

spaced too widely may fail to detect individuals if they occupy home ranges that fall between trap locations, breaking the assumption of conventional CR models that the probability of capture of every individual is greater than zero. Regular arrays of traps, which maximize the ratio of study area to its perimeter, are preferable for conventional C-R methods using ad hoc density, as such an array will reduce the edge effect caused by the overlap of home ranges of the study animals outside the study area. Camera traps can fail through various reasons, which results in variable trap effort and needs to be reported. Camera locations should be randomized, and stratification is sometimes needed when there is spatial variation in capture probability and different habitats. Care is needed in using the same survey design for multiple species as different species may have different trap spacing requirements.

Sollmann et al. (2012) investigated how the spatial arrangement and size of the trapping array influenced parameter estimates for SCR models. Their simulated study using black bear data showed that SCR models performed well as long as the extent of the trap array was similar to or larger than the extent of individual movement during the study period, and movement was at least half the distance between traps. SCR models performed well across a range of spatial trap setups and animal movements. Contrary to non-spatial capture-recapture models, they do not require the trapping grid to cover an area several times the average home range of the studied species.

Sun et al. (2014) simulated black bear (*Ursus americanus*) populations and spatial capture-recapture data to evaluate the influence of trap configuration and trap spacing on estimates of population size and a spatial scale parameter, σ , that relates to home range size. For sampling over K sampling periods, let y_{ij} be the number of encounters for an individual i in trap j , and assume $y_{ij} \sim \text{Binomial}(K, p_{ij})$, where p_{ij} is the encounter probability and satisfies

$$p_{ij} = p_0 \exp[-d_{ij}^2/(2\sigma^2)].$$

Here d_{ij} is the distance between individual i and trap j and

$$p_0 = \exp[\alpha_0]/(1 + \exp[\alpha_0]).$$

A maximum likelihood approach was used to estimate N the population size and σ the spatial scale parameter. Three trap configurations were considered, and it was found that the root mean squared error depended on the relationship of the effective trap spacing to σ with traps ideally spaced no more than 2σ . The clustered configuration worked well and was sometimes the best. They concluded that in determining the sampling design, attention needs to be given to the size of the study area, the range of individual movement, and home range size.

Size of Study Area

Generally the study area should be large, but may be constrained by the costs and logistics of camera trapping. Various studies have reported that small survey areas

give inflated density estimates, which can be due to a number of factors relating to greater animal movements for wide ranging species, home range overlaps, and lack of habitat variation.

Effective Trapping Area (ETA)

Methods of adding a buffer strip of width w to the boundary of the sampled area have been considered elsewhere. However, the authors recommended merging a circular buffer (radius w) around each camera location to give a more realistic estimate of the ETA. Telemetry can help to base w on the average home range size. Otherwise camera trap studies have tended to use $w = 0.5 \times \text{mean maximum distance moved}$ of individuals captured at multiple camera stations. Clearly suitable methods for estimating w for such a method are needed. The authors, Burton et al. (2015), noted that the assumption of independent activity centers is perhaps unrealistic for territorial carnivores.

14.6.4 Partial Identity

Augustine et al. (2018) focussed on partial identity where, for some of the individuals, identity is only known from a single flank photo taken from single cameras, instead of deleting uncertain information. We now consider their model. Given N activity centers \mathbf{s}_i distributed uniformly over region \mathcal{S} of N individuals which are represented by an $N \times 2$ matrix, and let \mathbf{x}_j be the coordinates of the j th trap, giving a $J \times 2$ matrix \mathbf{X} . Let \mathbf{v} be a vector of length J containing the number of cameras deployed at each trap (either 1 or 2), and define the events $m \in \{B, L, R\}$ to correspond to both-side simultaneous capture, left-only capture, and right-only capture, respectively. Then, the binomial capture process for the m th capture type is $Y_{ijk}^{(m)} \sim \text{Bernoulli}(p_{ijk}^{(m)})$, where $p_{ijk}^{(m)}$ is the capture probability of individual i at trap j on occasion k ($k = 1, \dots, K$) for event type m . This process produces the partially latent complete capture history, a set of binomial frequencies $\mathbf{Y}_{ijk} = (\mathbf{Y}_{ijk}^{(B)}, \mathbf{Y}_{ijk}^{(L)}, \mathbf{Y}_{ijk}^{(R)})$, all of dimension $N \times J \times K$.

The observed capture history is the set of binomial frequencies $\mathbf{y}_{ijk} = (\mathbf{y}_{ijk}^{(B)}, \mathbf{y}_{ijk}^{(L)}, \mathbf{y}_{ijk}^{(R)})$, almost certainly not in the same order along the i dimension as the complete capture history for the L and R capture types. The dimensions of these three binomial frequencies are $n_B \times J \times K$, $n_L \times J \times K$, and $n_R \times J \times K$, respectively, with n_m being the number of individuals for which at least one m event was observed. Because double cameras are typically positioned to fire together in order to get a both-side capture, it is unlikely that each camera operates independently. Rather than model the capture probabilities of each of the two cameras separately with some correlation between cameras, the authors specified different detection functions for both-side captures and single-side captures, assuming independence between the both-, left-, and right-side capture processes as follows.

The Gaussian hazard detection function for capture type m is

$$p^{(m)}(\mathbf{s}, \mathbf{x}) = 1 - \exp[h^{(m)}(\mathbf{s}, \mathbf{x})], \quad (14.3)$$

where \mathbf{s} and \mathbf{x} are generic activity centers and trap locations, respectively, and using the distance between them,

$$h^{(m)}(\mathbf{s}, \mathbf{x}) = \lambda_0^{(m)} \exp[-\|\mathbf{s} - \mathbf{x}\|^2 / (2\sigma^2)],$$

where $\lambda_0^{(m)}$ is the expected number of detections for capture type m for a zero distance. Because we do not expect any systematic difference in the probability of detecting one flank over the other, we can set $\lambda_0^{(S)} = \lambda_0^{(R)} = \lambda_0^{(L)}$, where S indicates a single-side capture. For single camera stations, we have the single-side detection function $p^{(1S)}(\mathbf{s}, \mathbf{x})$ for both L and R captures. At double camera stations, the single-side capture probability is

$$p^{(2S)}(\mathbf{s}, \mathbf{x}) = 1 - (1 - p^{(1S)}(\mathbf{s}, \mathbf{x}))^2$$

because there are now two ways to photograph a single side (camera 1 or 2). Here $p^{(1S)}(\mathbf{s}, \mathbf{x})$ corresponds to (14.3) for a single-side capture. We note that B captures can only occur at double camera stations; so, we introduce $\lambda_0^{(B)}$ as the expected number of both-side observations for zero distance of the activity center from trap. Finally, the single- and both-side capture probabilities at each trap depend on the number of cameras deployed, giving

$$p^{(S)}(\mathbf{s}, \mathbf{x}) = p^{(1S)}(\mathbf{s}, \mathbf{x})I_{(v_j=1)} + p^{(2S)}(\mathbf{s}, \mathbf{x})I_{(v_j=2)}$$

and

$$p^{(B)}(\mathbf{s}, \mathbf{x}) = 0I_{(v_j=1)} + p^{(B)}(\mathbf{s}, \mathbf{x})I_{(v_j=2)}.$$

Data augmentation was used to both estimate the unknown N and resolve the unknown complete identities of the partial identity observations. Further details of the computation are given by the authors.

Augustine et al. (2019) noted how the above models of spatial mark-resight (SMR) and the previous spatial partial identity model (SPIM) extend SCR models that do not always allow for individual identity to be determined with certainty. They probabilistically resolved the uncertainty in individual identity using the spatial location where samples were collected. The level of information about individual identity that a spatial location contains is related to two key ecological concepts, population density and home range size, particularly related to the magnitude of home range overlap, which can be quantified using a proposed Identity Diversity Index (IDI). They showed that the SPIM models can produce imprecise and biased density estimates in many high IDI scenarios when data are sparse. They extended

such models to so-called “categorical SPIM” models that incorporated categorical, partially identifying covariates, which reduce the level of uncertainty in individual identity, increasing the reliability and precision of density estimates, and allowing reliable density estimation in scenarios with higher IDI values and with more sparse data.

14.6.5 Some Applications

Rowcliffe et al. (2008) developed a random encounter model (REM) where the individual recognition of animals is not required. It uses a two-dimensional ideal gas model in which a particle moving in space covers an area that is the product of the width of the particle and the total distance moved in a given time. The method depends on two key biological variables (average animal group size and day range) and two characteristics of the camera sensor (distance and angle within which it detects animals). The method requires camera sensitivity to be quantified so as to estimate the area effectively monitored by cameras, which was considered by Rowcliffe et al. (2011).

Nakashima et al. (2018) also considered a model where there is no individual recognition called the random encounter and staying time (REST) model, to estimate animal density. The REST model describes the relationship among population density, mean number of detections by a camera trap during a survey period, and staying time of individual animals in a predetermined detection zone in which individuals are certain to be detected by the camera trap.

Parsons et al. (2015) described using additional ancillary information from continuous video and radio telemetry to evaluate the assumptions of mark-resight models for abundance estimation on a barrier island raccoon (*Procyon lotor*) population using camera traps. The study site was geographically closed, allowing the estimation of real survival and *in situ* recruitment in addition to population size. They found that bait presence, home range length, marsh habitat, camera location relative to home range center, sex, and age class all contributed to individual heterogeneity in resighting probabilities.

Popescu et al. (2014) discussed the question of when can different data types be combined and used generalized linear models and generalized linear mixed-effects models to relate camera trap probabilities for marked animals to independent space use from telemetry relocations using 2 years of data for fishers (*Pekania pennanti*) as a case study. Home range and proximity analyses were carried out.

Furnas et al. (2017) used detection/non-detection data from baited camera stations with auxiliary information on home range sizes to estimate the population size and density of the mammal fisher, which is of conservation concern. They used the detection RN model of Royle and Nichols (2003), considered in Sect. 4.1.5.

Under this model, the probability of detecting occupancy of a species at site i and survey j is modeled as

$$P_{ij} = 1 - (1 - r_{ij})^{N_i},$$

where N_i is the local site abundance and r is the probability that a particular individual is detected. Abundance N_i is assumed to be modeled as Poisson(λ), while λ was modeled as a log-linear model and r_{ij} as a logistic linear model using covariates. The home range was modeled from previous data using the 100% minimum convex polygon method with at least 10 data points (and 30 later). They fitted a generalized linear mixed model with a log-link function and normally distributed errors to assess how home range size varied by gender and distance-to-coast covariates using **R** statistical software with the glmmPQL function. This was used to estimate the effective survey area, and bootstrapping was used to compute average density. Some model assumptions were tested.

14.6.6 Community Method

Tobler et al. (2015) developed a multi-session multi-species occupancy model for obtaining estimates for species richness and occupancy, combining data from multiple camera trap surveys (sessions). By estimating species presence at the session level and modelling detection probability and occupancy for each species and sessions as nested random effects, parameter estimates could be improved for each session, especially for species with sparse data. Previous multi-species approaches were extended by including one additional level of random session effects, thus allowing for heterogeneity of detection and occupancy for each species across sessions. The authors developed two variations of their model, a binary latent states model and a Royle-Nichols (RN) formulation for the relationship between detection probability and abundance.

Method 1

Let y_{ij} be the number of sampling occasions out of a total of k_j sampling occasions when the station was operating, that the species i was detected at camera station j . The authors used the term “session” for each primary sampling period (camera trap survey, i.e., the total duration cameras were active at a particular study site in a particular year/season), while “sampling day” will be used for the secondary period. Since every station can only be part of one session, they did not include the session index t for y and k but instead modeled session as a random effect on occupancy and detection. The model for the observed data contained three nested hierarchical levels: (1) the session-level community (i.e., all species occurring at a study site during a session), (2) the station-level community (occurrence process determining the species present in the vicinity of each camera trap location within a study site), and (3) the detection process. Let $w_{it} = 1$ if species i is present during session

t with probability Ω_i , and let $z_{ij} = 1$ if species i uses the area around station j , given available, with probability ψ_{ij} . We now have the following sequence of distributions:

$$\begin{aligned} w_{it} &\sim \text{Bernoulli}(\Omega_i) \\ z_{ij} &\sim \text{Bernoulli}(w_{it} \psi_{ij}) \\ y_{ij} &\sim \text{Binomial}(k_j, z_{ij} p_{ij}), \end{aligned}$$

where through the variable k_j one can use the actual number of days each camera was operating and accommodate sessions of different lengths. As ψ_{ij} and p_{ij} are species specific, they will rarely be estimable for all species in an inventory, because of a lack of data. To overcome this limitation, the authors treated both of these parameters as random effects having a community-level prior distribution. Covariates were used to distinguish between cameras set on trails and off trails for $\text{logit}(p_{ij})$ and $\text{logit}(\psi_{ij})$ with random effects. To estimate the derived parameter N_t , the total number of species present during each session, we can use

$$N_t = \sum_i w_{it}.$$

Method 2

Instead of using z_{ij} , abundance a_{ij} is modeled by a Poisson process with rate parameter λ_{ij} representing the number of species using the area around a station, or more generally an indicator for an unexplained preference for a specific station by one or multiple individuals. Also p_{ij} is replaced by a function of r_{ij} , the per-individual detection probability, and of a_{ij} , the total number of individuals of species i exposed to detection at site j . This leads to the sequence:

$$\begin{aligned} w_{it} &\sim \text{Bernoulli}(\Omega_i) \\ a_{ij} &\sim \text{Poisson}(\lambda_{ij}) \\ y_{ij} &\sim \text{Binomial}(k_j, 1 - (1 - r_{ij})^{a_{ij} w_{it}}), \end{aligned}$$

along with covariate models for $\text{logit}(r_{ij})$ and $\text{logit}(\lambda_{ij})$ with random effects. We can also express $\psi_{ij} = 1 - \exp[-\lambda_{ij}]$.

All models were implemented in a Bayesian framework using the BUGS language and run in software JAGS (Plummer 2003). To estimate the derived parameter N_t , the total number of species present during each session, we can use

$$N_t = \sum_i w_{it}.$$

A goodness-of-fit test indicated that the second RN model performed better, especially with pooling. The poorer fit was due to overdispersion for several species,

generally due to a few cameras having a much larger than expected number of detections than the rest.

14.6.7 Spatial Point Process Model

As described by Miller et al. (2019), a spatial point process describes the distribution of event locations across some spatial domain (e.g., locations of individuals of the same species). Points arise from a random process, described by the local intensity λ_s , which measures the expected density of points at a given location, s , in space. If points arise independently and at random, we can describe the local density as a homogeneous Poisson distribution called the Poisson point process. If event locations are independent but the intensity varies spatially, the distribution is referred to as an inhomogeneous point process as λ_s varies. If all events are not observed, such as is the case when sampling and detection are incomplete, the observed points come from a thinned point process.

An advantage of a spacial model is that it is possible to link observations across different spatial scales. A useful result is that if occurrence is modelled as a function of local intensity, occupancy is no longer scale dependent, but instead can be determined based on cell area. The authors gave a broad-brush approach for setting up models with the structuring of joint likelihoods for different types of data. They incorporated spatial covariates, species distributions, and correlation and surveyed a large number of papers combining various types of data.

Royle and Gardner (2011) commented that “the deficiency with classical closed population models for estimating density from trapping arrays is that ‘space’ has no explicit manifestation in such models” as trap spacing will play an important role and suggested using spatial point process models (Efford 2004). Let y_{ijk} be the detection frequency of catching individual i ($i = 1, \dots, n$) in trap j ($j = 1, \dots, J$) during occasion k ($k = 1, \dots, K$); for example y_{ijk} may be binary taking values 0 or 1. Otherwise the data could arise from bear hair snare studies (and other DNA-based sampling) where an individual can be encountered a number of times during any period, and the biological material (hair, etc.) accumulates but cannot be partitioned into distinct visits after it is collected.

A natural choice for the observation model is the Poisson distribution $y_{ijk} \sim \text{Poisson}(\lambda_0 g_{ij})$, where g_{ij} is a function of d_{ij} the distance of the activity center of individual i from trap j , and λ_0 is the expected number of captures in a trap located precisely at an individual’s activity center. A popular choice for g_{ij} is the half normal, namely, $g_{ij} = \exp[-d_{ij}^2/\sigma^2]$. We then have a generalized linear model (GLM)

$$\log(E[y_{ijk}]) = \alpha + \beta d_{ij}^2,$$

where $\alpha = \log(\lambda_0)$, $\beta = -(1/\sigma^2)$, and d_{ij} is a random effect.

When the data are actually binary, where the y_{ijk} take values 0 or 1, the authors assumed that

$$\Pr(y_{ijk} = 1) = \pi_{ijk} = 1 - \exp[-\lambda_0 g_{ij}],$$

the probability $y > 0$ under the Poisson encounter frequency model. Here we have a binomial GLM model

$$\log(-\log(1 - \pi_{ijk})) = \alpha + \beta d_{ij}^2,$$

where α and β are the same as before.

Another model the authors considered is a multinomial model, which applies to the situation where an individual can only be captured at most a single time during a sample occasion. Using the Poisson encounter model with independent traps, the total number of captures for each individual during each period (i.e., when we sum over traps) is also a Poisson random variable (by the additivity of Poisson random variables), namely,

$$y_{i \cdot k} = \sum_j y_{ijk} \sim \text{Poisson}(\lambda_0 \sum_j g_{ij}).$$

Using another property of the Poisson distribution implies that the $y_{i1k}, y_{i2k}, \dots, y_{iJk}$ conditional on $y_{i \cdot k}$ is multinomial with index $y_{i \cdot k}$ and probabilities $g_{ij} / \sum_j g_{ij}$. From the Poisson model, define

$$\bar{p}_i = \Pr(y_{i \cdot k} > 0) = 1 - \exp[-\lambda_0 \sum_j g_{ij}].$$

Then if \mathbf{y}_{ik} is a multinomial observation for individual i during occasion k , we can describe the cell probabilities of this $J + 1$ dimensional multinomial distribution by

$$\pi_j = \bar{p}_i \frac{g_{ij}}{\sum_j g_{ij}} \quad j = 1, \dots, J,$$

and for the last cell corresponding to “not captured,” $1 - \bar{p}_i$. The Poisson, binomial, and multinomial modes were dealt with individually along with some extensions.

14.6.8 Modeling Identification Errors

Photographic tags have been used for a wide range of animals such as cheetahs, pythons, adders, snow leopards, crocodiles, tigers, tigers salamanders, leafy sea dragons, and marine mammals, usually from natural markings or from wear and tear such as scars. There is misidentification when natural marks from the same

individual cannot be matched. There are two main causes of this. The first relates to the quality of the photographs and the matching method (e.g., the photo is not taken from the right angle) and changes in the natural marks (e.g., loss of scars or further scars) referred to as the “evolving natural tags” model (EV), which is modeled below. Computer programs are available to aid the matching process (e.g., Arzoumanian et al. 2005). Before getting into details with evolving tags, suppose we have two samples. Denoting capture by a 1 (and no capture by a 0), then if an animal is captured in both samples, it has a real capture history (11). However, if it is correctly identified at sample 1 and misidentified at sample 2, two dependent capture histories (10) and (01) that cannot be distinguished from the data and having the same probability are generated, with the second history being a ghost capture history. There can of course be real histories (10) and (01) with their own ghost histories.

Yoshizaki et al. (2009) used the three-sample CR model (M_t) approach to deal with the situation where misidentification is due to changes in natural marks, but not to the quality of photographs. It is based on the following assumptions.

- (1) Marks are unique and there are no unmarkable animals in the population.
- (2) Misidentification can occur only at recapture events. (i.e., not at the first capture event).
- (3) Misidentification of an animal leads to creating a unique, new identity (the misidentification of two different animals never leads to the same new false identity).
- (4) Once a new identity is created, the old identity is not subject to recapture, but the new, false, identity is subject to recapture (and is different from other animals).
- (5) An animal is captured only once at a given sampling occasion.

Let β be the probability of correct identification for a captured animal, and assume that β is constant for all animals at each sampling occasion. This parameter applies only to recaptured animals and can also be viewed as the probability that the existing natural mark does not change between two consecutive sampling occasions. To consider misidentification in a statistical modeling framework, it is helpful to consider capture histories in terms of the true capture history (which describes the sequence of capture and non-capture events of an animal during the study) versus the generated capture history (which is the list of capture events that can be observed as data). The “real” capture history with incorrect identification and its “ghost” capture history are always observed together, but as if they were different animals.

Let ω denote a true capture history and Y_ω the number of animals with that true capture history. The generated capture history, denoted by γ , is the history we observe, and the corresponding random variable X_γ denotes the number of such histories observed. With misidentification possible, the observed capture history will agree with the true capture history only if identifications at all capture events are correct. For example, an animal captured on all sampling occasions will have true capture history (111). If misidentification occurred only at the second sampling occasion, i.e., the change in the natural mark occurred between the first and second sampling occasions, two observed histories will be recorded, namely (100, 011). For

an animal with true capture history (111), the probability of observing the original, real capture history (100) is $p_1 p_2 (1 - \beta) p_3 \beta$. The ghost capture history (011) that is created due to misidentification at the second capture event also has the same probability $p_1 p_2 (1 - \beta) p_3 \beta$. Note that when a 0 is recorded in the real history because of misidentification, the probability of the event is $p_i (1 - \beta)$ rather than $(1 - p_i)$ as in models without misidentification. Also, the real capture history (100) cannot include a recapture at the third sampling occasion because misidentification occurred at the second sampling occasion, whereas the ghost capture history does have the possibility of recapture at the third sampling occasion. In a similar manner, we can list every possible generated history γ for different true histories and obtain their probabilities, as in Table 1 of Yoshizaki et al. (2009). Note that the sum of the marginal probabilities of all the possible capture histories $Pr(\text{gamma})$ will exceed 1 because of the ghost capture histories.

The marginal probabilities for generated capture histories, $Pr(\gamma)$, are obtained by summing all probabilities for each distinct generated capture history (Table 2 for three samples from Yoshizaki et al. 2009). Let $P'_k = p_k \beta$ be the probability of observing an animal with correct identification at sampling occasion k ; then $Q'_k = (1 - p_k) Q'_{k+1} + p_k (1 - \beta)$ is the probability of not observing an animal after occasion $k - 1$, with $Q'_4 = 1$. For example, for (111), the probability is $p_1 P'_2 P'_3$, while for (101) we have $p_1 q_2 P'_3$. The authors gave expressions for $Pr(\gamma)$ in the general case of a K -sample experiment. Estimation was carried out using estimates obtained by either unweighted least squares or minimum chi-squared based on a function of the squared differences $(x_\gamma - E[X_\gamma])^2$ using $E[X_\gamma] = N Pr(\gamma)$, where x_γ is the observed count and N is the population size.

Alternatively, N can be replaced by x_{all}/p^* , where x_{all} is the total number of capture histories observed and, for $K = 3$, $p^* = \text{total} - Pr(\gamma = 000)$. Then N is estimated by $\widehat{N} = x_{\text{all}}/\widehat{p^*}$. The authors found that if misidentification is ignored (i.e., $\beta = 1$), the model M_t estimator of N is positively biased even if $1 - \beta$, the probability of misidentification, is only 0.05, and the magnitude of the bias increases rapidly with the probability above 0.1. Yoshizaki et al. (2009) concluded that for misidentification rates of 5% or greater and moderately high capture probabilities, it is clear that the above estimators perform better than the M_t estimator.

14.7 Adaptive Cluster Sampling

Considered integrating adaptive cluster sampling and spatial occupancy modelling by developing two models to handle the dependence induced by cluster sampling. These two models were first compared individually and then combined. As with adaptive sampling, there are two sampling stages with the second stage depending on what is found in the first stage (Seber & Salehi 2013). We begin with n spatial sites selected randomly from the population area, with J samples taken at each site. Let s_{i1} and $y_{i1} \in \{0, \dots, J\}$ denote the spatial location and number of detections, respectively, for sampling site $i = 1, \dots, n$. Additional sampling effort is dedicated

to areas (“neighborhoods”) near locations s_{i1} with detections ($y_{i1} > 0$). For each location with a detection, an additional $k_i - 1$ sites s_{i2}, \dots, s_{ik_i} near s_{i1} are surveyed, again with J sampling occasions at each site. Each selected sample location is surveyed the same number of times, even if the same location is selected more than once by adaptive cluster sampling. We allocate each site to the first cluster it is assigned to and resolve any issues with individual sites appearing in multiple clusters.

The first question is how do we define a “neighborhood” as this can be done in several ways. The authors suggested, for example, a neighborhood consisting the four rook neighbors (neighbors to the north, east, south, and west) of occupied sites in the second stage (with each cell attributed to at most one cluster). Using this case, let $z_{ij} = 1$ if site i is occupied and 0 otherwise, for $i = 1, \dots, n$ initial sites and $j = 2, \dots, 5$ second-stage sites. The y_{ij} detections satisfy

$$y_{ij} \mid z_{ij} \sim \text{Binomial}(J, pz_{ij}),$$

where p is the usual conditional probability of detection given present at a site, and $\Pr(Y_{ij} = 0) = 1$. It is assumed there is a latent continuous spatial process v_{ij} so that the site is occupied if $v_{ij} > 0$, and the site is not occupied if $v_{ij} \leq 0$.

For the spatial occupancy model, the latent process is modeled as a Gaussian process with $E[v_{ij}] = \mathbf{x}'_{ij}\boldsymbol{\beta}$, where the \mathbf{x}_{ij} are covariates. Also $\text{var}[v_{ij}] = 1$ and spatial correlation $\text{cov}[v_{ij}, v_{uv}] = M(\|s_{ij} - s_{uv}\|)$, where $\|s_{ij} - s_{uv}\|$ is the distance between the sites and M is the Matérn covariance function. The Matérn correlation function has two parameters: $\eta > 0$ controls the smoothness of v , and $\rho > 0$ controls the range of spatial dependence. Under this model, the occupancy probability at site s_{ij} has the usual probit regression form $\Phi(\mathbf{x}'_{ij}\boldsymbol{\beta})$, where Φ is the standard normal distribution function.

For the random-effects occupancy, the authors used a simpler model, which accounted for dependence between sites in the same sampling cluster using a shared random effect. In this model, the latent process is modelled as

$$v_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + b_i + \varepsilon_{ij},$$

where the cluster random effects are $b_i \sim N(0, \sigma^2)$ and $\varepsilon_{ij} \sim N(0, 1)$, both independent over i and j . Further details of this model using Bayesian methods are given by the authors.

We see from above that spatial autocorrelation, when detectability is more similar among neighboring than distant detectors, can be a problem. If ignored it can lead to considerable estimation bias (Moqanaki et al. 2021)

14.8 DNA Methods

With technological advances, DNA typing has become easier and cheaper to apply and has been used as a natural genetic tag to investigate a wide range of animals. Lukacs and Burnham 2005, Beja-Pereira et al. 2009, and Lampa et al. 2013, gave a number of practical details on carrying out a DNA experiment. It can be particularly useful with difficult-to-access sites. However, the method does have a few problems such as deterioration of the material that it has been taken from through environmental factors such as moisture and UV radiation, as well as some possible matching problems. For example, errors can occur if too few or insufficiently variable loci are used resulting in different animals seeming to be the same animal, the so-called “shadow” effect, so that the population size is underestimated. With low quality and quantity of DNA available, there can be genotyping errors due to false alleles as well as specimens from the same individual seeming to have different DNA, all leading to an overestimation of population size. There can also be (1) contamination (from other individuals), (2) sequencing errors that can arise through “amplification failure” when there is a copying failure of a piece of the genome, (3) allelic dropout when one allele at a locus fails to “amplify” or is not present in the pipetted DNA sample, (4) mutations during amplification, and (5) individual heterogeneity such as variability in cell-shedding rates.

Capture-recapture methods involve taking DNA samples at several points in time often noninvasively such as simply collecting samples like animal droppings (passively) or through contact (actively) like collecting hair samples using barbed wire around trees and posts. There can be some problems with noninvasive methods such as with hair snares as some animals may encounter a snare but not leave a fair sample. Also poor quality hair samples will not be genotyped, and not all genotyped hair samples will provide a unique identification. This type of model with a spatial adaption has been referred to as the spatially explicit capture-recapture (SECR) model, as previously described.

Boulanger et al. (2018) applied the SECR model using large-scale DNA mark-recapture inventories and hair snag sites that were completed for 5 of 7 bear management areas in Alberta from 2004 to 2008. Their approach modeled the distribution of estimated home range centers within the sample grid using scale of movement parameters estimated from repeat detections of animals during the time of sampling. Closed telemetry models were also considered. Gerber et al. (2012) used photographic-sampling data of the carnivore Malagasy civet *Fossa fossana* collected with and without a lure to assess the effects of the lure and to compare the use of four density estimators that varied in methods of area estimation. They raised two potential problems with capture-recapture, namely, uncertainties arising from bait or lure to attract animals to the detection device and ad hoc boundary-strip methods to compensate for edge effects in area estimation. The second problem can be avoided with an SECR model with the consequent added advantage of direct density estimation, even though precision of estimation is reduced because spatial variation is incorporated.

14.8.1 Comparing DNA and Camera Methods

Cheng et al. (2017) compared traditional live trapping with noninvasive genetic sampling (NGS) to see how they compared for a relatively common species like the snowshoe hare (*Lepus americanus* Erxleben). They found that NGS provided population estimates similar to those derived from live trapping, but a higher density of sampling plots was required for NGS. Also, while live trapping does not incur laboratory costs, its field costs can be considerably higher than for NGS, especially when study sites are difficult to access. They concluded that “NGS can work for common species, but that it will require field and laboratory pilot testing to develop cost-effective sampling protocols.”

Burgar et al. (2018) compared both camera trap and DNA spatial models using data from a small, recovering mammal population of the mesocarnivore fisher (*Pekania pennanti*) building on a previous study (Linden et al. 2017). They used genetic tagging via hair sampling combined with infra-red remote cameras. Their study goals were to estimate fisher density from a spatial capture-recapture (SCR) model using noninvasive genetic survey data, to evaluate the ability of a spatial count (SC) model using concurrent camera trap survey data to produce comparable estimates, and to compare the costs and benefits associated with both sampling methods. The SC density estimates were influenced by their priors so they strongly recommended informative priors should be avoided, recommending instead using a range of unweighted prior knowledge. Thin detection data was problematic for both SCR and SC models, potentially producing biased low estimates. The total cost of the genetic survey was two-thirds of the camera trap survey. They ran models using JAGS (Plummer 2003), interfacing through R using the *jagsUI* (Kellner 2015) and the *rjags* (Plummer 2016) packages for the SCR and SC models, respectively.

14.8.2 Combining DNA and Camera Methods

Gopalaswamy et al. (2012) used two approaches, a direct approach and a two-step approach, to estimate the density of the rare and elusive tiger, a sparsely populated predator. They combined both photographic capture-recapture and fecal DNA collected during the same season at a single site using a spatial capture-recapture (SCR) model. The direct approach was developed for the situation in which raw data from two independent surveys are available. Looking at the problem from a very general point of view, they supposed D_P and D_G are data from the photographic and genetic surveys, respectively, which are expected to be informative about three sets of parameters, θ_P , θ_G , and θ_B , where θ_P and θ_G are parameters exclusively informed by data sets D_P and D_G , respectively, and θ_B is the set of parameters informed by both data sets. A joint distribution is given by

$$f(D_P, D_G | \theta_P, \theta_G, \theta_B) = f_P(D_P | \theta_P, \theta_B) f_G(D_G | D_P, \theta_P, \theta_G, \theta_B).$$

In the two-step approach, where the focus is on prior information, the authors begin with a prior distribution for the parameters

$$\pi(\boldsymbol{\theta}_P, \boldsymbol{\theta}_G, \boldsymbol{\theta}_B) = \pi_{PB}(\boldsymbol{\theta}_P, \boldsymbol{\theta}_B)\pi_G(\boldsymbol{\theta}_G),$$

and obtain the joint posterior density

$$\pi(\boldsymbol{\theta}_P, \boldsymbol{\theta}_G, \boldsymbol{\theta}_B | D_P, D_G) \propto \pi(\boldsymbol{\theta}_P, \boldsymbol{\theta}_B | D_P)\pi_G(\boldsymbol{\theta}_G)f_G(D_G | D_P, \boldsymbol{\theta}_P, \boldsymbol{\theta}_G, \boldsymbol{\theta}_B).$$

Here the product $\pi(\boldsymbol{\theta}_P, \boldsymbol{\theta}_B | D_P)\pi_G(\boldsymbol{\theta}_G)$ represents what we know about all the model parameters at the end of the first phase (i.e., the photographic survey). The data D_G from the second phase is then incorporated. The two survey types represent two observation processes from a common state process involving activity centers. However, there were some problems as it was not possible to associate the photograph of an individual tiger with its corresponding fecal DNA sample, and there was a time gap of approximately six weeks between the two surveys. Data was sparse, and Bayesian methods were used along with data augmentation; some details are given by Gopalaswamy et al. (2012).

Sollmann et al. (2013) also developed a combined model using separate Poisson models for observed data, with means that are sex specific and that depend on distance to activity centers and on effort. A Bayesian method is used along with data augmentation for analysis. Efford et al. (2013) considered a model with known variable effort. Examples of possible effort variables are the number of days that each automatic camera was operated in a tiger study, or the number of rub trees sampled for DNA in each grid cell of a grizzly bear study.

Open population methods for tigers have also been used by, for example, Karanth et al. (2006) using Pollock's robust design (Seber & Schofield 2019, chapter 11).

In Sect. 4.11 we considered a paper by Fisher et al. (2016) about occupancy to estimate missed detections when an animal does not leave a hair sample with viable DNA by sampling the hair trap and the area around it using camera-trapping.

14.9 Summary

Estimating abundance using capture-recapture methods for animals that move around in a closed population can be difficult. One way of dealing with this, as well as for some stationary populations, is to record the locations or centers of activity of the animals or objects (e.g., animal signs) using so-called spatial capture-recapture (SRC) methods. The probability of capture can then be modeled on the distance of an animal from a trap location.

The “trap” can be any kind of recording device such as a personal observation or camera or an array of proximity or acoustical detectors. It can also be a single- or multi-catch trap. Sometimes telemetry can be used at the same time and can be combined with the capture data. SRC can also be combined with occupancy and

distance sampling data, as well as with adaptive cluster sampling. Information on activity centers can be used to study animal interactions.

Bayesian models are extensively used and can also be combined with frequency methods. Other extensions to SCR are the use of stratification, presence-absence data only, and, in particular, spatial resight models. Different sightings for marked and unmarked can be allowed for as well as the possible lack of individual recognition.

With technological advances, camera methods are being increasingly used to estimate densities of elusive terrestrial mammals, animals with low densities, and those animals difficult to capture or detect. Along with DNA methods, which are also considered, they have many advantages such as being noninvasive. The design of SCR models is considered. Many examples and applications are given throughout the chapter.

Appendix A

Some General Results

A.1 Poisson Process

We look at one- and two-dimensional versions of this process.

A.1.1 One-Dimensional

A homogeneous Poisson process is a point process where the number of events in a time interval of length T is $\text{Poisson}(\lambda T)$, where λ is the “intensity” of the process. Given there are n events in the time interval, those events form an independent random sample from a uniform distribution on the interval. For the interested reader, the following is a technical definition of an inhomogeneous Poisson process (IPP) where λ is not constant, but given by $\lambda(t)$.

We consider a counting process $\{N(t), t \geq 0\}$ in time with the following properties:

- (1) $N(0) = 0$.
- (2) Has independent increments.
- (3) $\Pr(N(t+h) - N(t) = 1) = \lambda(t)h + o(h)$, where $o(h)/h \rightarrow 0$ as $h \rightarrow 0$.
- (3) $\Pr(N(t+h) - N(t) \geq 2) = o(h)$.

Here $\lambda(t)$ is called the intensity function. The above properties imply that $N(t+h) - N(t)$ is a Poisson random variable with parameter (mean)

$$\mathbb{E}[N(t+h) - N(t)] = \int_t^{t+h} \lambda(s)ds,$$

which implies

$$\mathbb{E}[N(t)] = \int_0^t \lambda(s)ds.$$

We also have for interval $(a, b]$

$$\Pr(N(a, b] = n) = \frac{\Lambda(a, b)}{n!} \exp[-\Lambda(a, b)],$$

where

$$\Lambda(a, b) = \int_a^b \lambda(t)dt.$$

If $\lambda(t) = \lambda$, we have a homogeneous Poisson process.

A.1.2 Two-Dimensional

An IPP defined for a spatial Poisson process has an intensity function $\lambda(x, y)$ with coordinates (x, y) . If A is a bounded area, we define

$$\Lambda(A) = \int_A \lambda(x, y)dxdy.$$

Then the number $N(A)$ in area A is Poisson with mean $\Lambda(A) = \mathbb{E}[N(A)]$. Also, the numbers of points occurring in disjoint regions are independent.

A.2 N-Mixture Models

An example of an N-mixture model is as follows, where there is a latent (unknown) process such as

$$N_i \sim \text{Poisson}(\lambda),$$

where N_i is the abundance at site i , and an observation process that builds on it such as

$$y_{ij} \mid N_i \sim \text{Binomial}(N_i, p),$$

where y_{ij} is the j th count on site i and p is the probability of detection. The model is also referred to as a hierarchical model.

A.3 Multinomial Distribution

A multinomial distribution arises from N independent trials where a trial has k mutually exclusive outcomes. Here, the x_i , the number of counts occurring in the cell with probability p_i ($i = 1, 2, \dots, k$), have a joint probability function

$$\begin{aligned}\Pr[\{x_i\}] &= \frac{N!}{\prod_{i=1}^k x_i!} \prod_{i=1}^k p_i^{x_i} \\ &= \binom{N}{x_1, x_2, \dots, x_k} \prod_{i=1}^k p_i^{x_i}, \quad \sum_{i=1}^k x_i = N, \quad \sum_{i=1}^k p_i = 1.\end{aligned}\quad (\text{A.1})$$

Here $\Pr[\{x_i\}]$ stands for

$$\Pr[\{x_i\}] = \Pr[X_1 = x_1, X_2 = x_2, \dots, X_k = x_k],$$

the “singular” (and symmetrical version) of the distribution, as we need the identifiability constraint $\sum_{i=1}^k p_i = 1$. We can express this distribution briefly as the singular distribution $(x_1, x_2, \dots, x_k) \sim \text{Multinomial}(N, \mathbf{p})$.

We obtain the nonsingular distribution by writing $p_k = 1 - \sum_{i=1}^{k-1} p_i = 1 - P$, say, and $x_k = N - \sum_{i=1}^{k-1} x_i = N - r$, say. In this respect, the notations for the singular and non-singular distributions are sometimes confused in the literature, especially the definition

$$\binom{N}{\{x_i\}} = \frac{N!}{\prod_{i=1}^k x_i!}.$$

The binomial distribution, written $\text{Binomial}(n, p)$, has $k = 2$, and the Bernoulli distribution, written $\text{Bernoulli}(x, p)$, is $\text{Binomial}(1, p)$ for a discrete random x taking the values of 1 and 0.

A.3.1 Some Properties

By adding appropriate cells together, we see that the marginal distribution of any subset of a multinomial distribution is also multinomial with the appropriate p_i 's added together. If we finally end up with just two pooled cells, we have the binomial distribution. We now show how a nonsingular multinomial distribution can be represented by the product of conditional binomial distributions. To do this we

note first that if x_1 and x_2 have a multinomial distribution and x_1 has a binomial distribution, we find that

$$\begin{aligned}\Pr[x_2 \mid x_1] &= \frac{\Pr[x_2, x_1]}{\Pr[x_1]} \\ &= \binom{N - x_1}{x_2} \left(\frac{p_2}{1 - p_1} \right)^{x_2} (1 - \frac{p_2}{1 - p_1})^{n - x_1 - x_2}.\end{aligned}$$

Now

$$\Pr[x_1, x_2, \dots, x_k] = \Pr[x_1] \Pr[x_2 \mid x_1] \Pr[x_3 \mid x_1, x_2] \cdots \Pr[x_k \mid x_1, x_2, \dots, x_{k-1}]$$

and we have shown that for $k = 2$ we have

$$\Pr[x_1, x_2] = \Pr[x_1] \Pr[x_2 \mid x_1]$$

where both distributions are binomial. We can then show by induction that the factorization of $\Pr[x_1, x_2, \dots, x_k]$ gives a product of conditional binomial distributions.

A.3.2 Conditional Multinomial Distribution

Given a nonsingular multinomial distribution, then

$$\Pr[x_1, x_2] = \frac{N}{x_1! x_2! (N - x_1 - x_2)!} p_1^{x_1} p_2^{x_2} (1 - p_1 - p_2)^{N - x_1 - x_2}.$$

If $y = x_1 + x_2$, then y has probability function

$$\Pr[y] = \binom{N}{y} (p_1 + p_2)^y (1 - p_1 - p_2)^{N-y}$$

and

$$\begin{aligned}\Pr[x_1 \mid y] &= \frac{\Pr[x_1, y]}{\Pr[y]} \\ &= \frac{\Pr[x_1, x_2]}{\Pr[y]} \\ &= \binom{y}{x_1} \left(\frac{p_1}{p_1 + p_2} \right)^{x_1} \left(\frac{p_2}{p_1 + p_2} \right)^{x_2},\end{aligned}$$

which is a singular binomial distribution. We now show that this result holds generally for a nonsingular multinomial distribution

$$\Pr[x_1, \dots, x_{k-1}] = \frac{N}{\prod_{i=1}^{k-1} x_i!(N-r)!} \prod_{i=1}^{k-1} p_i^{x_i} (1-P)^{N-r}.$$

Then, r has a binomial distribution

$$\Pr(r) = \binom{N}{r} P^r (1-P)^{N-r},$$

and we find that

$$\Pr[x_1, \dots, x_{k-1} \mid r] = \frac{r}{\prod_{i=1}^{k-1} x_i!} \prod_{i=1}^{k-1} \left(\frac{p_i}{P} \right),$$

a singular distribution that does not depend on N , and

$$\Pr[x_1, \dots, x_{k-1}] = \Pr[x_1, \dots, x_{k-1} \mid r] \Pr[r].$$

A.4 Delta Method

We consider general ideas only without getting too involved with technical details about limits (see also Agresti (2013, section 16.1)). Let X be a random variable with mean μ and variance σ^2 , and let $Y = g(X)$ be a “well-behaved” function of X that has a Taylor expansion

$$g(X) - g(\mu) = (X - \mu)g'(\mu) + \frac{1}{2}(X - \mu)^2 g''(X_0),$$

where X_0 lies between X and μ , $g'(\mu)$ is the derivative of g evaluated at $X = \mu$, and $g''(X_0)$ is the second derivative of g evaluated at $X = X_0$. Then taking expected values,

$$\mathbb{E}[g(X)] \approx g(\mu) + \frac{1}{2}\sigma^2 g''(\mu). \quad (\text{A.2})$$

Assuming second- order terms can be neglected, we have $\mathbb{E}[Y] \approx g(\mu)$ and

$$\begin{aligned} \text{var}(Y) &\approx \mathbb{E}[(g(X) - g(\mu))^2] \\ &\approx \mathbb{E}[(X - \mu)^2][g'(\mu)]^2 \\ &= \sigma^2 [g'(\mu)]^2. \end{aligned} \quad (\text{A.3})$$

For example, if $g(X) = \log X$ then, for large μ ,

$$\text{var}(\log X) \approx \sigma^2/\mu^2. \quad (\text{A.4})$$

If $\mathbf{X} = (X_1, X_2, \dots, X_k)'$ is a vector with mean $\boldsymbol{\mu}$, then for suitable g , we have the first-order Taylor expansion

$$Y = g(\mathbf{X}) - g(\boldsymbol{\mu}) \approx \sum_{i=1}^k (X_i - \mu_i) g'_i(\boldsymbol{\mu}), \quad (\text{A.5})$$

where $g'_i(\boldsymbol{\mu})$ is $\partial g / \partial X_i$ evaluated at $\mathbf{X} = \boldsymbol{\mu}$. Then

$$\begin{aligned} \text{var}[Y] &\approx \text{E}[(g(\mathbf{X}) - g(\boldsymbol{\mu}))^2] \\ &\approx \text{E} \left[\sum_{i=1}^k \sum_{j=1}^k (X_i - \mu_i)(X_j - \mu_j) g'_i(\boldsymbol{\mu}) g'_j(\boldsymbol{\mu}) \right] \\ &= \sum_{i=1}^k \sum_{j=1}^k \text{cov}[X_i, X_j] g'_i(\boldsymbol{\mu}) g'_j(\boldsymbol{\mu}). \end{aligned} \quad (\text{A.6})$$

A “quick and dirty” method for a product or ratio of two variables is as follows. If $Y = X_1/X_2$ then taking logs and differentials, we get

$$\frac{\Delta Y}{Y} = \frac{\Delta X_1}{X_1} - \frac{\Delta X_2}{X_2}.$$

Squaring and taking expected values gives us

$$\text{var} \left[\frac{X_1}{X_2} \right] \approx \mu_y^2 \left(\frac{\text{var}[X_1]}{\mu_1^2} + \frac{\text{var}[X_2]}{\mu_2^2} - 2 \frac{\text{cov}[X_1, X_2]}{\mu_1 \mu_2} \right), \quad (\text{A.7})$$

where $\mu_y \approx \mu_1/\mu_2$. For a product $X_1 X_2$ we simply replace the minus sign by a plus sign and μ_y by $\mu_1 \mu_2$.

Sometimes we wish to derive asymptotic variances and covariances of parameters using Taylor expansions and the delta method. This can be very onerous, but there are a few shortcuts or “rules” that have been brought together by Jolly (1965, section 5).

A.5 Parameter-Expanded Data Augmentation

Royle, Dorazio et al. (2007a) introduced an estimation method for handling a multinomial model with unknown index N , which arises frequently in this book, e.g., Sect. 5.5.2. Bayesian analysis of such a model has proved difficult as the dimension of the parameter space may not be fixed, being in some cases a function of N , for example, if p_h is the probability of capturing individual h , where $h = 1, \dots, N$. The augmentation idea is to augment, say, the n observed capture histories with a fixed known number, say M , of all-zero capture histories (“pseudo” individuals), and to model the augmented dataset as a zero-inflated version of the complete-data model using an unknown, but estimable, zero-inflation parameter. We require $M > N$, which is usually easily achieved a priori. In fact M can often be interpreted as the size of a super-community of species from which the local community is drawn. The key is to fix the size of the data set. The proposed data augmentation yields a data set composed of the observed detection frequencies y_1, \dots, y_n and the augmented zeros y_{n+1}, \dots, y_M .

Given the augmented dataset, we may adopt a conventional likelihood-based framework for inference wherein the likelihood of the augmented data set turns out to be a zero-inflated binomial mixture. Alternatively we can use a Bayesian framework, where, for example, Monte Carlo Markov chain (MCMC) methods such as Gibbs sampling may be used to compute inferences for N and other model parameters. Royle and Dorazio (2012) described a wider application of the above method under the title of “parameter-expanded data augmentation” extending the method beyond just capture-recapture.

A.6 Conditional Expectations

If x and y are any pair of random variables, then

$$\mathbb{E}[x] = \mathbb{E}_y \{\mathbb{E}[x | y]\} \quad (\text{A.8})$$

and

$$\text{var}[x] = \mathbb{E}_y \{\text{var}[x | y]\} + \text{var}_y[\{\mathbb{E}[x | y]\}], \quad (\text{A.9})$$

where \mathbb{E}_y , etc. denotes taking the expected value with respect to the distribution of y . We note that if $\mathbb{E}[x | y]$ does not depend on y (e.g., just a constant), then the second term of (A.9) is zero and, by the delta method above,

$$\begin{aligned} \text{var}[x] &= \mathbb{E}_y[\{\text{var}[x | y]\}] \\ &= \mathbb{E}_y[g(y)], \text{ say,} \end{aligned}$$

$$\begin{aligned} &\approx g(\theta) \\ &= \{\text{var}[x \mid y]\}_{y=\theta}, \end{aligned} \tag{A.10}$$

where $\theta = E[y]$.

If z is independent of x and y , then we have

$$\text{cov}[x, yz] = E_z\{\text{cov}[x, yz \mid z]\} = E[z]\text{cov}[x, y]. \tag{A.11}$$

A.7 Profile Likelihood Intervals

Confidence intervals for a parameter θ are generally based on $\hat{\theta}$, the maximum likelihood estimate of θ . In the past, since $\hat{\theta}$ is asymptotically normal with mean θ and variance estimate $\hat{\sigma}^2(\hat{\theta})$, the 95% confidence interval used has been $\hat{\theta} \pm 1.96\hat{\sigma}$, sometimes referred to as a Wald confidence interval. However, it is well-known that this interval can perform badly for small samples in closed populations, and in other contexts, due to such things as the bias in $\hat{\theta}$, a poor variance estimate, and asymmetry in the sampling distribution of $\hat{\theta}$. Various transformations of θ have been suggested such as the logarithm and, if θ is a probability, the logit transformation $\eta = \text{logit}(\theta) = \log[\theta/(1 - \theta)]$ (though it is usually expressed linearly in terms of other parameters). We then back-transform the interval using $\theta = 1 - (1 + e^\eta)^{-1}$.

An alternative method of constructing confidence intervals that is generally better than the Wald method is to use the so-called profile likelihood, and we first consider having just a single parameter θ . If $L(\theta)$ is the likelihood function of θ , the likelihood ratio test for the null hypothesis $H_0 : \theta = \theta_0$ does not reject H_0 at the α level of significance if, for large samples,

$$-2[\log L(\theta_0) - \log L(\hat{\theta})] < \chi_1^2(1 - \alpha),$$

or

$$\log L(\theta_0) > \log L(\hat{\theta}) - \frac{1}{2}\chi_1^2(1 - \alpha) \quad (= \ell^*),$$

where $\chi_1^2(1 - \alpha)$ is the upper $100(1 - \alpha)$ th quantile of the chi-square distribution with one degree of freedom. The $100(1 - 2\alpha)\%$ profile confidence interval for θ is then the set of all θ_0 not rejected by the test, that is, satisfies the above inequality, namely, $\log L(\theta_0) > \ell^*$. The bisection method is commonly used to find the endpoints of the interval by finding the solutions of $\log L(\theta_0) - \ell^* = 0$.

For a vector parameter $\theta = (\theta_1, \dots, \theta_p)'$, we can find the profile interval for a single element, say $\beta = \theta_i$, as follows. We first find the maximum likelihood

estimate of $\boldsymbol{\theta}$ subject to $\theta_i = \beta$, say $\tilde{\boldsymbol{\theta}}(\beta)$. Then our profile interval for θ_i is the set of all β satisfying

$$\log L(\tilde{\boldsymbol{\theta}}(\beta)) - \ell^* = 0.$$

The bisection method is very time consuming as it requires solutions of the equation $\log L(\tilde{\boldsymbol{\theta}}(\beta)) - \ell^* = 0$ for each element of $\boldsymbol{\theta}$. However, Gimenez et al. (2005) used a better algorithm proposed by Venzon and Moolgavkar (1988) as follows. The endpoints are solutions of the nonlinear equations $\mathbf{F}(\boldsymbol{\theta}) = \mathbf{0}$, where

$$\mathbf{F}(\boldsymbol{\theta}) = \left(\begin{array}{c} \log L(\boldsymbol{\theta}) - \ell^* \\ \frac{\partial \log L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_{-i}} \end{array} \right),$$

where $\boldsymbol{\theta}_{-i}$ is $\boldsymbol{\theta}$ without θ_i . The reader is referred to Gimenez et al. (2005) for further details such as finding the best starting value of $\boldsymbol{\theta}$ for the iterative process, coping with problems like boundary estimates, and allowing for over-dispersion. They found from their simulation results that both the profile and Wald-type intervals achieve the nominal coverage as long the size of the release cohort exceeded a hundred. Furthermore, when it is less than this, there was no gain in using the profile instead of the Wald intervals. An algorithm is available for computing the profile intervals in M-SURGE (Choquet et al., 2004). Clearly this topic needs further investigation.

A.8 Large Sample Hypothesis Tests

For a general theory of large sample tests, see Seber (2015), which shows that the usual large sample tests are asymptotically equivalent to a test of a linear hypothesis. Applications to the multinomial distribution are given by Seber (2013, chapter 4).

In the past there has been two standard general procedures for testing hypotheses in capture-recapture modeling, the Pearson goodness-of-fit test and the likelihood-ratio test that leads to the concept of deviance. More recently the Score test has also been used, which we first describe generally. Given the likelihood function $L(\boldsymbol{\theta})$, the Score test for testing a null hypothesis H_0 about $\boldsymbol{\theta}$ is

$$\left(\frac{\partial \log L(\boldsymbol{\theta})'}{\partial \boldsymbol{\theta}} \mathbf{B} \frac{\partial \log L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_0},$$

where $\hat{\boldsymbol{\theta}}_0$ is the maximum likelihood estimate of $\boldsymbol{\theta}$ when H_0 is true, \mathbf{B} is $-n^{-1}\mathbb{E}[\partial^2 \log L(\boldsymbol{\theta})/\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}']$, the expected information matrix, and n is the sample size. It is not recommended to replace \mathbf{B} by the observed information matrix

$-\partial^2 \log L(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$ (cf. Morgan et al., 2007, for a simple illustration of what can go wrong if this is done).

There is a general principle involved with exponential families, which includes the multinomial. If $\boldsymbol{\theta}$ is the unknown vector parameter and \mathbf{T} denotes a vector of sufficient statistics, then the likelihood function can be factorized so that

$$L(\boldsymbol{\theta}; \text{data}) = \Pr[\text{data} | \mathbf{T}] \times \Pr[\mathbf{T}; \boldsymbol{\theta}],$$

where the second component is used to estimate $\boldsymbol{\theta}$, while the first component is used to test for model adequacy, usually involving hypergeometric distributions and contingency tables.

A.8.1 Diagnostic Residuals for Frequency Models

In regression, residuals of the form observation minus the fitted value, or scaled versions, have been used effectively to provide a check on the assumptions underlying the model. In multinomial models we focus on scaled versions of the raw residuals $x_i - np_i(\hat{\boldsymbol{\theta}})$ or $O_i - E_i$. For example, the Pearson residuals

$$e_i = (x_i - np_i(\hat{\boldsymbol{\theta}})) / \sqrt{np_i(\hat{\boldsymbol{\theta}})}$$

or $(O_i - E_i)/\sqrt{E_i}$ are called the (scaled) Pearson residuals (though terminology is not consistent in the literature) and satisfy $\sum_i e_i^2 = X^2$, the Pearson goodness-of-fit test. If the data come from Poisson distributions, for which the mean equals the variance, then the e_i are approximately distributed as $N(0, 1)$. For multinomial data, the variance is smaller than the mean (compare npq for the binomial distribution with np for a Poisson) so that the e_i will be too small in this case. Instead we can divide each e_i by its estimated standard deviation to get the so-called adjusted (or standardized) residual (Haberman, 1973), namely, $r_i = e_i / \sqrt{\hat{v}_{ii}}$ where \hat{v}_{ii} is an estimate of the variance of e_i . An expression for r_i and a derivation are given by Agresti (2013, section 16.3.2, equation (16.18)), which we summarize as follows.

Assuming H_0 is true, let $\boldsymbol{\theta}_T$ be the true value of $\boldsymbol{\theta}$ and let $\mathbf{p}_T = \mathbf{p}(\boldsymbol{\theta}_T)$. Let

$$\mathbf{A} = (\text{diag } \mathbf{p}(\boldsymbol{\theta}))^{-1/2} \frac{\partial \mathbf{p}}{\partial \boldsymbol{\theta}'}, \quad \text{where } \frac{\partial \mathbf{p}}{\partial \boldsymbol{\theta}'} = \left(\frac{\partial p_r(\boldsymbol{\theta})}{\partial \theta_s} \right),$$

and $\text{diag } \mathbf{p}$ is a diagonal matrix with elements p_i ($i = 1, 2, \dots, k$). We then find that for large n , \mathbf{e} is asymptotically multivariate normal $N_k(\mathbf{0}, \mathbf{V}_T)$, where

$$\mathbf{V}_T = \left[\mathbf{I}_k - \mathbf{p}(\boldsymbol{\theta})^{1/2} (\mathbf{p}(\boldsymbol{\theta})^{1/2})' - \mathbf{A} (\mathbf{A}' \mathbf{A})^{-1} \mathbf{A}' \right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_T}.$$

Estimating θ_T by $\widehat{\theta}$, we get

$$\widehat{v}_{ii} = \left[1 - p_i(\boldsymbol{\theta}) - \sum_{r=1}^k \sum_{s=1}^k \frac{1}{p_i(\boldsymbol{\theta})} \frac{\partial p_i}{\partial \theta_r} \frac{\partial p_i}{\partial \theta_s} v^{rs} \right]_{\widehat{\theta}},$$

where v^{rs} is the (r, s) th element of $(\mathbf{A}'\mathbf{A})^{-1}$.

Another type of residual is the so-called *deviance residual* that is defined by $G^2 = \sum_{i=1}^k \epsilon_i^2$, where

$$\epsilon_i = \text{sign}(x_i - np_i(\widehat{\boldsymbol{\theta}})) \sqrt{|2x_i \log \frac{x_i}{np_i(\widehat{\boldsymbol{\theta}})}|},$$

where $\text{sign}(x_i - np_i(\widehat{\boldsymbol{\theta}}))$ is $+1$ if the difference is positive, 0 if the difference is 0 , and -1 if the difference is negative. If an x_i is zero and all the p_i 's are positive, we ignore the residual for x_i in G^2 .

In capture-recapture studies, we record y_ω the number of individuals with capture history ω . For example, if $s = 4$ and $\omega' = (1010)$, we consider those individuals caught in the first and third samples, but not in the second and fourth. If $E[y_\omega] = \mu_\omega$, then we can obtain the Pearson residual $(y_\omega - \widehat{\mu}_\omega)/\sqrt{\widehat{\mu}_\omega}$ and the deviance residual

$$\text{sign}(y_\omega - \widehat{\mu}_\omega)[2y_\omega \log(y_\omega/\widehat{\mu}_\omega)]^{1/2}.$$

Various box plots can be constructed with these residuals, for example, all the residuals with frequency of capture i . When carrying out an omnibus goodness-of-fit test for a contingency table and obtaining significance, we need to look at cell residuals to see where the departures lie.

A.9 Bayesian Methods

After a slow start in closed population studies, Bayesian methods have become more popular and are used extensively in this book. Some knowledge of prior distributions is assumed, but some background theory is given by Seber and Schofield (2019, Appendices A9–A10); Markov chain Monte Carlo sampling, model averaging, and some further diagnostics are also described in their (Appendix 11), along with bootstrap and Monte Carlo estimates (A.12).

An important part of Bayesian theory is the use of posterior distributions, and sampling them. For example, suppose x is a random variable with probability or probability density function $f(x; \theta)$, usually referred to as the likelihood function, that depends on an unknown parameter θ with prior distribution $f(\theta)$. Here we use

f generically for all distributions. The posterior distribution of θ is then given by

$$\begin{aligned} f(\theta | x) &= \frac{f(\theta, x)}{f(x)} = \frac{f(x | \theta)f(\theta)}{f(x)} \\ &\propto f(x | \theta)f(\theta). \end{aligned}$$

We can use the above, for example, to maximize the posterior distribution with respect to θ as we can omit anything that does not involve θ . We therefore have the general expression

$$\text{posterior} \propto \text{likelihood} \times \text{prior}. \quad (\text{A.12})$$

In practice we shall sometimes write the likelihood as $L(\theta)$ or $L(\theta; x)$, when the emphasis is on θ and we omit any product terms not involving θ . We then write

$$f(\theta | x) \propto L(\theta; x)f(\theta).$$

We note that $f(x)$ is the marginal distribution of x given by

$$f(x) = \int f(x | \theta)f(\theta)d\theta.$$

The percentiles of a posterior distribution can be used to construct an interval for θ called a highest-posterior-density interval. If $\ell_{0.05}$ and $u_{0.95}$ are the lower and upper 5% percentiles of the posterior distribution, then

$$\Pr[\ell_{0.05} < \theta < u_{0.95} | x] = 0.9$$

and θ lies in the interval $[\ell_{0.05}, u_{0.95}]$, usually called a credible interval, with a probability of 0.90, or loosely 90%. We could also use what we might describe as one-sided 90% intervals such as $(0, u_{0.90})$ and $(\ell_{0.10}, 1)$.

A.9.1 Sequential Updating

In some situations we may wish to update our knowledge as data are collected. This can be done using Bayesian updating; the posterior distribution at time t is used as the prior distribution at time $t + 1$. Suppose that we observe sequential data y_1, \dots, y_T , where y_t is observed at time t . The model for y_1, \dots, y_T is

$$f(y_1, \dots, y_T | \theta) = f(y_1 | \theta)f(y_2 | y_1, \theta) \dots f(y_T | y_1, \dots, y_{t-1}, \theta)$$

The standard approach to finding the posterior distribution for θ is to wait until all data are observed at time T and find

$$f(\theta|y_1, \dots, y_T) = \frac{f(y_1, \dots, y_T|\theta)f(\theta)}{f(y_1, \dots, y_T)}.$$

Bayesian updating sequentially updates the posterior distribution as data arrives. The posterior at time t is

$$f(\theta|y_1, \dots, y_t) = \frac{f(y_1, \dots, y_t|\theta)f(\theta)}{f(y_1, \dots, y_t)}. \quad (\text{A.13})$$

At time $t+1$ we use (A.13) as the prior distribution and update our knowledge based on y_{t+1} ,

$$f(\theta|y_1, \dots, y_{t+1}) = \frac{f(y_{t+1}|y_1, \dots, y_t, \theta)f(\theta|y_1, \dots, y_t)}{f(y_{t+1}|y_1, \dots, y_t)}. \quad (\text{A.14})$$

The posterior in (A.14) is equivalent to that found if we had waited to observe all observations y_1, \dots, y_{t+1} and found the posterior based on the full data. To see this, we substitute $f(\theta|y_1, \dots, y_t)$ from (A.13) into (A.14),

$$\begin{aligned} f(\theta|y_1, \dots, y_{t+1}) &= \frac{f(y_{t+1}|y_1, \dots, y_t, \theta)f(y_1, \dots, y_t|\theta)f(\theta)}{f(y_{t+1}|y_1, \dots, y_t)f(y_1, \dots, y_t)} \\ &= \frac{f(y_1, \dots, y_{t+1}|\theta)f(\theta)}{f(y_1, \dots, y_{t+1})}. \end{aligned}$$

A.9.2 Bayes Factors and Posterior Model Probabilities

One approach to model uncertainty is to consider Bayes factors, or alternatively posterior model probabilities. Suppose that there are K models under consideration. A model variable M is included as another unknown. We use M_j as a shorthand for $M = j$, representing that model j is true. The Bayes factor between two models, j and h , is then given by

$$BF_{jh} = \frac{f(y|M_j)}{f(y|M_h)},$$

where y represents the data, and

$$f(y|M_i) = \int f(y|\theta, M_i)f(\theta|M_i)d\theta, \quad (\text{A.15})$$

is the marginal likelihood of the data y given the model M_i , where $f(\theta|M_i)$ is the prior for θ in model i . The Bayes factor BF_{jh} is a ratio of the evidence for model j compared to model h . If we let $f(M_j|y)$ denote the posterior probability of model j and the $f(M_j)$ represent the prior probability, then Bayes factor can also be represented as

$$\frac{f(M_j|y)}{f(M_h|y)} = BF_{jh} \frac{f(M_j)}{f(M_h)}.$$

It is a multiplier that changes the prior odds into posterior odds. The posterior probability of model j can be found in terms of Bayes factors and the prior model probabilities. Taking model h as a reference, $h \in \{1, \dots, K\}$,

$$f(M_j|y) = \frac{BF_{jh} f(M_j)}{\sum_{k=1}^K BF_{kh} f(M_k)},$$

where $BF_{ii} = 1$.

There are several appealing features of Bayes factors and posterior model probabilities, including:

- They follow from Bayes rule and are coherent.
- They allow us to evaluate evidence in favor of a null hypothesis.
- They do not require models to be nested.
- They can be used for Bayesian model averaging. For a parameter common in all K models, we can find a posterior distribution that accounts for uncertainty in the model choice; see Hoeting et al. (1999) for more details.

We also consider two downsides to the use of Bayes factors. The first is that Bayes factors (and equivalently posterior model probabilities) can be sensitive to the choice of priors chosen for the parameters θ , particularly when the prior is diffuse. This is best seen with Jeffreys-Lindley paradox; see, for example, Robert (2014). This has led (Gelman & Shalizi, 2013) to state

We do not see Bayesian methods as generally useful for giving the posterior probability that a model is true, or the probability for preferring model A over model B.

The second downside is that Bayes factors are difficult to evaluate. The marginal likelihood (A.15) used to define the Bayes factor is the same calculation that slowed the progress of Bayesian data analysis until modern computational approaches allowed us to avoid it. While various approximations and algorithms have been proposed that assist in the evaluation of Bayes factors, they remain difficult quantities to compute. Further information about the Bayes factor was given by Kass and Raftery (1995), including rules of thumb on how to interpret them.

References

- Aars, J., Marques, T. A., & et al. (2009). Estimating the Barents Sea polar bear subpopulation size. *Mammal Marine Science*, 25(1), 35–52.
- Acevedo, M., Corrada-Bravo, C., Corrada-Bravo, H., Villnueva-Rviera, L., & Aide, T. (2009). Automated classification of bird and amphibian calls using machine learning: A comparison of methods. *Ecological Informatics*, 4(4), 206–214.
- Agarwal, D. K., Gelfand, A. E., & Citron-Pousty S. (2002). Zero-inflated models with application to spatial count data. *Environmental and Ecological Statistics*, 9(4), 341–355.
- Agresti, A. (1994). Simple capture-recapture models permitting unequal catchability and effort. *Biometrics*, 50(2), 494–500.
- Agresti, A. (2013). *Categorical data analysis* (3rd ed.). Hoboken, NJ: Wiley.
- Aing, C., Halls, S., Oken, K., Dobrow, R., & Fieberg, J. (2011). A Bayesian hierarchical occupancy model for track surveys conducted in a series of linear, spatially correlated, sites. *Journal of Applied Ecology*, 48(6), 1508–1517.
- Ait Kaci Azzou, S., Singer, L., Aebsicher, T., Caduff, M., Wolf, B., & Wegmann, D. (2021). A sparse observation model to quantify species distributions and their overlap in space and time. *Ecography*, 44(6), 928–940.
- Alho, J. M. (1990). Logistic regression in capture-recapture models. *Biometrics*, 46(3) 623–635.
- Alho, J. M., Mulry, M. H., Wurdean, K., & Kim, J. (1993). Estimating heterogeneity in the probabilities of enumeration for dual-system estimation. *Journal of the American Statistical Association*, 88(423), 1130–1136.
- Alisauskas, R. T., & Conn, P. B. (2019). Effects of distance on detectability of Arctic waterfowl using double-observer sampling during helicopter surveys. *Ecology and Evolution*, 9(2), 859–867.
- Alldredge, M. W., Pacifici, K., Simons, T. R., & Pollock, K. H. (2008). A novel field evaluation of the effectiveness of distance and independent observer sampling to estimate aural avian detection probabilities. *Journal of Applied Ecology*, 45(5), 1349–1356.
- Alldredge, M. W., Pollock, K. H., & Simons, T. S. (2006). Estimating detection probabilities from multiple-observer point counts. *Auk*, 123(4), 1172–1182.
- Alldredge, M. W., Pollock, K. H., Simons, T. S., & Shriner, S. A. (2007a). Multiple-species analysis of point count data: A more parsimonious modeling framework. *Journal of Applied Ecology*, 44(2), 281–290.
- Alldredge, M. W., Pollock, K. H., Simons, T. R., Collazo, J. A., & Shriner, S. A. (2007b). Time-of-detection method for estimating abundance from point-count surveys. *Auk*, 124(2), 653–664.

- Alldredge, M. W., Simons, T. R., & Pollock, K. H. (2007a). Factors affecting aural detections of songbirds. *Ecological Applications*, 17(3), 948–955.
- Alldredge, M. W., Simons, T. R., Pollock, K. H., & Pacifici, K. (2007c). A field evaluation of the time-of-detection method to estimate population size and density for aural avian point counts. *Avian Conservation and Ecology*, 2(2), 13.
- Alonso, R. S., McClintock, B. T., Lyren, L. M., Boydston, E. E., & Crooks, K. R. (2015). Mark-recapture and mark-resight methods for estimating abundance with remote cameras: A carnivore case study. *PLoS One*, 10(3), e0123032.
- Alpizar-Jara, R., & Pollock, K. H. (1996). A combination line transect and capture-recapture sampling model for multiple observers in aerial surveys. *Environmental and Ecological Statistics*, 3(4), 311–327.
- Alpizar-Jara, R., & Pollock, K. H. (1999). Combining line transect and capture-recapture for mark-resighting studies. In *Marine mammal survey and assessment methods: Proceedings of the Symposium on Surveys, Status & Trends of Marine Mammal Populations*. Seattle, Washington (pp. 99–114). Rotterdam; Brookfield, VT: Balkema.
- Alpizar-Jara, R., & Smith, C. E. (2008). A continuous time version and a generalization of a Markov-recapture model for trapping experiments. *Mathematical Biosciences*, 214(1), 11–19.
- Altieri, L., Farcomeni, A., & Fegatelli, D. A. (2022). Continuous time-interaction processes for population size estimation, with an application to drug dealing in Italy. *Biometrics*, 79(2), 1254–1267.
- Altwegg, R., Wheeler, M., & Erni, B. (2008). Climate and the range dynamics of species with imperfect detection. *Biology Letters*, 4(5), 581–584.
- Amundson, C. L., Royle, J. A., & Handel, C. M. (2014). A hierarchical model combining distance sampling and time removal to estimate detection probability during avian point counts. *Auk*, 31(4), 476–494.
- Anan, O., Böhning, D., & Maruotti, A. (2007). *Journal of Statistical Computation and Simulation*, 87(10), 2094–2114.
- Anderson, C. W., Nielsen, C. K., Hester, C. M., Hubbard, R. D., Stroud, J. K., & Schauber, E. M. (2013). Comparison of indirect and direct methods of distance sampling for estimating density of white-tailed deer. *Wildlife Society Bulletin*, 37(1), 146–154.
- Anderson, D. R. (2001). The need to get the basics right in wildlife field studies. *Wildlife Society Bulletin*, 29(4), 1294–1297.
- Anderson, D. R., Burnham, K. P., & Crain, B. R. (1979). Line transect estimation of population size: The exponential case with grouped data. *Communications in Statistics—Theory and Methods A*, 8(5), 487–507.
- Anderson, D. R., Burnham, K. P., White, G. C., & Otis, D. L. (1983). Density estimation of small-mammal populations using a trapping web and distance sampling methods. *Ecology*, 64(4), 674–680.
- Anderson, D. R., & Pospahala, R. S. (1970). Correction of bias in belt transect studies of immotile objects. *Journal of Wildlife Management*, 34(1), 141–146.
- Anderson, M. J., & Fienberg, S. E. (1999). *Who counts? The politics of census-taking in contemporary America* (Vol. ADE 4.5). Russell Sage Foundation.
- Andrieu, C., & Roberts, G. O. (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *Annals of Statistics*, 37(2), 697–725.
- Arnold, R., Hayakawa, Y., & Yip, P. (2010). Capture-recapture estimation using finite mixtures of arbitrary dimension. *Biometrics*, 66(2), 644–655.
- Araújo, M. B., & Guisan, A. (2006). Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, 33, 1677–1688.
- Askins, R. A., Philbrick, M. J., & Sugeno, D. S. (1987). Relationship between the regional abundance of forest and the composition of forest bird communities. *Biological Conservation*, 39(2), 129–152.
- Arzoumanian, Z., Holmberg, J., & Norman, B. (2005). An astronomical pattern-matching algorithm for computer-aided identification of whale sharks *Rhincodon typus*. *Journal of Applied Ecology*, 42(6), 999–1011.

- Augustine, B. C., Royle, J. A., & et al. (2018). Spatial capture-recapture with partial identity: An application to camera traps. *Annals of Applied Statistics*, 12(1), 67–95.
- Augustine, B. C., Royle, J. A., Linden, D. W., & Fuller, A. K. (2020). Spatial proximity moderates genotype uncertainty in genetic tagging studies. *Proceedings of the National Academy of Sciences*, 117, 17903–17912.
- Augustine, B. C., Royle, J. A., Murphy, S. M., Chandler, R. B., Cox, J. J., & Kelly, M. J. (2019). Spatial capture-recapture for categorically marked populations with an application to genetic capture-recapture. *Ecosphere*, 10(4), e02627.
- Ausband, D. E., Rich, N. R., & et al. (2014). Monitoring gray wolf populations using multiple survey methods. *Journal of Wildlife Management*, 78(2), 335–346.
- Azaele, S., Muneepeerakul, R., Rinaldo, A., & Rodriguez-Iturbe, I. (2010). Inferring plant ecosystem organization from species occurrences. *Journal of Theoretical Biology*, 262(2), 323–329.
- Azuma, D. L., Baldwin, J. A., & Noon, B. R. (1990). Estimating the occupancy of spotted owl habitat areas by sampling and adjusting for bias. In *USDA General Technical Report PSW-124*. Berkeley, CA.
- Bailey, L. L., Hines, J. E., Nichols, J. D., & MacKenzie, D. I. (2007). Sampling design trade-offs in occupancy studies with imperfect detection: Examples and software. *Ecological Applications*, 17(1), 281–290.
- Bailey, L. L., Reid, J. A., Forsman, E. D., & Nichols, J. D. (2009). Modeling co-occurrence of northern spotted and barred owls: Accounting for detection probability differences. *Biological Conservation*, 142(12), 2983–2989.
- Bailey, L. L., MacKenzie, D. I., & Nichols, J. D. (2014). Advances and applications of occupancy models. *Methods in Ecology and Evolution*, 5(12), 1269–1279.
- Baillargeon, S., & Rives, L.-P. (2007). Recapture: Loglinear models for capture-recapture in R. *Journal of Statistical Software*, 19(5), 1–31.
- Baker, L. F., Artym, K. J., & Swanson, H. K. (2017). Optimal sampling methods for modelling the occupancy of Arctic grayling (*Thymallus arcticus*) in the Canadian Barrenlands. *Canadian Journal of Fisheries and Aquatic Sciences*, 74(10), 1564–1574.
- Balderama, E., Gardner, B., & Reich, B. J. (2016). A spatial-temporal double-hurdle model for extremely over-dispersed avian count data. *Spatial Statistics*, 18(part A), 263–275.
- Banks-Leite, C., Pardini, R., & et al. (2014). Assessing the utility of statistical adjustments for imperfect detection in tropical conservation science. *Journal of Applied Ecology*, 51(4), 849–859.
- Barabesi, L., & Marcheselli, M. (2008). Improved strategies for coverage estimation by using replicated line-intercept sampling. *Environmental and Ecological Statistics*, 15(2), 215–239.
- Barbet-Massin, M., Jiguet, F., Albert, C. H., & Thuiller, W. (2012). Selecting pseudo-absences for species distribution models: How, where and how many. *Methods in Ecology and Evolution*, 3(2), 327–328.
- Bardos, D. C., Guillera-Arroita, G., & Wintle, B. A. (2015). Valid auto-models for spatially autocorrelated occupancy and abundance data. *Methods in Ecology and Evolution*, 6(10), 1137–1149.
- Barker, R. J., Forsyth, D. M., & Wood, M. (2014). Modeling sighting heterogeneity and abundance in spatially replicated multiple-observer surveys. *Journal of Wildlife Management*, 78(4), 701–708.
- Barker, R. J., & Link, W. A. (2013). Bayesian multimodel inference by RJMCMC: A Gibbs sampling approach. *American Statistician*, 67(3), 150–156.
- Barker, R. J., Schofield, M. R., Link, W. A., & Sauer, J. R. (2018). On the reliability of N-mixture models for count data. *Biometrics*, 74(1), 369–377.
- Barker, R. J., Schofield, M. R., Wright, J. A., Frantz, A. C., & Stevens, S. (2014). Closed-population capture–recapture modeling of samples drawn one at a time. *Biometrics*, 70(4), 775–782.

- Barlow, J., & Taylor, B. L. (2005). Estimates of sperm whale abundance in the northeastern temperate Pacific from a combined acoustic and visual survey. *Marine Mammal Science*, 21(3), 429–445.
- Bart, J., Droege, S., Geissler, P., Peterjohn, B., & Ralph, C. J. (2004). Density estimation in wildlife surveys. *Wildlife Society Bulletin*, 32(4), 1242–1247.
- Bart, J., & Earnst, S. (2002). Double sampling to estimate density and population trends in birds. *Auk*, 119(1), 36–45.
- Bartolucci, F., & Lupparelli, M. (2008). Focused information criterion for capture-recapture models for closed population. *Scandinavian Journal of Statistics*, 35(4), 629–649.
- Bartolucci, F., Mira, A., & Scaccia, L. (2004). Answering two biological questions with a latent class model via mcmc applied to capture-recapture data. In M. Di Bacco, G. D'Amore, & F. Scalpari (Eds.), *Applied Bayesian statistical studies in biology and medicine* (pp. 7–23). Boston, MA: Springer US.
- Basiri, R., Moradi, M., Kiani, B., & Maasumi Babaarabi, M. (2018). Evaluation of distance methods for estimating population density in *Populus euphratica* Olivier natural stands (case study: Maroon riparian forests Iran). *Journal of Forest Science*, 64(5), 230–244.
- Basu, S. (1998). Bayesian estimation of the number of undetected errors when both reviewers and errors are heterogeneous. In A. P. Basu & S. Mukhopadhyay (Eds.), *Frontiers in reliability analysis* (pp. 19–36). Hackensack, New Jersey: World Scientific Publishing.
- Basu, S., & Ebrahimi, N. (2001). Bayesian capture-recapture methods for error detection and estimation of population size: Heterogeneity and dependence. *Biometrika*, 88(1), 269–279.
- Baumgartner, L. J. (2004). Population estimation methods to quantify temporal variation in fish accumulations downstream of a weir. *Fisheries Management and Ecology*, 13(6), 355–364.
- Becker, E. F., & Christ, A. M. (2015). A unimodal model for double observer distance sampling surveys. *PLoS One*, 10(8), e0136403.
- Becker, E. F., Golden, H. N., & Gardner, C. L. (2004). Using probability sampling of animal tracks in snow to estimate population size. In W. L. Thompson (Ed.), *Sampling rare or elusive species: Concepts, designs, and techniques for estimating population parameters* (pp. 248–270). Washington, D.C.: Island Press.
- Becker, E. F., & Quang, P. X. (2009). A gamma-shaped detection function for line-transect surveys with mark-recapture and covariate data. *Journal of Agricultural, Biological, and Environmental Statistics*, 14(2), 207–223.
- Becker, N. G. (1984). Estimating population size from capture-recapture in continuous time. *Australian Journal of Statistics*, 26(1), 1–7.
- Bedrick, E. J. (1994). Maximum-likelihood estimation for the removal method. *Canadian Journal of Statistics*, 22(2), 285–293.
- Beissinger, S. R., Iknayan, K. J., & et al. (2016). Incorporating imperfect detection into joint models of communities: A response to Warton et al. *Trends in Ecology and Evolution*, 31(10), 736–737.
- Beja-Pereira, A., Oliveira, R., Alves, P. C., Schwartz, M. K., & Luikart, G. (2009). Advancing ecological understandings through technological transformations in noninvasive genetics. *Molecular Ecology Resources*, 9(5), 1279–1301.
- Belant, J. L., Bled, F., Wilton, C. M., Fyumagwa, R., Mwampeta, S. B., & Beyer, Jr., D. E. (2016). Estimating lion abundance using n-mixture models for social species. *Scientific Reports*, 6(1), 35920–35920.
- Bell, M. C., Eaton, D. R., Bannister, R. C. A., & Addison, J. T. (2003). A mark-recapture approach to estimating population density from continuous trapping data: Application to edible crabs, *Cancer pagurus*, on the east coast of England. *Fisheries Research*, 65(1), 361–378.
- Benejam, L., Alcaraz, C., & et al. (2011). Fish catchability and comparison of four electrofishing crews in Mediterranean streams. *Fisheries Research*, 123–124, 9–15.
- Berger, J. (2010). *Statistical decision theory and Bayesian analysis*. New York: Springer-Verlag.
- Berger, J. O., Bernardo, J. M., & Sun, D. (2012). Objective priors for discrete parameter spaces. *Journal of the American Statistical Association*, 107(498), 636–648.

- Besag, J. E., & Gleaves, J. T. (1973). On the detection of spatial pattern in plant communities. *Bulletin of the International Statistical Institute*, 45, 153–158.
- Besag, J., York, J., & Mollie, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43(1), 1–49.
- Binns, M. R. (1986). Behavioural dynamics and the negative binomial distribution. *Oikos*, 47(3), 315–318.
- Bishop, J. (2006). Standardizing fishery-dependent catch and effort data in complex fisheries with technology change. *Reviews in Fish Biology and Fisheries*, 16(1), 21–38.
- Bivand, R. S., Pebesma, E., & Gómez-Rubio, V. (2013). *Applied spatial data analysis* (2nd ed.). New York: Springer.
- Blanc, L., Marboutin, E., Gatti, S., Zimmermann, F., & Gimenez, O. (2014). Improving abundance estimation by combining capture-recapture and occupancy data: Example with a large carnivore. *Journal of Applied Ecology*, 51(6), 1733–1739.
- Blaschke, W. (1949). *Vorlesungen über integralgeometrie*. New York: Chelsea.
- Bled, F., Nichols, J. D., & Altwegg, R. (2013). Dynamic occupancy models for analyzing species' range dynamics across large geographic scales. *Ecology and Evolution*, 3(15), 4896–4909.
- Bled, F., Royle, J. A., & Cam, E. (2011). Hierarchical modeling of an invasive spread: The Eurasian Collared-Dove *Streptopelia decaocto* in the United States. *Ecological Applications*, 21(1), 290–302.
- Bled, F., Sauer, J., Pardieck, K., Doherty, P., & Royle, J. A. (2013). Modeling trends from North American breeding bird survey data: A spatially explicit approach. *PLoS One*, 8(12), e81867.
- Boender, C. G. E., & Rinnooy Kan, A. H. G. (1987). A multinomial Bayesian approach to the estimation of population and vocabulary size. *Biometrika*, 74(4), 849–856.
- Bohk, C., Rau, R., & Cohen, J. E. (2015). Taylor's power law in human mortality. *Demographic Research*, 33(1), 589–610.
- Böhning, D., van Der Heijden, P. G. M., & Bunge, J. (2018). *Capture-recapture methods for the social sciences*. Chapman and Hall.
- Bohrmann, T. F., & Christman, M. C. (2012). Evaluating sampling efficiency in depletion surveys using hierarchical Bayes. *Canadian Journal of Fisheries and Aquatic Sciences*, 69(6), 1080–1090.
- Bohrmann, T. F., & Christman, M. C. (2013). Optimal allocation of sampling effort in depletion surveys source. *Journal of Agricultural, Biological, and Environmental Statistics*, 18(2), 218–233.
- Bolfarine, H., Leite, J. G., & Rodrigues, J. (1992). On the estimation of the size of a finite and closed population. *Biometrical Journal*, 34(5), 577–593.
- Bollinger, E. K., Gavin, T. A., & McIntyre, D. C. (1988). Comparison of transects and circular-plots for estimating bobolink densities. *Journal of Wildlife Management*, 52(4), 777–786.
- Bonner, S. J., & Holmberg, J. (2013). Mark-recapture with multiple non-invasive marks. *Biometrics*, 69(3), 766–775.
- Bonner, S. J., & Schofield, M. R. (2014). MC(MC)MC: Exploring Monte Carlo integration within MCMC for mark-recapture models with individual covariates. *Methods in Ecology and Evolution*, 5(12), 1305–1315.
- Bonner, S. J., Schofield, M. R., Noren, P., & Price, S. J. (2016). Extending the latent multinomial model with complex error processes and dynamic Markov bases. *Annals of Applied Statistics*, 10(1), 246–263.
- Bonnet, A., Martinez Herrera, M., & Sangnier, M. (2021). *Statistics & Probability Letters*, 179, 109214, Article 109214.
- Borchers, D. L. (2012). A non-technical overview of spatially explicit capture-recapture models. *Journal of Ornithology*, 152(Suppl 2), 1–10.
- Borchers, D., Buckland, S., Goedhart, P., Clarke, E., & Hedley, S. (1998). Horvitz-Thompson estimators for double-platform line transect surveys. *Biometrics*, 54(4), 1221–1237.
- Borchers, D. L., Buckland, S. T., & Zucchini, W. (2002). *Estimating animal abundance: Closed populations*. London: Springer-Verlag.

- Borchers, D. L., & Efford, M. G. (2008). Spatially explicit maximum likelihood methods for capture-recapture studies. *Biometrics*, 64(2), 377–385.
- Borchers, D., Distiller, G., Foster, R., Harmsen, B., & Milazzo, L. (2014). Continuous-time spatially explicit capture–recapture models, with an application to a jaguar camera-trap. *Methods in Ecology and Evolution*, 5(7), 656–665.
- Borchers, D., & Fewster, R. (2016). Spatial capture–recapture models. *Statistical Science*, 31(2), 219–232.
- Borchers, D. L., Laake, J. L., Southwell, C., & Paxton, C. G. M. (2006). Accommodating unmodeled heterogeneity in double-observer distance sampling surveys. *Biometrics*, 62(2), 372–378.
- Borchers, D. L., & Langrock, R. (2015). Double-observer line transect surveys with Markov-modulated Poisson process models for animal availability. *Biometrics*, 71(4), 1060–1069.
- Borchers, D. L., & Marques, T. A. (2017). From distance sampling to spatial capture–recapture. *Advances in Statistical Analysis*, 101(4), 475–494.
- Borchers, D., Marques, T., Gunnlaugsson, T., & Jupp, P. (2010). Estimating distance sampling detection functions when distances are measured with errors. *Journal of Agricultural, Biological, and Environmental Statistics*, 15(3), 346–361.
- Borchers, D. L., Nightingale, P., Stevenson, B.C., & Fewster, R. M. (2022). A latent capture history model for digital aerial surveys. *Biometrics*, 78(1), 274–285.
- Borchers, D. L., Stevenson, B. C., Kidney, D., Thomas, L., & Marques, T. A. (2015). A unifying model for capture–recapture and distance sampling surveys of wildlife populations. *Journal of the American Statistical Association*, 110(509), 195–204.
- Borchers, D. L., Zucchini, W., & Fewster, R. M. (1998). Mark–recapture models for line transect surveys. *Biometrics*, 54(4), 1207–1220.
- Bord, S., Bioche, C., & Druilhet, P. (2018). A cautionary note on Bayesian estimation of population size by removal sampling with diffuse priors. *Biometrical Journal*, 60(3), 450–462.
- Bord, S., Druilhet, P., Gasqui, P., Abrial, D., & Vourc'h, G. (2014). Bayesian estimation of abundance based on removal sampling under weak assumption of closed population with catchability depending on environmental conditions. Application to tick abundance. *Ecological modelling*, 274, 72–79.
- Bornand, C. N., K'ery, M., Bueche, L., & Fischer, M. (2014). Hide-and-seek in vegetation: Time-to-detection is an efficient design for estimating detectability and occurrence. *Methods in Ecology and Evolution*, 5(5), 413–442.
- Boulanger, J., Kendall, K. C., Stetz, J. B., Roon, D. A., Waits, L.P., & Paetkau, D. (2008). Multiple data sources improve DNA-based mark–recapture population estimates of grizzly bears. *Ecological Applications*, 18(3), 577–589.
- Boulanger, J., Nielsen, S. E., & Stenhouse, G. B. (2018). Using spatial mark–recapture for conservation monitoring of grizzly bear populations in Alberta. *Scientific Reports*, 8(1), 5204–5215.
- Boulinier, T., Nichols, J. D., Sauer, J. R., Hines, J. E., & Pollock, K. H. (1998). Estimating species richness: The importance of heterogeneity in species detectability. *Ecology*, 79(3), 1018–1028.
- Braaten, D. O. (1969). Robustness of the De Lury population estimator. *Journal of the Fisheries Research Board of Canada*, 26(2), 339–355.
- Bradley, D., Conklin, E., & et al. (2017). Resetting predator baselines in coral reef ecosystems. *Science Report*, 7(1), 43131.
- Brandes, T. S. (2008). Automated sound recording and analysis techniques for bird surveys and conservation. *Bird Conservation International*, 18(S1), S163–S173.
- Bravington, M.V., Miller, D. L.,& Hedley, S. L.(2021). Variance propagation for density surface models. *Journal of Agricultural, Biological, and Environmental Statistics*, 26(2), 306–323.
- Brierley, A. S., Axelsen, B. E., & et al. (2004). Single-target echo detections of jellyfish. *ICES Journal of Marine Science*, 61(3), 383–393.
- Brintz, B., Fuentes, C., & Madsen, L. (2018). An asymptotic approximation to the N-mixture model for the estimation of disease prevalence. *Biometrics*, 74(4), 1512–1518.

- Britzke, E. R., Gillam, E. H., & Murray, K. L. (2013.) Current state of understanding of ultrasonic detectors for the study of bat ecology. *Acta Theriologica*, 58(2), 109–117.
- Broms, K. M., Johnson, D. S., Altwegg, R., & Conquest, L. L. (2014). Spatial occupancy models applied to atlas data show Southern Ground Hornbills strongly depend on protected area. *Ecological Applications* 2(2), 363–374.
- Broms, K. M., Hooten, M. B., & Fitzpatrick, R. M. (2016). Model selection and assessment for multi-species occupancy models. *Ecology*, 97(7), 1759–1770.
- Broms, K. M., Hooten, M. B., Johnson, D. S., Altwegg, R., & Conquest, L. L. (2016a). Dynamic occupancy models for explicit colonization processes. *Ecology*, 97(1), 194–204.
- Brooks, S. P., Morgan, B. J. T., Ridout, M. S., & Pack, S. E. (1997). Finite mixture models for proportions. *Biometrics*, 53(3), 1097–1115.
- Brotons, L., Thuiller, W., Araújo, M. B., & Hirzel, A. H. (2004). Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. *Ecography*, 27(4), 437–448.
- Buck, R. C. (1965). *Advanced calculus*. New York: Wiley.
- Buckland, S. T. (1987). On the variable circular plot method of estimating animal density. *Biometrics*, 43(2), 363–384.
- Buckland, S. T. (2006). Point transect surveys for songbirds: Robust methodologies. *Auk*, 123(2), 345–357.
- Buckland, S. T., Anderson, D. R., Burnham, K. P., & Laake, J. L. (1993). *Distance sampling: Estimating abundance of biological populations*. London: Chapman and Hall.
- Buckland, S. T., Anderson, D. R., Burnham, K. P., Laake, J. L., Borchers, D. L., & Thomas, L. (2001). *Introduction to distance sampling*. Oxford, U. K.: Oxford University Press.
- Buckland, S. T., Anderson, D. R., Burnham, K. P., Laake, J. L., Borchers, D. L., & Thomas, L. (Eds.). (2004). *Advanced distance sampling*. Oxford, UK: Oxford University Press. Paperback, 2007
- Buckland, S. T., Magurran, A. E., Green, R. E., & Fewster, R. M. (2005). Monitoring changes in biodiversity through composite indices. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1454), 243–254.
- Buckland, S. T., Borchers, D. L., Johnston, A., Henrys, P. A., & Marques, T. A. (2007). Line transect methods for plant surveys. *Biometrics*, 63(4), 989–998.
- Buckland, S. T., & Johnston, A. (2017). Monitoring the biodiversity of regions: Key principles and possible pitfalls. *Biological Conservation*, 214, 23–34.
- Buckland, S. T., Laake, J. L., & Borchers, D. L. (2010). Double-observer line transect methods: Levels of independence. *Biometrics*, 66(1), 169–177.
- Buckland, S. T., Marsden, S. J., & Green, R. E. (2008). Estimating bird abundance: Making methods work. *Bird Conservation International*, 18(S1), S91–S108.
- Buckland, S. T., Oedekoven, C. S., & Borchers, D. L. (2016). Model-based distance sampling. *Journal of Agricultural, Biological, and Environmental Statistics*, 21(1), 58–75.
- Buckland, S. T., Rexstad, E. A., Marques, T. A., & Oedekoven, C. S. (2015). *Distance sampling: Methods and applications*. Springer International Publishing: Switzerland.
- Buckland, S. T., Russell, R. E., Dickson, B. G., Saab, V. A., Gorman, D. G., & Block, W. M. (2009). Analysing designed experiments in distance sampling. *Journal of Agricultural, Biological and Environmental Statistics*, 14(4), 432–442.
- Buckland, S. T., Summers, R. W., Borchers, D. L., & Thomas, L. (2006). Point transect sampling with traps or lures. *Journal of Applied Ecology*, 43(2), 377–384.
- Buckland, S. T., Yuan, Y., & Marcon, E. (2017). Measuring temporal trends in biodiversity. *Advanced Statistical Analysis*, 101(4), 461–474.
- Burch, B. D., & Sánchez Meador, A. J. (2018). Comparison of forest age estimators using k-tree, fixed-radius, and variable-radius plot sampling. *Canadian Journal of Forest Research*, 48(8), 942–951.
- Burgar, J. M., Stewart, F. E. C., Volpe, J. P., Fisher, J. T., & Burton, A. C. (2018). Estimating density for species conservation: Comparing camera trap spatial count models to genetic spatial capture-recapture models. *Global Ecology and Conservation*, 15, e0411.

- Burgman, M. A., & Fox, J. C. (2003). Bias in species range estimates from minimum convex polygons: Implications for conservation and options for improved planning. *Animal Conservation*, 6(1), 19–28.
- Burnham, K. P., & Anderson, D. R. (1976). Mathematical models for nonparametric inferences from line transect data. *Biometrics*, 32(2), 325–326.
- Burnham, K. P., Anderson, D. R., & Laake, J. L. (1980). Estimation of density for line transect sampling of biological populations. *Wildlife Monographs*, 72, 3–202.
- Burnham, K. P., Buckland, S. T., & et al. (2004). Design of distance sampling surveys and geographic information systems. In S. T. Buckland, D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers, & L. Thomas (Eds.), *Advanced distance sampling*. Oxford, UK: Oxford University Press.
- Burnham, K. P., Buckland, S. T., Laake, & et al. (2004). Further topics in distance sampling. In S. T. Buckland, D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers, & L. Thomas (Eds.), *Advanced distance sampling*, chap 11 (pp. 307–392). Oxford, UK: Oxford University Press.
- Burnham, K. P., & Overton, W. S. (1978). Estimation of the size of a closed population when capture probabilities vary among animals. *Biometrika*, 65(3), 625–633.
- Burnham, K. P., & Overton, W. S. (1979). Robust estimation of population size when capture probabilities vary among animals. *Ecology*, 60(5), 927–936.
- Burt, M. L., Borchers, D. L., Jenkins, K. J., & Marques, T. A. (2014). Using mark–recapture distance sampling methods on line transect surveys. *Methods in Ecology and Evolution*, 5(11), 1180–1191.
- Burton, A. C., Neilson, E., & et al. (2015). Wildlife camera trapping: A review and recommendations for linking surveys to ecological processes. *Journal of Applied Ecology*, 52(3), 675–685.
- Butler, M., Ballard, W., Wallace, M., Demaso, S., & Mcgee, B. (2007). Aerial surveys for estimating wild turkey abundance in the Texas Rolling Plains. *Journal of Wildlife Management*, 71(5) 1639–1645.
- Butler, S., & Bird, F. L. (2007). Estimating density of intertidal ghost shrimps using counts of burrow openings. Is the method reliable? *Hydrobiologia*, 589(1), 303–314.
- Buxton, A., Matechou, E., Griffin, J., Diana, A., & Griffiths, R. A. (2021). Optimising sampling and analysis protocols in environmental DNA studies. *Scientific Reports*, 11(1), No. 11637.
- Caldwell, Z. R., Zgliczynski, B. J., Williams, G. J., & Sandin, S. A. (2016). Reef fish survey techniques: Assessing the potential for standardizing methodologies. *PLoS One*, 11(4), e0153066.
- Callaghan, C. T., Bowler, D. E., Blowes, S. A., Chase, J. M., Lyons, M. B., & Pereira, H. M. (2022). Quantifying effort needed to estimate species diversity from citizen science data. *Ecosphere*, 13(4), e3966.
- Callaghan, C. T., Ozeroff, I., Hitchcock, C., & Chandler, M. (2020). Capitalizing on opportunistic citizen science data to monitor urban biodiversity: A multi-taxa framework. *Biological Conservation*, 251, article108753.
- Camacho, C. (2016). Birding trip reports as a data source for monitoring rare species. *Animal Conservation*, 19(5), 430–435.
- Caravaggi, A., Banks, P. B., & et al. (2017). A review of camera trapping for conservation behaviour research. *Remote Sensing in Ecology and Conservation*, 3(3), 109–122.
- Carle, F. L., & Strub, M. R. (1978). A new method for estimating population size from removal data. *Biometrics*, 34(4), 621–630.
- Carlin, B., & Chib, S. (1995). Bayesian model choice via Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society, Series B*, 57(3), 473–484.
- Carpenter, B., Gelman, A., & et al. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1).
- Carrier, P., Rosenfeld, J. S., & Johnson, R. M. (2009). Dual-gear approach for calibrating electric fishing capture efficiency and abundance estimates. *Fisheries Management and Ecology*, 16(2), 139–146.
- Carroll, R. J., & Lombard, F. (1985). A note on N estimators for the binomial distribution. *Journal of the American Statistical Association*, 80(390), 423–426.

- Casner, K., Forister, M., Karthik, R., & Shapiro, A. (2014). The utility of repeated presence data as a surrogate for counts: A case study using butterflies. *Journal of Insect Conservation*, 18(1), 13–27.
- Castledine, B. J. (1981). A Bayesian analysis of multiple-recapture sampling for a closed population. *Biometrika*, 68(1), 197–210.
- Castellón, T. D., Rothermel, B. B., & Nomaní, S. Z. (2015). A comparison of line-transect distance sampling methods for estimating gopher tortoise population densities. *Wildlife Society Bulletin*, 39(4), 804–812.
- Chakraborty, A., Gelfand, A. E., Wilson, A. M., Latimer, A. M., & Silander, J. A. (2011). Point pattern modelling for degraded presence-only data over large regions. *Applied Statistics*, 60(5), 757–776.
- Chambert, T., Kendall, W. L., & et al. (2015). Testing hypotheses on distribution shifts and changes in phenology of imperfectly detectable species. *Methods in Ecology and Evolution*, 6(6), 638–647.
- Chambert, T., Miller, D. A. W., & Nichols, J. D. (2015). Modeling false positive detections in species occurrence data under different study designs. *Ecology*, 96(2), 332–333.
- Chambert, T., Waddle, J. H., Miller, D. A. W., Walls, S. C., Nichols, J. D., & Yoccoz, N. (2018). A new framework for analysing automated acoustic species detection data: Occupancy estimation and optimization of recordings post-processing. *Methods in Ecology and Evolution*, 9(3), 560–570.
- Chandler, R. B., & Hepinstall-Cyberman, J. (2016). Estimating the spatial scales of landscape effects on abundance. *Landscape Ecology*, 31(6), 1383–1394.
- Chandler, R. B., Muths, E., Sigafus, B. H., Schwalbe, C. R., Jarchow, C. J., & Hossack, B. R. (2015). Spatial occupancy models for predicting metapopulation dynamics and viability following reintroduction. *Journal of Applied Ecology*, 52(5), 1325–1333.
- Chandler, R. B., & Royle, J. A. (2013). Spatially explicit models for inference about density in unmarked or partially marked populations. *Annals of Applied Statistics*, 7(2), 936–954.
- Chandler, R. B., Royle, J. A., & King, D. I. (2011). Inference about density and temporary emigration in unmarked populations. *Ecology*, 92(7), 1429–1435. Errata: *Ecology*, 95(3), 794.
- Chao, A. (1984). Nonparametric estimation of the number of classes in a population. *Scandinavian Journal of Statistics*, 11(4), 265–270.
- Chao, A. (1987). Estimating the population size for capture-recapture data with unequal catchability. *Biometrics*, 43(4), 783–791.
- Chao, A. (1989). Estimating population size for sparse data in capture-recapture experiments. *Biometrics*, 45(2), 427–438.
- Chao, A. (2001). An overview of closed capture-recapture models. *Journal of Agricultural, Biological and Environmental Statistics*, 6(2), 158–175.
- Chao, A. (2005). Species estimation and applications. In N. Balakrishnan, C. B. Read, & B. Vidakovic (Eds.), *Encyclopedia of statistical sciences* (pp. 7907–7916). New York: Wiley.
- Chao, A. (2015). Capture-recapture for human populations. In *Wiley Stats Ref: Statistics Reference Online* (pp. 1–16). <https://doi.org/10.1002/9781118445112.stat04855.pub2>.
- Chao, A., & Chang, S.-H. (1999). An estimating function approach to the inference of catch-effort models. *Environmental and Ecological Statistics*, 6(3), 313–334.
- Chao, A., & Chiu, C. H. (2012). Estimation of species richness and shared species richness. In N. Balakrishnan (Ed.), *Methods and applications of statistics in the atmospheric and earth sciences* (pp. 76–111). New York: NY, USA.
- Chao, A., & Chiu, C.-H. (2016a). Species richness: Estimation and comparison. *Wiley StatsRef: Statistics Reference Online*, 1, 26.
- Chao, A., & Chiu, C.-H. (2016b). Nonparametric estimation and comparison of species richness. *Wiley StatsRef: Statistics Reference Online*, 1, 11.
- Chao, A., Chiu, C.-H., Hsieh, T. C., Davis, T., Nipperess, D. A., & Faith, D. P. (2015a). Rarefaction and extrapolation of phylogenetic diversity. *Methods in Ecology and Evolution*, 6(4), 380–388.

- Chao, A., Chiu, C.-H., & Jost, L. (2014). Unifying species diversity, phylogenetic diversity, functional diversity, and related similarity and differentiation measures through Hill numbers. *Annual Review of Ecology, Evolution, and Systematics*, 45(1), 297–324.
- Chao, A., Chiu, C.-H., Wu, S.-H., Huang, C.-L., & Lin, Y. C. (2019). Comparing two classes of alpha diversities and their corresponding beta and (dis)similarity measures, with an application to the Formosan sika deer *Cervus nippon taiouanus* reintroduction programme. *Methods in Ecology and Evolution*, 10(8), 1286–1297.
- Chao, A., Chu, W., & Hsu, C.-H. (2000). Capture-recapture when time and behavioral response affect capture probabilities. *Biometrics*, 56(2), 427–433.
- Chao, A., & Colwell, R. K. (2017). Thirty years of progeny from Chao's inequality: Estimating and comparing richness with incidence data and incomplete sampling (invited article). *Statistics and Operation Research Transactions*, 41(1), 3–54.
- Chao, A., Colwell, R. K., Chiu, C.-H., Townsend, D., & Murrell, D. (2017). Seen once or more than once: Applying Good-Turing theory to estimate species richness using only unique observations and a species list. *Methods in Ecology and Evolution*, 8(10), 1221–1232.
- Chao, A., Colwell, R. K., Lin, C.-W., & Gotelli, N. J. (2009). Sufficient sampling for asymptotic minimum species richness estimators. *Ecology*, 90(4), 1125–1133.
- Chao, A., Gotelli, N. J., & et al. (2014). Rarefaction and extrapolation with Hill numbers: A framework for sampling and estimation in species diversity studies. *Ecological Monographs*, 84(1), 45–67.
- Chao, A., Henderson, P. A., & et al. (2021). Measuring temporal change in alpha diversity. A framework integrating taxonomic, phylogenetic and function diversity and the next iNEXT.3D standardization. *Methods in Ecology and Evolution*, 12(10), 1926–1940.
- Chao, A., Hsieh, T. C., Chazdon, R. L., Colwell, R. K., & Gotelli, N. J. (2015b). Unveiling the species-rank abundance distribution by generalizing the Good-Turing sample coverage theory. *Ecology*, 96(5), 1189–1201.
- Chao, A., Hwang, W.-H., Chen, Y.-C., & Kuo, C.-Y. (2000). Estimating the number of shared species in two communities. *Statistica Sinica*, 10(1), 227–246.
- Chao, A., & Jost, L. (2012). Coverage-based rarefaction and extrapolation: Standardizing samples by completeness rather than size. *Ecology*, 93(12), 2533–2547.
- Chao, A., Kubota, Y., & et al. (2020). Quantifying sample completeness and comparing diversities among assemblages. *Ecological Research*, 35(2), 292–314.
- Chao, A., Lee, S. M., & Jeng, S. L. (1992). Estimating population size for capture-recapture data when capture probabilities vary by time and individual animal. *Biometrics*, 48(1), 201–216.
- Chao, A., & Lin, C.-W. (2012). Nonparametric lower bounds for species richness and shared species richness under sampling without replacement. *Biometrics*, 68(3), 912–921.
- Chao, A., Ma, K. H., Hsieh, T. C., & Chiu, C.-H. (2015c). User's guide for online program SpadeR (Species-richness Prediction and Diversity Estimation in R).
- Chao, A., Pan, H. Y., & Chiang, S. C. (2008). The Petersen-Lincoln estimator and its extension to estimate the size of a shared population. *Biometrical Journal*, 50(6), 957–970.
- Chao, A., & Shen, T. J. (2010). *User's guide for program SPADE (Species Prediction And Diversity Estimation)*. Program and User's Guide published at http://chao.stat.nthu.edu.tw/wordpress/software_download/.
- Chao, A., & Tsay, P. (1998). A sample coverage approach to multiple-system estimation with application to census undercounts. *Journal of the American Statistical Association*, 93(441), 283–293.
- Chao, A., Tsay, P. K., Lin, S. H., Shau, W. Y., & Chao, D. Y. (2001a). The applications of capture-recapture models to epidemiological data. *Statistics in Medicine*, 20(20), 3123–3157.
- Chao, A., Wang, Y. T., & Jost, L. (2013). Entropy and the species accumulation curve: A novel estimator of entropy via discovery rates of new species. *Methods in Ecology and Evolution*, 4(11), 1091–1110.
- Chao, A., Yip, P., Lee, S. M., & Chu, W. (2001b). Population size estimation based on estimating functions for closed capture recapture models. *Journal of Statistical Planning Inference*, 92(1), 213–232.

- Chapman, D. G. (1951). Some properties of the hypergeometric distribution with applications to zoological sample censuses. *University of California Publications in Statistics*, 1, 131–160.
- Chapman, D. G. (1952). Inverse multiple and sequential samples. *Biometrics*, 8(4), 286–306.
- Chapman, D. G. (1954). The estimation of biological populations. *Annals of Mathematical Statistics*, 25(1), 1–15.
- Chapman, D. G. (1955). Population estimation based on change of composition caused by selective removal. *Biometrika*, 42(3–4), 279–290.
- Chapman, D. G. (1961). Statistical problems in the dynamics of exploited fish population. In *Proceedings of the 4th Berkeley Symposium, 1960* (Vol. 4, pp. 153–168).
- Chapman, D. G. & Murphy, G. I. (1965). Estimates of population and mortality from survey-removal records. *Biometrics*, 21, 921–935.
- Chase, S. K., & Levine, A. (2016). A framework for evaluating and designing citizen science programs for natural resources monitoring. *Conservation Biology*, 30(3), 456–466.
- Chatfield, C. (1995). Model uncertainty, data mining and statistical inference. *Journal of the Royal Statistical Society, Series A*, 158(3), 419–466.
- Chatterjee, K., & Bhuyan, P. (2017). A new capture-recapture model in dual-record system. Technical report no. ASU/2017/8. <https://www.isical.ac.in/~asu/TR/TechRepASU201708.pdf>.
- Chatterjee, K., & Bhuyan, P. (2020a). On the estimation of population size from a post-stratified two-sample capture-recapture data under dependence. *Journal of Statistical Computation and Simulation*, 90(5), 819–838.
- Chatterjee, K., & Bhuyan, P. (2020b). On the estimation of population size from a dependent triple-record system. *Journal of the Royal Statistical Society, Series A*, 182(4), 1487–1501.
- Chatterjee, K., & Mukherjee, D. (2016a). An improved integrated likelihood population size estimation in dual-record system. *Statistics and Probability Letters*, 110, 146–154.
- Chatterjee, K., & Mukherjee D. (2016b). On the estimation of homogeneous population size from a complex dual-record system. *Journal of Statistical Computation and Simulation*, 86(17), 3562–3581.
- Chatterjee, K., & Mukherjee, D. (2018). A new integrated likelihood for estimating population size in dependent dual-record system. *Canadian Journal of Statistics*, 46(4), 577–592.
- Chatterjee, K., & Mukherjee, D. (2021). On the estimation of population size under dependent dual-record system: An adjusted profile-likelihood approach. *Journal of Statistical Computation and Simulation*, 91(13), 2740–2763.
- Chaudhary, V., Wisely, S. M., Hernandez, F. A., Hines, J. E., Nichols, J. D., & Oli, M. K. (2020). A multi-state occupancy modelling framework for robust estimation of disease prevalence in multi-tissue disease systems. *Journal of Applied Ecology*, 57(12), 2463–2474.
- Chee, Y. E., & Wintle, B. A. (2010). Linking modelling, monitoring and management: An integrated approach to controlling overabundant wildlife. *Journal of Applied Ecology*, 47(6), 1169–1178.
- Chelgren, N. D., Adams, M. J., Bailey, L. L., & Bury, R. B. (2011). Using multilevel spatial models to understand salamander site occupancy patterns after wildfire. *Ecology*, 92(2), 408–421.
- Chen, C.-L., Pollock, K. H., & Hoenig, J. M. (1998). Combining change-in-ratio, index-removal, and removal models for estimating population size. *Biometrics*, 54(3), 815–827.
- Chen, G., Kéry, M., Plattner, M., Ma, K., & Gardner, B. (2013). Imperfect detection is the rule rather than the exception in plant distribution studies. *Journal of Ecology*, 101(1), 183–191.
- Chen, S. X. (1999). Estimation in independent observer line transect surveys for cluster populations. *Biometrics*, 55(3), 754–759.
- Chen, S. X., & Lloyd, C. J. (2000). A nonparametric approach to the analysis of two-stage mark-recapture experiments. *Biometrika*, 87(3), 633–649.
- Cheng, E., Hodges, K. E., Sollmann, R., & Mills, L. S. (2017). Genetic sampling for estimating density of common species. *Ecology and Evolution*, 7(16), 6210–6219.
- Chiarucci, A., Bacaro, G., Rocchini, D., & Fattorini, L. (2008). Discovering and rediscovering the sample-based rarefaction formula in the ecological literature. *Community Ecology*, 9(1), 121–123.

- Chipperfield, J., Brown, J., & Bell, P. (2017). Estimating the count error in the Australian census. *Journal of Official Statistics*, 33(1), 43–59.
- Chiu, C.-H., Wang, Y. T., Walther, B. A., & Chao, A. (2014). An improved nonparametric lower bound of species richness via a modified Good-Turing frequency formula. *Biometrics*, 70(3), 671–682.
- Choquet, R., Reboulet, A.-M., and 3 others (2004). M-Surge: New software specifically designed for multistate capture-recapture models. *Animal Biodiversity and Conservation*, 27(1), 207–207.
- Cimprich, D. A. (2009). Effect of count duration on abundance estimates of Black-capped Vireos. *Journal of field ornithology*, 80(1), 94–100.
- Clare, J., McKinney, S. T., Depue, J. E., & Loftin, C. S. (2017). Pairing field methods to improve inference in wildlife surveys while accommodating detection covariance. *Ecological Applications*, 27(7), 2031–2047.
- Clark, A. E., & Altweig, R. (2019). Efficient Bayesian analysis of occupancy models with logit link functions. *Ecology and Evolution*, 9(2), 756–768.
- Clark, A. E., Altweig, R., & Ormerod, J. T. (2016). A variational Bayes approach to the analysis of occupancy models. *PloS One*, 11(2), e0148966.
- Clark, J. S., Gelfand, A. E., Woodall, C. W., & Zhu, K. (2014). More than the sum of the parts: Forest climate response from joint species distribution models. *Ecological Applications*, 24(5), 990–999.
- Clark, J. S., Nemergut, D., Seyednasrollah, B., Turner, P. J., & Zhang, S. (2017). Generalized joint attribute modeling for biodiversity analysis: Median-zero, multivariate, multifarious data. *Ecological Monographs*, 87(1), 34–56.
- Clark, N. J., Wells, K., & Lindberg, O. (2018). Unravelling changing interspecific interactions across environmental gradients using Markov random fields. *Ecology*, 99(6), 1277–1283.
- Clark, P. J., & Evans, F. C. (1954). Distance to nearest neighbor as a measure of spatial relationships in populations. *Ecology*, 35(4), 445–453.
- Claytor, R., & Allard, J. (2003). Change-in-ratio estimates of lobster exploitation rate using sampling concurrent with fishing. *Canadian Journal of Fisheries and Aquatic Sciences*, 60(10), 1190–1203.
- Cleasby, I. R., & Nakagawa, S. (2011). Neglected biological patterns in the residuals: A behavioural ecologist's guide to co-operating with heteroscedasticity. *Behavioral Ecology and Sociobiology*, 65(12), 2361–2372.
- Clement, M. J. (2016). Designing occupancy studies when false-positive detections occur. *Methods in Ecology and Evolution*, 7(12), 1538–1547.
- Clement, M. J., Rodhouse, T. J., Ormsbee, P. C., Szewczak, J. M., & Nichols, J. D. (2014). Accounting for false-positive acoustic detections of bats using occupancy models. *Journal of Applied Ecology*, 51(5), 1460–1467.
- Clobert, J., Baguette, M., Benton, T. G., & Bullock, J. M. (2012). *Dispersal ecology and evolution*. Oxford: Oxford University Press.
- Cogbill, C. V., Thurman, A. L., Williams, J. W., Zhu, J., Mladenoff, D. J., & Goring, S. J. (2018). A retrospective on the accuracy and precision of plotless forest density estimators in ecological studies. *Ecosphere*, 9(4), e02187.
- Cohn, J. (2008). Citizen science: Can volunteers do real research? *BioScience*, 58(3), 192–197.
- Coleman, B. D. (1981). On random placement and species-area relations. *Mathematical Biosciences*, 54(3), 191–215.
- Coleman, B. D., Mares M. A., Willig, M. R., & Hsieh, Y.-H. (1982). Randomness, area, and species richness. *Ecology*, 63(4), 1121–1133.
- Colwell, R. K., Chao, A., & et al. (2012). Models and estimators linking individual-based and sample-based rarefaction, extrapolation and comparison of assemblages. *Journal of Plant Ecology*, 5(1), 3–21.
- Colwell, R. K., & Elsensohn, J. E. (2014). EstimateS turns 20: Statistical estimation of species richness and shared species from samples, with non-parametric extrapolation. *Ecography*, 37(6), 609–613.

- Comer, S., Speldewinde, P., & et al. (2018). Evaluating the efficacy of a landscape scale feral cat control program using camera traps and occupancy models. *Scientific Reports*, 8(1), 5335–5339.
- Comeña, N., & Guerra, Á. (2019). Application of the Gómez-muñoz model to estimate catch and effort in the artisanal fishery for the squid *Loligo vulgaris* (Mollusca: Cephalopoda) in the Ría de Vigo (NW Spain). *Scientia Marina*, 83(4), 305–316.
- Conn, P. B. (2012). hierarchicalDS: Functions for performing hierarchical analysis of distance sampling data. R package version 2.0.1.
- Conn, P. B., & Alisauskas, R. T. (2018). Simultaneous modelling of movement, measurement error, and observer dependence in mark-recapture distance sampling: An application to Arctic bird surveys. *Annals of Applied Statistics*, 12(1), 96–122.
- Conn, P. B., Cooch, E. G., & Caley, P. (2012). Accounting for detection probability when estimating force-of-infection from animal encounter data. *Journal of Ornithology*, 152(Sup2), 511–520.
- Conn, P. B., Hoef, J. M. V., & et al. (2014). Estimating multispecies abundance using automated detection systems: Ice-associated seals in the Bering Sea. *Methods in Ecology and Evolution*, 5(12), 1280–1293.
- Conn, P. B., Johnson, D. S., Williams, P. J. Melin, S. R., & Hooten, M. B. (2018). A guide to Bayesian model checking for ecologists. *Ecological Monographs*, 88(4), 526–542.
- Conn, P., Kendall, W., & Samuel, M. (2004). A general model for the analysis of mark-resight, mark-recapture, and band-recovery data under tag loss. *Biometrics*, 60(4), 900–909.
- Conn, P. B., Laake, J. L., & Johnson, D. S. (2012). A hierarchical modeling framework for multiple observer transect surveys. *PLoS One*, 7(8), e42294.
- Cooch, E. G., Conn, P. C., Ellner, S. P., Dobson, A. P., & Pollock, K. H. (2010). Disease dynamics in wild populations: Modeling and estimation: A review. *Journal of Ornithology*, 152(Sup2), 485–509.
- Cook, R. D., & Jacobson, J. O. (1979). A design for estimating visibility bias in aerial surveys. *Biometrics*, 35, 735–742.
- Cormack, R. M. (1966). A test for equal catchability. *Biometrics*, 22(2), 330–342.
- Cormack, R. M. (1981). Loglinear models for capture-recapture experiments on open populations. In R. W. Hiorns, & D. Cooke (Eds.), *The mathematical theory of the dynamics of biological populations* (pp. 197–215). London: Academic Press.
- Cormack, R. M. (1985). Examples of the use of GLIM to analyse capture-recapture studies. In B. J. T. Morgan, & P. M. North (Eds.), *Statistics in Ornithology* (pp. 243–273). New York: Springer-Verlag.
- Cormack, R. M. (1989). Log-Linear models for capture-recapture. *Biometrics*, 45(2), 395–413.
- Cormack, R. M. (1992). Interval estimation for mark-recapture studies of closed populations. *Biometrics*, 48(2), 567–576.
- Cormack, R. M. (1993). The flexibility of GLIM analyses of multiple recapture or resighting data. In J.-D. Lebreton, & P. North (Eds.), *Marked Individuals in the study of bird population* (pp. 39–49). Basel, Switzerland: Birkhauser Verlag.
- Cormack, R. M. (1994). Unification of mark-recapture analyses by loglinear modelling. In D. J. Fletcher & B. F. J. Manly (Eds.), *Statistics in Ecology and Environmental Monitoring* (pp. 19–32). Dunedin, New Zealand: University of Otago Press.
- Cormack, R. M., & Jupp, P. E. (1991). Inference for Poisson and multinomial models for capture-recapture experiments. *Biometrika*, 78(4), 911–916.
- Cottam, G., Curtis, B. W., & Hale, B. (1953). Some sampling characteristics of a population of randomly dispersed individuals. *Ecology*, 34(4), 741–757.
- Coull, B. A., & Agresti, A. (1999). The use of mixed logit models to reflect heterogeneity in capture-recapture studies. *Biometrics*, 55(1), 294–301.
- Cox, D. (1975). Partial likelihood. *Biometrika*, 62(2), 269–276.
- Cox, D. R., & Hinkley, D. V. (1974). *Theoretical statistics*. Boca Raton, FL: Chapman and Hall/CRC.

- Cox, K. D., Black, M. J., & et al. (2017). Community assessment techniques and the implications for rarefaction and extrapolation with Hill numbers. *Ecology and Evolution*, 7(24), 1213–1226.
- Cragg, J. G. (1971). Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica*, 39(5), 829–844.
- Craig, C.C. (1953). On the utilization of marked specimens in estimating populations of flying insects. *Biometrika*, 40(1–2), 170–176.
- Cressie, N. A. C. (1993). *Statistics for spatial data*. New York: Wiley.
- Cressie, N., Calder, C. A., Clark, J. S., Hoefer, J. M. V., & Wikle, C. K. (2009). Accounting for uncertainty in ecological analysis: The strengths and limitations of hierarchical statistical modeling. *Ecological Applications*, 19(3), 553–570.
- Crum, N. J., Fuller, A. K., Sutherland, C. S., Cooch, E. G., & Hurst, J. (2017). Estimating occupancy probability of moose using hunter survey data. *Journal of Wildlife Management*, 81(3), 521–534.
- Crum, N. J., Neyman, L. C., & Gowan, T. A. (2021). Abundance estimation for line transect sampling: A comparison of distance sampling and spatial capture-recapture models. *PLoS One*, 16(5), e0252231.
- Dacey, M. F. (1963). Order neighbor statistics for a class of random patterns in multidimensional space. *Annals of the Association of American Geographers*, 53(4), 505–515.
- Dacey, M. F. (1964). Modified poisson probability law for point pattern more regular than random. *Annals of the Association of American Geographers*, 54(4), 559–565.
- Dacey, M. F. (1965). Order distance in an inhomogeneous random point pattern. *Canadian Geographer*, 9(3), 144–153.
- Dacey, M. F. (1966). A compound probability law for a pattern more dispersed than random and with areal inhomogeneity. *Economic Geography*, 42(2), 172–179.
- Dail, D., & Madsen, L. (2011). Models for estimating abundance from repeated counts of an open metapopulation. *Biometrics*, 67(2), 577–587.
- D'Amien, M., Rahbek, C., Zimmermann, N. E., & Guisan, A. (2017). Spatial predictions at the community level: From current approaches to future frameworks. *Biological Reviews*, 92(1), 169–187.
- Darling, J. A., & Blum, M. J. (2007). DNA-based methods for monitoring invasive species: A review and prospectus. *Biological Invasions*, 9(7), 751–765.
- Darras, K., Furnas, B., Fitriawan, I., Mulyani, Y., & Tscharntke, T. (2018). Estimating bird detection distances in sound recordings for standardizing detection ranges and distance. *Methods in Ecology and Evolution*, 9(9), 1928–1938.
- Darroch, J. N. (1958). The multiple recapture census. I: Estimation of a closed population. *Biometrika*, 45(3–4), 343–359.
- Darroch, J. N., Fienberg, S. E., Glonek G. F. V., & Junker B. W. (1993). A three-sample multiple-recapture approach to census population estimation with heterogeneous catchability. *Journal of the American Statistical Association*, 88(423), 1137–1148.
- Dauwalter, D. C., & Fisher, W. L. (2007). Electrofishing capture probability of smallmouth bass in streams. *North American Journal of Fisheries Management*, 27(1), 162–171.
- Davis, A. J., Hooten, M. B., & et al. (2016). Inferring invasive species abundance using removal data from management actions. *Ecological Applications*, 26(7), 2339–2346.
- Davis, D. E. (1963). Estimating the numbers of game populations. In H. S. Mosby (Ed.), *Wildlife investigational techniques* (2nd ed., pp. 89–118). Washington: The Wildlife Society.
- Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application*. Cambridge, UK: Cambridge University Press.
- Dawe, E. G., Hoenig, J. M., & Xu, X. (1993). Change-in-ratio and index-removal methods for population assessment and their application to snow crab (*Chionoecetes opilio*). *Canadian Journal of Fisheries and Aquatic Sciences*, 50(7), 1467–1476.
- Dawid, A. P., & Lauritzen, S. L. (1993). Hyper Markov laws in the statistical analysis of decomposable graphical models. *Annals of Statistics*, 21(3), 1272–1317.
- Dawson, D. K., & Efford, M. G. (2009). Bird population density estimated from acoustic signals. *Journal of Applied Ecology*, 46(6), 1201–1209.

- De Lima, L. F., Pelissari, A. L., Rodrigues, C. L., Sousa, N. J., & Corte, A. P. D. (2020). Quality assessment of pine wood harvesting by residue inventory using line intercept cluster sampling. *International Journal of Forest Engineering*, 31(3), 205–210.
- De Lury, D. B. (1947). On the estimation of biological populations. *Biometrics*, 3(4), 145–167.
- De Lury, D. B. (1951). On the planning of experiments for the estimation of fish populations. *Journal of the Fisheries Research Board of Canada*, 8(4), 281–307.
- De Rivera, O. R., McCrea, R., & Margalida, A. (2021). Removal modelling in ecology: A systematic review. *PloS One*, 16(3), e0229965–e0229965.
- Dejean, T., Valentín, A., Mique, C., Taberlet, P., Bellemain, E., & Miaud, C. (2012). Improved detection of an alien invasive species through environmental DNA barcoding: The example of the American bullfrog *Lithobates catesbeianus*. *Journal of Applied Ecology*, 49(4), 953–959.
- Dellaportas, P., & Forster, J. J. (1999). Markov chain Monte Carlo model determination for hierarchical and graphical log-linear models. *Biometrika*, 86(3), 615–633.
- Dénes, F. V., Silveira, L. F., & Beissinger, S. R. (2015). Estimating abundance of unmarked animal populations: Accounting for imperfect detection and other sources of zero inflation. *Methods in Ecology and Evolution*, 6(5), 543–556.
- Dennis, E. B., Brereton, T. M., & et al. (2019). Trends and indicators for quantifying moth abundance and occupancy in Scotland. *Journal of Insect Conservation*, 23(2), 369–380.
- Dennis, E. B., Freeman, S. M., Brereton, T., & Roy, D. B. (2013). Indexing butterfly abundance whilst accounting for missing counts and variability in seasonal pattern. *Methods in Ecology and Evolution*, 4(7), 637–645.
- Dennis, E. B., Morgan, B. J. T., Brereton, T. M., Roy, D. B., & Fox, R. (2017). Using citizen science butterfly counts to predict species population trends. *Conservation Biology*, 31(6), 1350–1361.
- Dennis, E. B., Morgan, B. J. T., Freeman, S. N., Brereton, T. M., & Roy, D. B. (2016a). A generalized abundance index for seasonal invertebrates. *Biometrics*, 72(4), 1305–1314.
- Dennis, E. B., Morgan, B. J. T., Freeman, S. N., Roy, D. B., & Brereton, T. M. (2016b). Dynamic models for longitudinal butterfly data. *Journal of Agricultural, Biological, and Environmental Statistics*, 21(1), 1–21.
- Dennis, E. B., Morgan, B. J. T., & Ridout, M. S. (2015). Computational aspects of N-mixture models. *Biometrics*, 71(1), 237–246.
- Denwood, M. J. (2016). runjags: An R package providing interface utilities, model templates, parallel computing methods and additional distributions for MCMC models in JAGS. *Journal of Statistical Software*, 71(1), 1–25.
- Desmet, P., & Cowling, R. (2004). Using the species-area relationship to set baseline targets for conservation. *Ecology and Society*, 9(2), article 11.
- de Valpine, P., Turek, D., Paciorek, C. J., Anderson-Bergman, C., Lang, D. T., & Bodik, R. (2017). Programming with models: Writing statistical algorithms for general model structures with NIMBLE. *Journal of Computational and Graphical Statistics*, 26(2), 403–413.
- Devarajan, K., Morelli, T. L., & Tenan, S. (2020). Multi-species occupancy models: Review, roadmap, and recommendations. *Ecography*, 43(11), 1612–1624.
- Dey, S., Delampady, M., & Gopalaswamy, A. M. (2019). Bayesian model selection for spatial capture–recapture models. *Ecology and Evolution*, 9(20), 11569–11583.
- Dey, S., Delampady, M., Parameshwaran, R., Kumar, N. S., Srivaths, A., & Karanth, K. U. (2017). Bayesian methods for estimating animal abundance at large spatial scales using data from multiple sources. *Journal of Agricultural, Biological, and Environmental Statistics*, 22(2), 111–139.
- Dice, L. R. (1938). Some census methods for mammals. *Journal of Wildlife Management*, 2(3), 119–130.
- Dickinson, J. L., Shirk, J., & et al. (2012). The current state of citizen science as a tool for ecological research and public engagement. *Frontiers in Ecology and the Environment*, 10(6), 291–297.
- Diefenbach, D. R., Marshall, M. R., Mattice, J. A., Brauning, D. W., & Johnson, D. (2007). Incorporating availability for detection in estimates of bird abundance. *Auk*, 124(1), 96–106.
- Diggle, P. J. (1975). Robust density estimation using distance methods. *Biometrika*, 62(1), 39–48.

- Diggle, P. J. (1977). A note on robust density estimation for spatial point patterns. *Biometrika*, 64(1), 91–95.
- Dillon, A., & Kelly, M. J. (2007). Ocelot (*Leopardus pardalis*) in Belize: The impact of trap spacing and distance moved on density estimates. *Oryx*, 41(4), 469–477.
- Distiller, G., & Borchers, D. L. (2015). A spatially explicit capture–recapture estimator for single-catch traps. *Ecology and Evolution*, 5(21), 5075–5087.
- Dodd, Jr., C. K., & Dorazio, R. M. (2004). Using counts to simultaneously estimate abundance and detection probabilities in a salamander community. *Herpetologica*, 60(4), 468–478.
- Doi, H., Fukaya, K., Oka, S.-A., Sato, K., Kondoh, M., & Miya, M. (2019). Evaluation of detection probabilities at the water-filtering and initial PCR steps in environmental DNA metabarcoding using a multispecies site occupancy model. *Scientific Reports*, 9(1), 3581–358.
- Dorazio, R. M. (2007). On the choice of statistical models for estimating occurrence and extinction from animal surveys. *Ecology*, 88(11), 2773–2782.
- Dorazio, R. M. (2012). Predicting the geographic distribution of a species from presence-only data subject to detection errors. *Biometrics*, 68(4), 1303–1312.
- Dorazio, R. M. (2013). Bayes and Empirical Bayes estimators of abundance and density from spatial capture-recapture data. *PLoS One*, 8(12), e84017.
- Dorazio, R. M. (2014). Accounting for imperfect detection and survey bias in statistical analysis of presence-only data. *Global Ecology and Biogeography*, 23(12), 1472–1484.
- Dorazio, R. M., & Connor, E. F. (2014). Estimating abundances of interacting species using morphological traits, foraging guilds, and habitat. *PLoS One*, 9(4), e94323.
- Dorazio, R. M., Connor, E. F., & Askins, R. A. (2015). Estimating the effects of habitat and biological interactions in an avian community. *PLoS One*, 10(8), e0135987.
- Dorazio, R. M., Gotelli, N. J., & Ellison, A. M. (2011). Modern methods of estimating biodiversity from presence-absence surveys. In G. Venora, O. Grillo, & J. Lopez-Pujol (Eds.), *Biodiversity loss in a changing planet* (pp. 277–302). Rijeka, Croatia: InTech.
- Dorazio, R. M., Jelks, H. L., & Jordan, F. (2005). Improving removal-based estimates of abundance by sampling a population of spatially distinct subpopulations. *Biometrics*, 61(4), 1093–1101.
- Dorazio, R. M., & Karanth, K. U. (2017). A hierarchical model for estimating the spatial distribution and abundance of animals detected by continuous-time recorders. *PLoS One*, 12(5), e0176966.
- Dorazio, R. M., Kéry, M., Royle, J. A., & Plattner, M. (2010). Models for inference in dynamic metacommunity systems. *Ecology*, 91(8), 2466–2475.
- Dorazio, R. M., & Royle, J. A. (2003). Mixture models for estimating the size of a closed population when capture rates vary among individuals. *Biometrics*, 59(2), 351–364.
- Dorazio, R. M., & Royle, J. A. (2005a). Estimating size and composition of biological communities by modeling the occurrence of species. *Journal of the American Statistical Association*, 100(470), 389–398.
- Dorazio, R. M., & Royle, J. A. (2005b). Rejoinder to ‘The performance of mixture models in heterogeneous closed population capture–recapture.’ *Biometrics*, 61(3), 874–876.
- Dorazio, R. M., Royle, J. A., Söderström, B., & Glimskär, A. (2006). Estimating species richness and accumulation by modeling species occurrence and detectability. *Ecology*, 87(4), 842–854.
- Dormann, C. F., Bobrowski, M., & et al. (2018). Biotic interactions in species distribution modelling: 10 questions to guide interpretation and avoid false conclusions. *Global Ecology and Biogeography*, 27(9), 1004–1016.
- Dormann, C. F., Calabrese, J. M., & et al. (2018). Model averaging in ecology: A review of Bayesian, information-theoretic, and tactical approaches for predictive inference. *Ecological Monographs*, 88(4), 485–504.
- Dormann, C. F., McPherson, J. M., & et al. (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: A review. *Ecography*, 30(5), 609–628.
- Downs, J. A., Horner, M. W., & Tucker, A. D. (2011). Time-geographic density estimation for home range analysis. *Annals of Geographic Information Sciences*, 17(3), 163–171.

- Du Preez, B. D., Loveridge, A. J., & Macdonald, D. W. (2014). To bait or not to bait: A comparison of camera-trapping methods for estimating leopard *Panthera pardus* density. *Biological Conservation*, 176, 153–116.
- Dunham, K. M., & du Toit, A. J. (2013). Using citizen-based survey data to determine densities of large mammals: A case study from Mana Pools National Park, Zimbabwe. *African Journal of Ecology*, 51(3), 431–440.
- Duan, X., Liu, L., & Zhao, P. (2009). Estimation of covariate distribution with capture-recapture data. *Communications in Statistics, Theory and Methods*, 38(20), 3705–3712.
- Dunn, P. K., & Smyth, G. K. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, 5(3), 236–244.
- Duquette, J. F., Belant, J. L., Svoboda, N. J., Beyer, Jr., D. E., & Albright, C. A. (2014). Comparison of occupancy modeling and radiotelemetry to estimate ungulate population dynamics. *Population Ecology*, 56(3), 481–492.
- Durban, J. W., & Elston, D. A. (2005). Mark-recapture with occasion and individual effects: Abundance estimation through Bayesian model selection in a fixed dimensional parameter spaces. *Journal of Agricultural, Biological, and Environmental Statistics*, 10(3), 291–305.
- Dusetzina, S. B., Tyree, S., Meyer, A. M., Meyer, A., Green, L., & Carpenter, W. R. (2014). *Linking data for health services research: A framework and instructional guide* (Internet) Rockville (MD): Agency for Healthcare Research and Quality (US); 2014 Sep. <https://www.ncbi.nlm.nih.gov/books/NBK253313>.
- Eberhardt, L. L. (1967) Some developments in ‘distance sampling’. *Biometrics*, 23(2), 207–216.
- Eberhardt, L. L. (1968). A preliminary appraisal of line transects. *Journal of Wildlife Management*, 32(1), 82–88.
- Eberhardt, L. L. (1978). Transect methods for population studies. *Journal of Wildlife Management*, 42(1), 1–31.
- Eberhardt, L. L. (1982). Calibrating an index using removal data. *Journal of Wildlife Management*, 46(3), 734–740.
- Eberhardt, L. L., & Simmons, M. A. (1987). Calibrating population indices by double sampling. *Journal of Wildlife Management*, 51(3), 665–675.
- Edgar, G. J., Barrett, N. S., & Morton, A. J. (2004). Biases associated with the use of underwater visual census techniques to quantify the density and size-structure of fish populations. *Journal of Experimental Marine Biology and Ecology*, 308(2), 269–290.
- Efford, M. G. (2004). Density estimation in live-trapping studies. *Oikos*, 106(3), 598–610.
- Efford, M. G. (2011). Estimation of population density by spatially explicit capture–recapture analysis of data from area searches. *Ecology*, 92(12), 2202–2207.
- Efford, M. G. (2013). SECR: Spatially explicit capture-recapture models. R package version 2.6.1. <http://CRAN.R-project.org/package=secr>.
- Efford, M. G. (2019). Non-circular home ranges and the estimation of population density. *Ecology*, 100(2), e02580.
- Efford, M. G., Borchers, D. L., & Byrom, A. E. (2009). Density estimation by spatially explicit capture-recapture: Likelihood-based methods. In Thomson, D. L., Cooch, E. G., & Conroy, M. J. (Eds.), *Modeling demographic processes in marked populations*. Environmental and Ecological Statistics (Vol. 3). Boston, MA: Springer.
- Efford, M. G., Borchers, D. L., Mowat, G., & Ginenez, O (2013). Varying effort in capture-recapture studies. *Methods in Ecology and Evolution*, 4(7), 629–636.
- Efford, M., & Dawson, D. (2009). Effect of distance-related heterogeneity on population size estimates from point counts. *Auk*, 126(1), 100–111.
- Efford, M. G., Dawson, D. K., & Borchers, D. L. (2009). Population density estimated from locations of individuals on a passive detector array. *Ecology*, 90(10), 2676–2682.
- Efford, M. G., & Fewster, R. M. (2013). Estimating population size by spatially explicit capture-recapture. *Oikos*, 122(6), 918–928.
- Efford, M. G., & Hunter, C. M. (2018). Spatial capture-mark-resight estimation of animal population density. *Biometrics*, 74(2), 411–420.

- Efford, M. G., & Mowat, G. (2014). Compensatory heterogeneity in spatially explicit capture-recapture data. *Ecology*, 95(5), 1341–1348.
- Efford, M., & Schofield, M. R. (2020). A spatial open-population capture- recapture model. *Biometrics*, 76(2), 392–402.
- Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.
- Elith, J., Graham, C. H., & et al. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 29(2), 129–151.
- Elith, J., & Leathwick, J. R. (2009) Species distribution models: Ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, 40(1), 677–697.
- Elith, J., Phillips, S. J., Hastie, T., Dudik, M., Chee, Y. E., & Yates, C. J. (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, 17(1), 43–57.
- Ellison, A. M. (2010). Partitioning diversity. *Ecology*, 91(7), 1962–1963.
- Emlen, J. T. (1971). Population densities of birds derived from transect counts. *Auk*, 88(2), 323–342.
- Engler, R., Guisan, A., & Rechsteiner, L. (2004). An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. *Journal of Applied Ecology*, 41(2), 263–274.
- Etterson, M. A., Niemi, G. J., & Danz, N. P. (2009). Estimating the effects of detection heterogeneity and overdispersion on trends estimated from avian point counts. *Ecological Applications*, 19(8), 2049–2066.
- Etterson, M., Schumaker, N., Garber, K., Lennartz, S., Kanarek, A., & Connolly, J. (2021). A spatially explicit model for estimating risks of pesticide exposure to bird populations. *PLoS One*, 16(6), e0252545.
- Evans, M. A., & Bonett, D. G. (1994). Bias reduction for multiple-recapture estimators of closed population size. *Biometrics*, 50(2), 388–395.
- Evans, M. A., Bonett, D. G., & McDonald, L. L. (1994). A general theory for modelling capture-recapture data from a closed population. *Biometrics*, 50(2), 396–405.
- Evans, N. T., Olds, B. P., & et al. (2016). Quantification of mesocosm fish and amphibian species diversity via environmental DNA metabarcoding. *Molecular Ecology Resources*, 16(1), 29–41.
- Facon, M., Pinault, M., & et al. (2016). A comparative study of the accuracy and effectiveness of Line and Point Intercept Transect methods for coral reef monitoring in the Southwestern Indian Ocean Islands. *Ecological Indicators*, 60, 1045–1055.
- Falke, J. A., Bailey, L. L., Fausch, K. D., & Bestgen, K. R. (2012). Colonization and extinction in dynamic habitats: An occupancy approach for a Great Plains stream fish assemblage. *Ecology*, 93(4), 858–867.
- Farcomeni, A. (2011). Recapture models under equality constraints for the conditional capture probabilities. *Biometrika*, 98(1), 237–242.
- Farcomeni, A. (2015). Latent class recapture models with flexible behavioural response. *Statistica*, 75(1), 5–17.
- Farcomeni, A. (2016). A general class of recapture models based on the conditional capture probabilities. *Biometrics*, 72(1), 116–124.
- Farcomeni, A., & Scacciatielli, D. (2013). Heterogeneity and behavioral response in continuous time capture-recapture, with application to street cannabis use in Italy. *Annals of Applied Statistics*, 7(4), 2293–2314.
- Farcomeni, A., & Tardella, L. (2010). Reference Bayesian methods for recapture models with heterogeneity. *Test*, 19(1), 187–208.
- Farnsworth, G. L., Nichols, J. D., & et al. (2005). Statistical approaches to the analysis of point count data: A little extra information can go a long way. In R. C. John & T. D. Rich (Eds.), *Bird conservation implementation and integration in the Americas*. Proceedings of the Third International Partners in Flight Conference, 2002. March 20–24; Asilomar, California (Vol. 2, pp. 736–743).
- Farnsworth, G. L., Pollock, K. H., & et al. (2002). A removal model for estimating detection probabilities from point-count surveys. *Auk*, 119(2), 414–425.

- Farzana, A., Parry, M., & Fletcher, D. (2020). Estimating overdispersion in sparse multinomial data. *Biometrics*, 76(3), 834–842.
- Fattorini, L., Puletti, N., & et al. (2016). Checking the performance of point and plot sampling on aerial photoimagery of a large-scale population of trees outside forests. *Canadian Journal of Forest Research*, 46(11), 1264–1274.
- Fegatelli, D. A., & Farcomeni, A. (2016). On the design of closed recapture experiments. *Biometrical Journal*, 58(6), 1273–1294.
- Fegatelli, D. A., & Tardella, L. (2012). Improved inference on capture recapture models with behavioural effects. *Statistical Methods and Applications*, 22(1), 45–66.
- Fehrmann, L., Gregoire, T. G., & Kleinn, C. (2012). Triangulation based inclusion probabilities: A design-unbiased sampling approach. *Environmental and Ecological Statistics*, 19(1), 107–123.
- Fewster, R. M. (2011). Variance estimation for systematic designs in spatial surveys. *Biometrics*, 67(4), 1518–1531.
- Fewster, R. M., & Buckland, S. T. (2004). Assessment of distance sampling estimators. In S. T. Buckland, D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers, & L. Thomas (Eds.), *Advanced distance sampling* (pp. 281–306). Oxford: Oxford University Press.
- Fewster, R. M., Buckland, S. T., & et al. (2009). Estimating the encounter rate variance in distance sampling. *Biometrics*, 65(1), 225–236.
- Fewster, R. M., & Jupp, P. E. (2009). Inference on population size in binomial detectability models. *Biometrika*, 96(4), 805–820.
- Fewster, R. M., & Pople, A. R. (2008). A comparison of mark–recapture distance-sampling methods applied to aerial surveys of eastern grey kangaroos. *Wildlife Research*, 35(4), 320–330.
- Fewster, R. M., Southwell, C., Borchers, D. L., Buckland, S. T., & Pople, A. R. (2008). The influence of animal mobility on the assumption of uniform distances in aerial line transect surveys. *Wildlife Research*, 35(4), 275–288.
- Fewster, R. M., Stevenson, B. C., & Borchers, D. L. (2008). The influence of animal mobility on the assumption of uniform distances in aerial line transect surveys. *Wildlife Research*, 35(4), 275–288.
- Fewster, R. M., Stevenson, B. C., & Borchers, D. L. (2016). Trace- contrast models for capture-recapture without capture histories. *Statistical Science*, 31(2), 245–258.
- Ficetola, G. F., Pansu J., & et al. (2015). Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. *Molecular Ecology Resources*, 15(3), 543–556.
- Ficetola, G. F., Taberlet, P., & Coissac, E. (2016). How to limit false positives in environmental DNA and metabarcoding? *Molecular Ecology Resources*, 16(3), 604–607.
- Fidino, M., Simonis, J. L., & Magle, S. B. (2019). A multistate dynamic occupancy model to estimate local colonization-extinction rates and patterns of co-occurrence between two or more interacting species. *Methods in Ecology and Evolution*, 10(2), 233–244.
- Fielding, A. H., & Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, 24(1), 38–49.
- Fienberg, S. E. (1972). The multiple recapture census for closed populations and incomplete 2^k contingency tables. *Biometrika*, 59(3), 591–603.
- Fienberg, S. E., Johnson, M. S., & Junker, B. W. (1999). Classical multilevel and Bayesian approaches to population size estimation using multiple lists. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 162(3), 383–405.
- Fienberg, S. E. (1992). Bibliography on capture-recapture modelling with application to census undercount adjustment. *Survey Methodology*, 18(1), 143–154.
- Fisher, J. T., & Bradbury S. A. (2014). A multi-method hierarchical modeling approach to quantifying bias in occupancy from noninvasive genetic tagging studies. *Journal of Wildlife Management*, 78(6), 1087–1095.
- Fisher, J. T., Heim, N., Code, S., & Paczkowski, J. (2016). Grizzly bear noninvasive genetic tagging surveys: Estimating the magnitude of missed detections. *PLoS One*, 11(9), e0161055.

- Fisher, R. A., Corbet, A. S., & Williams, C. B. (1943). The relation between the number of species and the number of individuals in a random sample of an animal population. *Journal of Animal Ecology*, 12, 42–58.
- Fiske, I. J., & Chandler, R. B. (2011). Unmarked: An R package for fitting hierarchical models of wildlife occurrence and abundance. *Journal of Statistical Software*, 43, 1–23.
- Fithian, W., Elith, J., Hastie, T., Keith, D. A., & O'Hara, R. B. (2015). Bias correction in species distribution models: Pooling survey and collection data for multiple species. *Methods in Ecology and Evolution*, 6(4), 424–438.
- Fithian, W., & Hastie, T. (2013). Finite-sample equivalence in statistical models for presence-only data. *Annals of Applied Statistics*, 7(4), 1917–1939.
- Fitzpatrick, M. C., Preisser, E. L., Ellison, A. M., & Elkinton, J. S. (2009). Observer bias and the detection of low density populations populations. *Ecological Applications*, 19(7), 1673–1679.
- Fletcher, D. (2018). *Model averaging*. Heidelberg: Springer Briefs in Statistics.
- Foley, K., Rosenberger, A., & Mueter, F. (2015). Effectiveness of single-pass backpack electrofishing to estimate juvenile coho salmon abundance in Alaskan headwater streams. *Fisheries Science*, 81(4), 601–610.
- Forbes, A. D. (1995). Classification algorithm evaluation: Five performance measures based on confusion matrices. *Journal of Clinical Monitoring*, 11(3), 189–206.
- Forcey, G. M., Anderson, J. T., Ammer, F. K., & Whitmore, R. C. (2006). Comparison of two double-observer point-count approaches for estimating breeding bird abundance. *Journal of Wildlife Management*, 70(6), 1674–1681.
- Foster, R. J., & Harmsen, B. J. (2012). A critique of density estimation from camera-trap data. *Journal of Wildlife Management*, 76(2), 224–236.
- Franklin, J. (2013). Species distribution models in conservation biogeography: Developments and challenges. *Diversity and Distributions*, 19(10), 1217–1223.
- Fraver, S., Ducey, M. J., Woodall, C. W., D'Amato, A. W., Milo, A. M., & Palik, B. J. (2018). Influence of transect length and downed woody debris abundance on precision of the line-intersect sampling method. *Forest Ecosystems*, 5(1), 1–10.
- Freeman, P. R. (1973). Sequential recapture. *Biometrika*, 60(1), 141–153.
- Froidevaux J. S. P., Zellweger, F., Bollmann, K., & Obрист, M. K. (2014). Optimizing passive acoustic sampling of bats in forests. *Ecology and Evolution*, 4(24), 4690–4700.
- Frusher, S. D., Hoenig, J. M., & Ihde, T. (2007). Evaluating catchability assumptions for change-in-ratio and index-removal estimators, with application to southern rock lobster. *Fisheries Research*, 84(2), 254–262.
- Fuller, A. K., Linden, D. W., & Royle, J. A. (2016). Management decision making for fisher populations informed by occupancy modeling. *Journal of Wildlife Management*, 80(5), 794–802.
- Fullman, R. L. (1953). Measurement of particle sizes in opaque bodies. *Journal of Metals*, 5(3), 447–452.
- Furnas, B. J., Landers, R. H., Callas, R. L., & Matthews, S. M. (2017). Estimating population size of Fishers (*Pekania pennanti*) using camera stations and auxiliary data on home range size. *Ecosphere*, 8(3), e01747.
- Galipaud, M., Gillingham, M. A. F., David, M., & Dechaume-Moncharmont, F.-X. (2014). Ecologists overestimate the importance of predictor variables in model averaging: A plea for cautious interpretations. *Methods in Ecology and Evolution*, 5(10), 983–991.
- Gálvez, N., Guillera-Arroita, G., Morgan, B. J. T., & Davies, Z. G. (2016). Cost-efficient effort allocation for camera-trap occupancy surveys of mammals. *Biological Conservation*, 204, 350–359.
- Gao, M. (2013). Detecting spatial aggregation from distance sampling: A probability distribution model of nearest neighbor distance. *Ecological Research*, 28(3), 397–405.
- García, M. B., Silva, J. L., Tejero, P., & Pardo, I. (2021). Detecting early-warning signals of concern in plant populations with a Citizen Science network. Are threatened and other priority species for conservation performing worse? *Journal of Applied Ecology*, 58(7), 1388–1398.

- Gardiner, M. M., Allee, L. L., Brown, P. M. J., Losey, J. E., Roy, H. E., & Smyth, R. R. (2012). Lessons from lady beetles: Accuracy of monitoring data from US and UK citizen-science programs. *Frontiers in Ecology and the Environment*, 10(9), 471–476.
- Gardner, B., Reppucci, J., Lucherini, M., & Royle, J. A. (2010). Spatially explicit inference for open populations: Estimating demographic parameters from camera-trap studies. *Ecology*, 91(11), 3376–3383.
- Gardner, B., Royle, J. A., & Wegan, M. T. (2009). Hierarchical models for estimating density from DNA mark-recapture studies. *Ecology*, 90(4), 1106–1115.
- Gardner, B., Royle, J. A., Wegan, M. T., Rainbolt, R. E., & Curtis, P. D. (2010). Estimating black bear density using DNA data from hair snares. *Journal of Wildlife Management*, 74(2), 318–325.
- Gardner, C. L., Lawler, J. P., Ver Hoef, J. M., Magoun, A. J., & Kellie, K.A. (2010). Coarse-scale distribution surveys and occurrence probability modeling for wolverine in interior Alaska. *Journal of Wildlife Management*, 74(8), 1894–1903.
- Garrard, G. E., Bekessy, S. A., McCarthy, M. A., & Wintle, B. A. (2008). When have we looked hard enough? A novel method for setting minimum survey effort protocols for flora surveys. *Austral Ecology*, 33(8), 986–998.
- Garrard, G. E., Bekessy, S. A., McCarthy, M. A., & Wintle, B. A. (2015). Incorporating detectability into environmental impact assessment for threatened species. *Conservation Biology*, 29(1), 216–225.
- Garrard, G. E., McCarthy, M. A., Williams, N. S. G., Bekessy, S. A., & Wintle, B. A. (2013). A general model of detectability using species traits. *Methods in Ecology and Evolution*, 4(1), 45–52.
- Garthwaite, P. H., Yu, K., & Hope, P. B. (1995). Bayesian analysis of a multiple-recapture model. *Communications in Statistics—Theory and Methods*, 24(9), 2229–2247.
- Gaskell, T. J., & George, B. J. (1972). A Bayesian modification of the Lincoln index. *Journal of Applied Ecology*, 9(2), 377–384.
- Gates, C. E., Marshall, W. H., & Olson, D. P. (1968). Line transect method of estimating grouse population densities. *Biometrics*, 24(1), 135–145.
- Gatz, A. J., Jr., & Loar, J. M. (1988). Petersen and removal population size estimates: Combining methods to adjust and interpret results when assumptions are violated. *Environmental Biology of Fishes*, 21(4), 293–307.
- Gazey, W. J., & Staley, M. J. (1986). Population estimation from mark- recapture experiments using a sequential Bayes algorithm. *Ecology*, 67(4), 941–951.
- Geissler, P. H., & Fuller, M. R. (1987). Estimation of the proportion of area occupied by an animal species. In *Proceedings of the Section on Survey Research Methods of the American Statistical Association 1986* (pp. 533–538).
- Gelfand, A. E., Schmidt, A. M., Wu, S., Silander, Jr., J. A., Latimer, A., & Rebelo, A. G. (2005). Modelling species diversity through species level hierarchical modelling. *Applied Statistics*, 54(1), 1–20.
- Gelfand, A. E., Schmidt, A. M., Banerjee, S., & Sirmans, C. (2004). Nonstationary multivariate process modeling through spatially varying coregionalization. *Test*, 13(2), 263–312.
- Gelfand, A. E., & Dey, D. K. (1994). Bayesian model choice: Asymptotics and exact calculations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 56(3), 501–514.
- Gelfand, A. E., & Shirota, S. (2019). Clarifying species dependence under joint species distribution modeling. arXiv:1908.09410 [stat.ME].
- Gelfand, A., Silander, Jr., J., Wu, S., Latimer, A., Lewis, P., Rebelo, A., & Holder, M. (2006). Explaining species distribution patterns through hierarchical modeling. *Bayesian Analysis*, 1(1), 41–92.
- Gelfand, A. E., & Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85(410), 398–409.
- Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. (Comment on an article by Browne and Draper). *Bayesian Analysis*, 1(3), 515–533.

- Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y. S. (2008). A weakly informative default prior distribution for logistic and other regression models. *Annals of Applied Statistics*, 2(4), 1360–1383.
- Gelman, A., & Shalizi, C. (2013). Philosophy and the practice of Bayesian statistics. *British Journal of Mathematical and Statistical Psychology*, 66(1), 8–38.
- George, E. I. (1992). Capture-recapture estimation via Gibbs sampling. *Biometrika*, 79(4), 677–683.
- George, J. C., Zeh, J., Suydam, R., & Clark, C. (2004). Abundance and population trend (1978–2001) of western Arctic bow-head whales surveyed near Barrow, Alaska. *Marine Mammal Science*, 20(4), 755–773.
- Gerber, B. D., Karpanty, S. M., & Kelly, M. J. (2012). Evaluating the potential biases in carnivore capture–recapture studies associated with the use of lure and varying density estimation techniques using photographic-sampling data of the Malagasy civet. *Population Ecology*, 54(1), 43–54.
- Germano, T. M., Field, K. J., & et al. (2015). Mitigation-driven translocations: Are we moving wildlife in the right direction? *Frontiers in Ecology and the Environment*, 13(2), 100–105.
- Getz, W. M., Fortmann-Roe, C., Cross, P. C., Lyons, A. J., Ryan, S. J., & Wilmers, C. C. (2007). LoCoH: Nonparametric kernel methods for constructing home ranges and utilization distributions. *PLoS One*, 2(2), e207.
- Getz, W. M., & Wilmers, C. C. (2004). A local nearest-neighbor convex-hull construction of home ranges and utilization distributions. *Ecography*, 27(4), 489–505.
- Ghosh, S. K., & Norris, J. L. (2005). Bayesian capture-recapture analysis and model selection allowing for heterogeneity and behavioral effects. *Journal of Agricultural, Biological, and Environmental Statistics*, 10(1), 35–49.
- Gibson-Reinemer, D. K., Stewart, D. R., Fritts, M. W., DeBoer, J. A., & Casper, A. F. (2016). Estimating the effects of environmental variables and gear type on the detection and occupancy of large-river fishes in a standardized sampling program using multiseason bayesian mixture models. *North American Journal of Fisheries Management*, 36(6), 1445–1456.
- Gilks, W. R., & Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics*, 41(2), 337–348.
- Gimenez, O., Blanc, L., & et al. (2014). Fitting occupancy models with E-SURGE: Hidden Markov modelling of presence–absence data. *Methods in Ecology and Evolution*, 5(6), 592–597.
- Gimenez, O., Bonner, S. J., & et al. (2009). WinBUGS for population ecologists: Bayesian modeling using Markov Chain Monte Carlo methods. In D. L. Thomson, E. G. Cooch, & M. J. Conroy (Eds.), *Modeling demographic processes in marked populations* (pp. 885–918). Environmental and Ecological Statistics (Vol. 3). Springer.
- Gimenez, O., Choquet, R., & et al. (2005). Efficient profile-likelihood confidence intervals for capture-recapture models. *Journal of Agricultural, Biological, and Environmental Statistics*, 10(2), 184–196.
- Giraud, C., Calenge, C., Coron, C., & Juliard, R. (2016). Capitalizing on opportunistic data for monitoring relative abundances of species. *Biometrics*, 72(2), 649–658.
- Givens, G. H., Edmondson, S. L., George, J. C., Tudor, B., DeLong, R. A., & Suydam, R. (2015). Weighted likelihood recapture estimation of detection probabilities from an ice-based survey of bowhead whales. *Environmetrics*, 26(1), 1–16.
- Glennie, R., Borchers, D. L., Murchie, M., Harmsen, B. J., & Foster, R. J. (2019). Open population maximum likelihood spatial capture-recapture. *Biometrics*, 75(4), 134–1355.
- Glennie, R., Buckland, S. T., & Thomas, L. (2015). The effect of animal movement on line transect estimates of abundance. *PLoS One*, 10(3), e0121333.
- Glennie, R., Buckland, S. T., & et al. (2021). Incorporating animal movement into distance sampling. *Journal of the American Statistical Association*, 116(533), 107–115.
- Godsoe, W. & Harmon, L. J. (2012). How do species interactions affect species distribution models? *Ecography* 35(9), 811–820.

- Goijman, A. P., Conroy, M. J., Bernardos, J. N., & Zaccagnini, M. E. (2015). Multi-season regional analysis of multi-species occupancy: Implications for bird conservation in agricultural lands in east-central Argentina. *PLoS One*, 10(6), e0130874.
- Gold, S. J., Wibert, W. N., Bondartsova, V., Biroscak, B. J., & Post, L. A. (2015). A capture-recapture approach to estimation of refugee populations. *International Migration*, 53(5), 3–25.
- Golding, N. & Harris, D. J. (2015). BayesComm: Bayesian community ecology analysis. – R package ver. 0.1–2.
- Golding, J. D., & Dreitz, V. J. (2016). Comparison of removal based methods for estimating abundance of five species of prairie songbirds. *Journal of Field Ornithology*, 87(4), 417–426.
- Golding, J. D., Nowak, J. J., & Dreitz, V. J. (2017). A multispecies dependent double-observer model: A new method for estimating multispecies abundance. *Ecology and Evolution*, 7(10), 3425–3435.
- Golding, N., Nunn, M. A., & Purse, B. V. (2015). Identifying biotic interactions which drive the spatial distribution of a mosquito community. *Parasites and Vectors*, 8(1), 367–377.
- Gómez de Segura, A., Toás, J., Pedraza, S. N., Crespo, E. A., & Raga, J. A. (2006). Abundance and distribution of the endangered loggerhead turtle in Spanish Mediterranean waters and its conservational implications. *Animal Conservation*, 9(2), 199–206.
- Gómez-Rodríguez, C., Bustamante, J., Díaz-Paniagua, C., & Guisan, A. (2012). Integrating detection probabilities in species distribution models of amphibians breeding in Mediterranean temporary ponds. *Diversity and Distributions*, 18(3), 260–272.
- Good, I. J. (1953). The population frequencies of species and the estimation of population parameters. *Biometrika*, 40(3–4), 237–264.
- Good, I. J. (1967). A Bayesian significance test for multinomial distributions. *Journal of the Royal Statistical Society: Series B*, 29(3), 399–418.
- Good, I. J. (2000). Turing's anticipation of empirical Bayes in connection with the cryptanalysis of the naval Enigma. *Journal of Statistical Computation and Simulation*, 66(2), 101–111.
- Good, I. J., Lewis, B. C., Gaskins, R. A., & Howell, L. W. (1979). Population estimation by the removal method assuming proportional trapping. *Biometrika*, 66(3), 485–494.
- Goodman, L. A. (1960). On the exact variance of products. *Journal of the American Statistical Association*, 55(292), 708–713.
- Goodsoe, W., & Harmon, L. J. (2012). How do species interactions affect species distribution models? *Ecography*, 35(9), 811–820.
- Gopalaswamy, A. M., Royle, J. A., Delampady, M., Nichols, J. D., Karanth, K. U., & Macdonald, D. W. (2012). Density estimation in tiger populations: Combining information for strong inference. *Ecology*, 93(7), 1741–1751.
- Gopalaswamy, A. M., Royle, J. A., Hines, J. E., & et al. (2012). Program SPACECAP: Software for estimating animal density using spatially explicit capture-recapture models. *Methods in Ecology and Evolution*, 3(6), 1067–1072.
- Gormley, A. M., Forsyth, D. M., & et al. (2011). Using presence-only and presence-absence data to estimate the current and potential distributions of established invasive species. *Journal of Applied Ecology*, 48(1), 25–34.
- Gorresen, P. M., Miles, A. C., Todd, C. M., Bonaccorso, F. J., & Weller, T. J. (2013). Assessing bat detectability and occupancy with multiple automated echolocation detectors. *Journal of Mammalogy*, 89(1), 11–17.
- Gotelli, N. J., & Chao, A. (2013). Measuring and estimating species richness, species diversity, and biotic similarity from sampling data. In S. A. Levin (Ed.), *Encyclopedia of biodiversity* (2nd ed., Vol. 5, pp. 195–211). Waltham, MA: Academic Press.
- Goudie, I. B. J., & Gormley, R. (2013). Maximum likelihood estimates for the Schnabel census with plants. *Communications in Statistics. Theory and Methods*, 42(20), 3704–3715.
- Goudie, I. B. J., Jupp, P. E., & Ashbridge, J. (2007). Plant-capture estimation of the size of a homogeneous population. *Biometrika*, 94(1), 243–248.
- Gould, W. R., Patla, D. A., & et al. (2012). Estimating occupancy in large landscapes: Evaluation of amphibian monitoring in the Greater Yellowstone Ecosystem. *Wetlands*, 32(2), 379–389.

- Gould, W. R., & Pollock, K. H. (1997). Catch-effort maximum likelihood estimation of important population parameters. *Biometrics*, 53(1), 207–216.
- Gould, W. R., Ray, A. M., Bailey, L. L., Thoma, D., Daley, R., & Legg, K. (2019). Multi-state occupancy modeling improves understanding of amphibian breeding dynamics in the Greater Yellowstone Area. *Ecological Applications*, 29(1), e01825.
- Graham, R. L., Knuth, D. E., & Patashnik, O. (1988). *Concrete mathematics*. Reading, MA, USA: Addison-Wesley.
- Graham, W. M., Martin, D. L., & Martin, J. C. (2003). In situ quantification and analysis of large jellyfish using a novel video profiler. *Marine Ecology Progress Series*, 254, 129–140.
- Gray, H. L., & Schucany, W. R. (1972). *The generalized jackknife statistic*. New York, N.Y: Marcel Dekker.
- Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82(4), 711–732.
- Gregoire, T. G. (1982). The unbiasedness of the mirage correction procedure for boundary overlap. *Forest Science*, 28(3), 504–508.
- Gregoire, T. G. & Scott, C. T. (1990). Sampling at the stand boundary: A comparison of the statistical performance among eight Methods. *Research in Forest Inventory, Monitoring, Growth and Yield*. Proceedings of the International Union of Forest Research Organizations XIX World Congress, Montreal, Canada, 5–11 August, 1990, eds. H. E. Burkhart, G. M. Bonnor, and J. J. Lowe, Publ. FWS-3-90, School of Forestry and Wildlife Resources, Virginia Polytechnic Institute and State University, pp. 78–85.
- Gregoire, T. G., & Scott, C. T. (2003). Altered selection probabilities caused by avoiding the edge in field surveys. *Journal of Agricultural, Biological, and Environmental Statistics*, 8(1), 36–47.
- Gregoire, T. G., & Valentine, H. T. (2008). *Sampling strategies for natural resources and the environment*. New York: Chapman and Hall/CRC.
- Greig-Smith, P. (1964). *Quantitative plant ecology* (2nd ed.). Washington, D.C.: Butterworth.
- Griffiths, M., & van Schaik, C. P. (1993). The impact of human traffic on the abundance and activity periods of Sumatran rain forest wildlife. *Conservation Biology*, 7(3), 623–626.
- Groce, M. C., Bailey, L. L., & Fausch, K. D. (2012). Evaluating the success of Arkansas darter translocations in Colorado: An occupancy sampling approach. *Transactions of the American Fisheries Society*, 141(3), 825–840.
- Grosenbaugh, L. R. (1952). Plotless timber estimates—new, fast, easy. *Journal of Forestry*, 50, 32–37.
- Grosenbaugh, L. R. (1958). Point-sampling and line-intercept sampling: Probability theory, geometric implications, synthesis. In *Southern Forest Experiment Station Occasional Paper SO-160*. Washington, D.C.: USDA Forest Service.
- Gu, W., & Swihart, R. K. (2004). Absent or undetected? Effects of non-detection of species occurrence on wildlife-habitat models. *Biological Conservation*, 116(2), 195–203.
- Guillera-Arroita, G. (2011). Impact of sampling with replacement in occupancy studies with spatial replication. *Methods in Ecology and Evolution*, 2(4), 401–406.
- Guillera-Arroita, G. (2017). Modelling of species distributions, range dynamics and communities under imperfect detection: Advances, challenges and opportunities. *Ecography*, 40(2), 281–295.
- Guillera-Arroita, G., & Lahoz-Monfort, J. J. (2012). Designing studies to detect differences in species occupancy: Power analysis under imperfect detection. *Methods in Ecology and Evolution*, 3(5), 860–869.
- Guillera-Arroita, J., Lahoz-Monfort, J., & et al. (2015). Is my species distribution model fit for purpose? Matching data and models to applications. *Global Ecology and Biogeography*, 24(3–4), 276–292.
- Guillera-Arroita, G., & Lahoz-Monfort, J. J. (2017). Species occupancy estimation and imperfect detection: Shall surveys continue after the first detection? *Advances in Statistical Analysis*, 101(4), 381–398.

- Guillera-Arroita, G., Lahoz-Monfort, J. J., MacKenzie, D. I., Wintle, B. A., & McCarthy, M. A. (2014). Ignoring imperfect detection in biological surveys Is dangerous: A response to 'Fitting and interpreting occupancy models'. *PLoS One*, 9(7), e99571.
- Guillera-Arroita, G., Lahoz-Monfort, J. J., van Rooyen, A. R., Weeks, A. R., & Tingley, R. (2017). Dealing with false positive and false negative errors about species occurrence at multiple levels. *Methods in Ecology and Evolution*, 8(9), 1081–1091.
- Guillera-Arroita, G., Morgan, B. J. T., Ridout, M. S., & Linkie, M. (2011). Species occupancy modeling for detection data collected along a transect. *Journal of Agricultural, Biological, and Environmental Statistics*, 16(3), 301–318.
- Guillera-Arroita, G., Morgan, B. J. T., Ridout, M. S., & Linkie, M. (2012). Models for species-detection data collected along transects in the presence of abundance-induced heterogeneity and clustering in the detection process. *Methods in Ecology and Evolution*, 3(2), 358–367.
- Guillera-Arroita, G., Ridout, M. S., & Morgan, B. J. T. (2010). Design of occupancy studies with imperfect detection. *Methods in Ecology and Evolution*, 1(2), 131–139.
- Guillera-Arroita, G., Ridout, M. S., & Morgan, B. J. T. (2014). Two-stage bayesian study design for species occupancy estimation. *Journal of Agricultural, Biological, and Environmental Statistics*, 19(2), 278–291.
- Guisan, A., Edwards, T. C., & Hastie, T. (2002). Generalized linear and generalized additive models in studies of species distributions: Setting the scene. *Ecological Modelling*, 157(2), 89–100.
- Guisan, A., Graham, C. H., Elith, J., Huettmann, F., & NCEAS Species Distribution Modelling Group. (2007). Sensitivity of predictive species distribution models to change in grain size. *Diversity and Distributions*, 13(3), 332–340.
- Guisan, A., Lehmann, A., & et al. (2006). Making better biogeographical predictions of species' distributions. *Journal of Applied Ecology*, 43(3), 386–392.
- Guisan, A., & Rahbek, C. (2011). SESAM—a new framework integrating macroecological and species distribution models for predicting spatio-temporal patterns of species assemblages. *Journal of Biogeography*, 38(8), 1433–1444.
- Guisan, A., & Thuiller, W. (2005). Predicting species distribution: Offering more than simple habitat models. *Ecology Letters*, 8(9), 993–1009.
- Guisan, A., Zimmermann, N. E., Elith, J., Graham, C. H., Phillips, S., & Peterson, A. T. (2007). What matters for predicting the occurrences of trees: Techniques, data, or species' characteristics? *Ecological Monographs*, 77(4), 615–630.
- Haberman, Shelby J. (1973). The analysis of residuals in cross-classified tables. *Biometrics*, 29(1), 205–220.
- Hackett, H. M., Lesmeister, D. B., Desanty-Combes, J., Montague, W. G., Millspaugh, J. J., & Gompper, M. E. (2007). Detection rates of eastern spotted skunks (*Spilogale putorius*) in Missouri and Arkansas using live-capture and non-invasive techniques. *American Midland Naturalist*, 158(1), 123–131.
- Haines, L. M. (2016). A note on the Royle–Nichols model for repeated detection-nondetection data. *Journal of Agricultural, Biological, and Environmental Statistics*, 21(3), 588–598.
- Haines, L. M. (2020). Multinomial N-mixture models for removal sampling. *Biometrics*, 76(2), 540–548.
- Halstead, B. J., Kleeman, P. M., & Rose, J. P. (2018). Time-to-detection occupancy modeling: An efficient method for analyzing the occurrence of amphibians and reptiles. *Journal of Herpetology*, 52(4), 416–425.
- Hammond, P. S., Berggren, P., & et al. (2002). Abundance of harbour porpoise and other cetaceans in the North Sea and adjacent waters. *Journal of Applied Ecology*, 39(2), 361–376.
- Hammond, P. S., Macleod, K., & et al. (2013). Cetacean abundance and distribution in European Atlantic shelf waters to inform conservation and management. *Biological Conservation*, 164, 107–122.
- Hanks, E. M., Hooten, M. B., & Baker, F. A. (2011). Reconciling multiple data sources to improve accuracy of large-scale prediction of forest disease incidence. *Ecological Applications*, 21(4), 1173–1188.

- Hardersen, S., & Corezzola, S. (2014). Plot-based butterfly surveys: Statistical and methodological aspects. *Journal of Insect Conservation*, 18(6), 1171–1183.
- Harding, C. M., Heathwood, A. W., Hunt, R. G., & Read, K. L. Q. (1984). The estimation of animal population size by the removal method. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 33 (2), 196–202.
- Harris, D. J. (2015). Generating realistic assemblages with a joint species distribution model. *Methods in Ecology and Evolution*, 6(4), 465–473.
- Harris, D. J. (2016). Inferring species interactions from co-occurrence data with Markov networks. *Ecology*, 97(12), 3308–3314.
- Harris, J. E., Jolley, J. C., Silver, G. S., Yuen, H., & Whitesel, T. A. (2016). An experimental evaluation of electrofishing catchability and catch depletion abundance estimates of larval lampreys in a wadeable stream: Use of a hierarchical approach. *Transactions of the American Fisheries Society*, 145(5), 1006–1017.
- Hartill, B. W., Taylor, S. M., Keller, K., & Weltersbach, M. S. (2020). Digital camera monitoring of recreational fishing effort: Applications and challenges. *Fish and Fisheries*, 21(1), 204–215.
- Hastie, T., & Fithian, W. (2013). Inference from presence-only data; the ongoing controversy. *Ecography*, 36(8), 864–867.
- Hawkes A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1), 83–90.
- Haxtema, Z., Temesgen, H., & Marquardt, T. (2012). Evaluation of n-tree distance sampling for inventory of headwater riparian forests of western Oregon. *Western Journal of Applied Forestry*, 27(3), 109–117.
- Hay, G., & Richardson, C. (2016). Estimating the prevalence of drug use using mark-recapture methods. *Statistical Science*, 31(2), 191–204.
- Hayes, R. J., & Buckland, S. T. (1983). Radial-distance models for the line-transect method. *Biometrics*, 39(1), 29–42.
- Hayne, D. W. (1949a). An examination of the strip census method for estimating animal populations. *The Journal of Wildlife Management*, 13(2), 145–157.
- Hayne, D. W. (1949b). Two methods for estimating populations from trapping records, *Journal of Mammalogy*, 39, 399–411.
- Hazen, E. L., Abrahms, B., Brodie, S., Carroll, G., Welch, H., & Bograd, S. J. (2021). Where did they not go? Considerations for generating pseudo-absences for telemetry-based habitat models. *Movement Ecology*, 9(1), 5.
- Heard, G. W., McCarthy, M. A., Scroggie, M. P., Baumgartner, J. B., & Parris, K. M. (2013). A Bayesian model of metapopulation viability, with application to an endangered amphibian. *Diversity and Distributions*, 19(5–6), 555–566.
- Hedley, S. H., & Buckland, S. T. (2004). Spatial models for line transect sampling. *Journal of Agricultural, Biological, and Environmental Statistics*, 9(2), 181–199.
- Hefley, T. J., & Hooten, M. B. (2016). Hierarchical species distribution models. *Current Landscape Ecology Reports*, 1(2), 87–97.
- Hefley, T. J., Hooten, M. B., Hanks, E. M., Russell, R. E., & Walsh, D. P. (2017a). Dynamic spatio-temporal models for spatial data. *Spatial Statistics*, 20, 206–220.
- Held, L., & Sabanés Bové, D. (2020). *Likelihood and Bayesian inference with applications in biology and medicine* (2nd ed.). Springer Nature.
- Henderson, P. A. (2021). *Southwoods ecological methods* (5th ed.). Chichester: Wiley.
- Henry, D. A. W., Lee, A. T. K., & Altwegg, R. (2020). Can time-to-detection models with fewer survey replicates provide a robust alternative to traditional site-occupancy models? *Methods in Ecology and Evolution*, 11 (5), 643–655.
- Herliansyah, R., King, R., & King, S. (2022). Laplace approximations for capture-recapture models in the presence of individual heterogeneity. *Journal of Agricultural, Biological and Environmental Statistics*, 27(3), 401–418.
- Hiby, L., & Krishna, M. B. (2001). Line transect sampling from a curving path. *Biometrics*, 57 (3), 727–731.
- Hiby, L., & Lovell, P. (1998). Using aircraft in tandem formation to estimate abundance of harbour porpoise. *Biometrics*, 54(4), 1280–289.

- Higa, M., Yamaura, Y., Koizumi, I., Yabuhara, Y., Senzaki, M., & Ono, S. (2015). Mapping large-scale bird distributions using occupancy models and citizen data with spatially biased sampling effort. *Diversity and Distributions*, 21(1), 46–54.
- Hilbe, J. M. (2011). *Negative binomial regression: Modelling overdispersed count data*. Cambridge: Cambridge University Press.
- Hill, B. M. (1968). Posterior distribution of percentiles: Bayes' theorem for sampling from a population. *Journal of the American Statistical Association*, 63(322), 677–691.
- Hill, M. O. (1973). Diversity and evenness: A unifying notation and its consequences. *Ecology*, 54(2), 427–432.
- Hill, B. M. (1979). Posterior moments of the number of species in a finite population and the posterior probability of finding a new species. *Journal of the American Statistical Association*, 74(36), 668–673.
- Hines, J. E., Nichols, J. D., & et al. (2010). Tigers on trails: Occupancy modeling for cluster sampling. *Ecological Applications*, 20(5), 1456–1466.
- Hines, J. E., Nichols, J. D., & Collazo, J. A. (2014). Multiseason occupancy models for correlated replicate surveys. *Methods in Ecology and Evolution*, 5(6), 583–591.
- Hirst, D. (1994). An improved removal method for estimating animal abundance. *Biometrics*, 50(2), 501–505.
- Hirzel, A. H., Helfer, V., & Metral, R. (2001). Assessing habitat-suitability models with a virtual species. *Ecological Modelling*, 145(2), 111–121.
- Hjort, N. L., & Claeskens, G. (2003). Frequentist model average estimators. *Journal of the American Statistical Association*, 98(464), 879–899.
- Hochachka, W. M., Fink, D., Hutchinson, R. A., Sheldon, D., Wong, W.-K., & Kelling, S. (2012). Data-intensive science applied to broad-scale citizen science. *Trends in Ecology and Evolution*, 27(2), 130–137.
- Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian model averaging: A tutorial. *Statistical Science*, 14(4), 382–417.
- Hogan, H. (2003). The accuracy and coverage evaluation: Theory and design. *Survey Methodology*, 29, 129–138.
- Holmes, S. B., McIlwrick, K. A., & Venier, L. A. (2014). Using automated sound recording and analysis to detect bird species-at-risk in southwestern Ontario woodlands. *Wildlife Society Bulletin*, 38(3), 591–598.
- Holzmann, H., Munk, A., & Zucchini, W. (2006). On identifiability in capture–recapture models. *Biometrics*, 62(3), 934–939.
- Hooten, M. B., & Hobbs, N. T. (2015). A guide to Bayesian model selection for ecologists. *Ecological Monographs*, 85(1), 3–28.
- Hooten, M. B., Wikle, C. K., Dorazio, R. M., & Royle, J. A. (2007). Hierarchical spatiotemporal matrix models for characterizing invasions. *Biometrics*, 63(2), 558–567.
- Hopper, K., Shinn, M., Laska, E., Meisner, M., & Wanderling, J. (2008). Estimating numbers of unsheltered homeless people through plant-capture and postcount survey methods. *American Journal of Public Health*, 98, 1438–1442.
- Horne, J. S., Garton, E. O., Krone, S. M., & Lewis, J. S. (2007). Analyzing animal movements using Brownian Bridges. *Ecology*, 88(9), 2354–2363.
- Hostetter, N. J., Gardner, B., Sillett, T., Pollock, K. H., & Simons, T. R. (2019). An integrated model decomposing the components of detection probability and abundance in unmarked populations. *Ecosphere*, 10(3), e02586.
- Hou, Z., Xu, Q., Hartikainen, S., & et al. (2015). Impact of plot size and spatial pattern of forest attributes on sampling efficacy. *Forest Science*, 61(5), 847–860.
- Houghton, J. D. R., Doyle, T. K., Davenport, J., & Hays, G. C. (2006). Developing a simple, rapid method for identifying and monitoring jellyfish aggregations from the air. *Marine Ecology Progress Series*, 314, 159–170.
- Howe, G. H. (1985). Use of computerized record linkage in follow-up studies of cancer epidemiology in Canada. *National Cancer Institute Monograph*, 67, 117–121.

- Hsieh, T. C., Ma, K. H., Chao, A., & McInerney, G. (2016). iNEXT: An R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods in Ecology and Evolution*, 7(12), 1451–1456.
- Hsu, Y.-H., Chen, Y., Yang, T. R., Kershaw, Jr., J. A., & Ducey, M. J. (2020). Sample strategies for bias correction of regional LiDAR-assisted forest inventory: Estimates on small woodlots. *Annals of Forest Science*, 77, 75.
- Huff, M. H., Bettinger, K. A., Ferguson, H. L., Brown, M. J., & Altman, B. (2000). *A habitat-based point-count protocol for terrestrial birds, emphasizing Washington and Oregon*. General Technical Report, PNW-GTR-501. Portland, OR; Pacific Northwest Research Station, Forest Service, U.S. Department of Agriculture.
- Huggins, R. M. (1989). On the statistical analysis of capture experiments. *Biometrika*, 76(1), 133–140.
- Huggins, R. (1991). Some practical aspects of a conditional likelihood approach to capture experiments. *Biometrics*, 47(2), 725–732.
- Huggins, R. (2001). A note on the difficulties associated with the analysis of capture–recapture experiments with heterogeneous capture probabilities. *Statistics and Probability Letters*, 54(2), 147–152.
- Huggins, R. (2004). Capture-recapture experiments with time heterogeneous capture probabilities in continuous time: A counting process approach. *Communications in Statistics. Theory and Methods*, 33(5), 1221–238.
- Huggins, R., & Chao, A. (2002). Asymptotic properties of an optimal estimating function approach to the analysis of mark recapture data. *Communications in Statistics—Theory and Methods*, 31(4), 575–595.
- Huggins, R. M., & Hwang, W. H. (2007). Non-parametric estimation of population size from capture-recapture data when capture probability depends on a covariate. *Journal of the Royal Statistical Society, Series C*, 56(4), 429–443.
- Huggins, R., & Hwang, W.-H. (2011). A review of the use of conditional likelihood in capture–recapture experiments. *International Statistical Review*, 79(3), 385–400.
- Hui, F. K. C. (2016). Bayesian ordination and regression analysis of multivariate abundance data in R. *Methods in Ecology and Evolution*, 7(6), 744–750.
- Hunter, M. E., Oyler-McCance, S. J., & et al. (2015). Environmental DNA (eDNA) sampling improves occurrence and detection estimates of invasive Burmese pythons. *PLoS One*, 10(4), e0121655.
- Hutchinson, R. A., Valente, J. J., Emerson, S. C., Betts, M. G., & Dietterich, T. G. (2015). Penalized likelihood methods improve parameter estimates in occupancy models. *Methods in Ecology and Evolution*, 6(8), 949–959.
- Hwang, W.-H., & Chao, A. (2002). Continuous-time capture-recapture models with covariates. *Statistica Sinica*, 12(4), 1115–1131.
- Hwang, W.-H., Chao, A., & Yip, P. S. F. (2002). Continuous-time capture–recapture models with time variation and behavioural response. *Australia and New Zealand Journal of Statistics*, 44(1), 41–45.
- Hwang, W.-H., & Huang, S. Y. H. (2003). Estimation in capture-recapture when covariates are subject to measurement errors. *Biometrics*, 59(4), 1113–1122.
- Hwang, W.-H., & Huggins, R. M. (2007). Application of semiparametric regression models in the analysis of capture-recapture experiments. *Australian and New Zealand Journal of Statistics*, 49(2), 191–202.
- Hwang, W.-H., & Huggins, R. (2011). A semiparametric model for a functional behavioural response to capture in capture-recapture experiments. *Australian and New Zealand Journal of Statistics*, 53(4), 403–421.
- Ihde, T. F., Hoenig, J. M., & Frusher, S. D. (2008a). Evaluation of a multi-year index-removal abundance estimator, with application to a Tasmanian rock lobster fishery. *Fisheries Research*, 89(1), 26–36.

- Ihde, T. F., Hoenig, J. M., & Frusher, S. D. (2008b). An index-removal abundance estimator that allows for seasonal change in catchability, with application to southern rock lobster *Jasus edwardsii*. *Transactions of the American Fisheries Society*, 137(3), 720–735.
- Iknayn K. J., Tingley, M. W., Furnas, B. J., & Beissinger, S. R. (2014). Detecting diversity: Emerging methods to estimate species diversity. *Trends in Ecology and Evolution*, 29(2), 97–106.
- Irigoyen, A. J., Rojo, I., Calò, A., Trobbiani, G., Sánchez-Carnero, N., & García-Charton, J. A. (2018). The “Tracked Roaming Transect” and distance sampling methods increase the efficiency of underwater visual censuses. *PLoS One*, 13(1), e0190990.
- Ivan, J. S., White, G. C., Shenk, T. M. & Cooch, E. G. (2013). Using simulation to compare methods for estimating density from capture-recapture data. *Ecology*, 94(4), 817–826.
- IWGDMF (International Working Group for Disease Monitoring and Forecasting). (1995a). Capture-recapture and multiple-record systems estimation: I. History and theoretical development. *American Journal of Epidemiology*, 142(10), 1047–1058.
- IWGDMF (International Working Group for Disease Monitoring and Forecasting). (1995b). Capture-recapture and multiple-record systems estimation: II. Applications in human diseases. *American Journal of Epidemiology*, 142(10), 1059–1068.
- Jamil, T., Ozinga, W. A., Kleyer, M., & ter Braak, C. J. F. (2013). Selecting traits that explain species–environment relationships: A generalized linear mixed model approach. *Journal of Vegetation Science*, 24(6), 988–1000.
- Jaro, M. A. (1989). Advances in record-linkage methodology as applied to matching the 1985 census of Tampa, Florida. *Journal of the American Statistical Association*, 84(406), 414–420.
- Jaro, M. A. (1995). Probabilistic linkage of large public health data files. *Statistics in Medicine*, 14(5–7), 491–498.
- Jarzyna, M. A., & Jetz, W. (2016). Detecting the multiple facets of biodiversity. *Trends in Ecology and Evolution*, 31(11), 527–538.
- Jathanna, D., Karanth, K. U., Kumar, N. S., Karanth, K. K., & Goswami, V. R. (2015). Patterns and determinants of habitat occupancy by the Asian elephant in the Western Ghats of Karnataka, India. *PLoS One*, 10(7), e0132323.
- Jaynes, E. T. (1957). Information theory and statistical mechanics. *Physical Review*, 106, 620–630.
- Jennelle, C. S., Cooch, E. G., Conroy, M. J., & Senar, J. C. (2007). State-specific detection probabilities and disease prevalence. *Ecological Applications*, 17(1), 154–167.
- Jerde, C. L., Mahon, A. R., Chadderton, W. L., & Lodge, D. M. (2011). ‘Sight-unseen’ detection of rare aquatic species using environmental DNA. *Conservation Letters*, 4(2), 150–157.
- Jewell, W. S. (1983). Bayesian estimation of undetected errors. Technical report, University of California, Berkley, Operations Research Center.
- Jeyam, A., McCrea, R. S., Bregnballe, T., Frederiksen, M., & Pradel, R. (2017). A test of positive association for detecting heterogeneity in capture for capture–recapture data. *Journal of Agricultural, Biological and Environmental Statistics*, 23(1), 1–19.
- Jiguet, F., Devictor, V., Julliard, R., & Couvet, D. (2012). French citizens monitoring ordinary birds provide tools for conservation and ecological sciences. *Acta Oecologica*, 44, 58–66.
- Johnson, C. J., Nielsen, S. E., Merrill, E. H., McDonald, T. L., & Boyce, M. S. (2006). Resource selection functions based on use-availability data: Theoretical motivation and evaluation method. *Journal of Wildlife Management*, 70(20), 347–357.
- Johnson, D. H. (2008). In defense of indices: The case of bird surveys. *Journal of Wildlife Management*, 72(4), 857–868.
- Johnson, D. S. (2012). stocc: Fit a spatial occupancy model via Gibbs sampling. R package version 1.0-7.
- Johnson, D. S., & Hoeting, J. A. (2003). Autoregressive models for capture- recapture data: A Bayesian approach. *Biometrics*, 59(2), 340–349.
- Johnson, D. S., Conn, P. B., Hooten, M. B., Ray, J. C., & Pond, B. A. (2013). Spatial occupancy models for large data sets. *Ecology*, 94(4), 801–808.
- Johnson, D. S., Laake, J., & Ver Hoef, J. (2010). A model-based approach for making ecological inference from distance sampling data. *Biometrics*, 66(1), 310–318.

- Johnson, D. S., Laake, J., & VerHoef, J. (2014). *DSpat: Spatial modelling for distance sampling data*. R package version 0.1.6.
- Jolly, G. M. (1965). Explicit estimates from capture-recapture data with both death and immigration-stochastic model. *Biometrika*, 52(1–2), 225–247.
- Jones, H. E., Hickman, M., Welton, N. J., De Angelis, D., Harris, R. J., & Ades, A. E. (2014). Recapture or precapture? Fallibility of standard capture-recapture methods in the presence of referrals between sources. *American Journal of Epidemiology*, 179(11), 1383–1393.
- Jordan, M. J., Barrett, R. H., & Purcell, K. L. (2011). Camera trapping estimates of density and survival of fishers (*Martes pennanti*). *Wildlife Biology*, 17(3), 266–276.
- Junge, C. O. (1963). A quantitative evaluation of the bias in population estimates based on selective samples. In *North Atlantic Fish Marking Symposium, I. C. N. A. F. Special publication No. 4* (pp. 26–28).
- Kaiser, L. (1983). Unbiased estimation in line-intercept sampling. *Biometrics*, 39(4), 965–976.
- Kamp, J., Oppel, S., Heldbjerg, H., Nyegaard, T., & Donald, P. F. (2016). Unstructured citizen science data fail to detect long-term population declines of common birds in Denmark. *Diversity and Distributions*, 22(10), 1024–1035.
- Kane, M. D., Morin, D. J., & Kelly, M. J. (2015). Potential for camera-traps and spatial mark-resight models to improve monitoring of the critically endangered West African lion (*Panthera leo*). *Biodiversity Conservation*, 24(14), 3527–3541.
- Kansanen, K., Packalen, P., Maltamo, M., & Mehtätalo, L. (2021). Horvitz-Thompson-like estimation with distance-based detection probabilities for circular plot sampling of forests. *Biometrics*, 77(2), 715–728.
- Karanth, K. U., Gopalaswamy, A. M., Kumar, N. S., Vaidyanathan, S., Nichols, J. D., & MacKenzie, D. I. (2011). Monitoring carnivore populations at the landscape scale: Occupancy modelling of tigers from sign surveys. *Journal of Applied Ecology*, 48(4), 1048–1056.
- Karanth, K. U., Nichols, J. D., & Kumar, N. S. (2004). Photographic sampling of elusive mammals in tropical forests. In W. L. Thompson (Ed.), *Sampling rare or elusive species: Concepts, designs, and techniques for estimating population parameters* (pp. 229–247). Washington, D.C.: Island Press.
- Karanth, K. U., Nichols, J. D., Kumar, N. S., & Hines, J. E. (2006). Assessing tiger population dynamics using photographic capture-recapture sampling. *Ecology*, 87(11), 2925–2937.
- Karavarsamis, N., & Huggins, R. M. (2019). Two-stage approaches to the analysis of occupancy data II. The heterogeneous model and conditional likelihood. *Computational Statistics and Data Analysis*, 133, 195–207.
- Karavarsamis, N., & Huggins, R. M. (2020). Two-stage approaches to the analysis of occupancy data I: The homogeneous case (analysis of occupancy data). *Communications in Statistics. Theory and Methods*, 49(19), 4751–4761.
- Karenyi, N., Nel, R., Altwegg, R., & Sink, K. (2016). Incorporating species detectability into conservation targets based on the species-area relationship. *Diversity and Distributions*, 22(7–8), 758–769.
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430), 773–795.
- Kass, R. E., & Wasserman, L. (1996). The selection of prior distributions by formal rules. *Journal of the American Statistical Association*, 91(435), 1343–1370.
- Katsanevakis, S. (2007). Density surface modelling with line transect sampling as a tool for abundance estimation of marine benthic species: The *Pinna nobilis* example in a marine lake. *Marine Biology*, 152(1), 77–85.
- Katsanevakis, S., Stelzenmüller V., & et al. (2011). Ecosystem-based marine spatial management: Review of concepts, policies, tools, and critical issues. *Ocean Coast Management*, 54(11), 807–820.
- Katsanevakis, S., Weber, A., & et al. (2012). Monitoring marine populations and communities: Methods dealing with imperfect detectability. *Aquatic Biology*, 16(1), 31–52.

- Keeping, D., Burger, J. H., & et al. (2018). Can trackers count free-ranging wildlife as effectively and efficiently as conventional aerial survey and distance sampling—Implications for citizen science in the Kalahari, Botswana. *Biological Conservation*, 223, 156–169.
- Keesing, F., Belden, L., & et al. (2010). Impacts of biodiversity on the emergence and transmission of infectious diseases. *Nature*, 468(7324), 647–652.
- Kellner, K. (2015). jagsUI: A wrapper around rjags to streamline JAGS analyses. R package.
- Kellner, K. F., Parsons, A. W., Kays, R., Millspaugh, J. J., & Rota, C. T. (2022). A two-species occupancy model with a continuous-time detection process reveals spatial and temporal interactions. *Journal of Agricultural, Biological, and Environmental Statistics*, 27(2), 321–338.
- Kellner, F. K., & Swihart, R. K. (2014). Accounting for imperfect detection in ecology: A quantitative review. *PLoS One*, 9(10), 1–8, e111436.
- Kendall, M. G., & Moran, P. A. P. (1963). *Geometrical probability*. London: Griffin.
- Kendall, W. L. (1999). Robustness of closed capture-recapture models to violations of the closure assumption. *Ecology*, 80(8), 2517–2525.
- Kendall, W. L. (2009). One size does not fit all: Adapting mark-recapture and occupancy models for state uncertainty. In D. L. Thomson, E. G. Cooch, & M. J. Conroy (Eds.), *Modeling demographic processes in marked populations* (pp. 765–780). New York: Springer.
- Kendall, W. L., Hines, J. E., Nichols, J. D., & Grant, E. V. H. (2013). Relaxing the closure assumption in occupancy models: Staggered arrival and departure times. *Ecology*, 94(3), 610–617.
- Kendall, W. L., & White, G. C. (2009). A cautionary note on substituting spatial subunits for repeated temporal sampling in studies of site occupancy. *Journal of Applied Ecology*, 46(6), 1182–1188.
- Kéry, M. (2004). Extinction rate estimates for plant populations in revisitation studies: The importance of detectability. *Conservation Biology*, 18(2), 570–574.
- Kéry, M. (2010). *Introduction to winbugs for Ecologists—a Bayesian approach to regression, ANOVA, mixed models and related analyses*. Burlington: Academic Press.
- Kéry, M. (2018). Identifiability in N-mixture models: A large-scale screening test with bird data. *Ecology*, 99(2), 281–288.
- Kéry, M., Dorazio, R., Soldaat, L., van Strien, A., Zuiderwijk, A., & Royle, J. A. (2009). Trend estimation in populations with imperfect detection. *Journal of Applied Ecology*, 46(6), 1163–1172.
- Kéry, M., Gardner, B., & Monnerat, C. (2010). Predicting species distributions from checklist data using site-occupancy models. *Journal of Biogeography*, 37(10), 1851–1862.
- Kéry, M., Guillera-Arroita, G., & Lahoz-Montfort, J. J. (2013). Analysing and mapping species dynamics using occupancy models. *Journal of Biogeography*, 40(8), 1463–1474.
- Kéry, M., & Plattner, M. (2007). Species richness estimation and determinants of species detectability in butterfly monitoring programmes. *Ecological Entomology*, 32(1), 53–61.
- Kéry, M., & Royle, J. A. (2008a). Hierarchical Bayes estimation of species richness and occupancy in spatially replicated surveys. *Journal of Applied Ecology*, 45(2), 589–598.
- Kéry, M., & Royle, J. A. (2009). Inference about species richness and community structure using species-specific occupancy models in the National Swiss Breeding Bird Survey MHB. In D. L. Thomson, E. G. Cooch, & M. J. Conroy (Eds.), *Modeling demographic processes in marked populations* (pp. 639–656). Boston, MA: Springer. Environmental and Ecological Statistics.
- Kéry, M., & Royle, J. A. (2010). Hierarchical modelling and estimation of abundance and population trends in metapopulation designs. *Journal of Animal Ecology*, 79(2), 453–461.
- Kéry, M., & Royle, J. A. (2016). *Applied hierarchical modeling in ecology: Analysis of distribution, abundance and species richness in R and BUGS*. Prelude and static models (Vol. 1). London, UK: Academic Press.
- Kéry, M., & Royle, J. A. (2020). *Applied hierarchical modeling in ecology: Analysis of distribution, abundance and species richness in R and BUGS*. Dynamic and advanced models (Vol. 2). Cambridge, Massachusetts: Academic Press.
- Kéry, M., Royle, J. A., Plattner, M., & Dorazio, R. M. (2009). Species richness and occupancy estimation in communities subject to temporary emigration. *Ecology*, 90(5), 1279–1290.

- Kéry, M., Royle, J. A., & Schmid, H. (2005). Modeling avian abundance from replicated counts using binomial mixture models. *Ecological Applications*, 15(4), 1450–1461.
- Kéry, M., & Schaub, M. (2012). *Bayesian population analysis using WinBUGS: A hierarchical perspective*. New York: Academic press, Elsevier.
- Kéry, M., & Schmid, H. (2004). Monitoring programs need to take into account imperfect species detectability. *Basic and Applied Ecology*, 5(1), 65–73.
- Kéry, M., & Schmid, H. (2006). Estimating species richness: Calibrating a large avian monitoring programme. *Journal of Applied Ecology*, 43(1), 101–110.
- Kéry, M., Spillmann, J. H., Truong, C., & Holderegger, R. (2006). How biased are estimates of extinction probability in revisit studies? *Journal of Ecology*, 94(5), 980–986.
- Khan, M. N. I., Hijbeek, R., & et al. (2016). An evaluation of the plant density estimator the point-centred quarter method (PCQM) using Monte Carlo simulation. *PLoS One*, 11(6), e0157985.
- King, R., Bird, S. M., Brooks, S. P., Hutchinson, S. J., & Hay, G. (2005). Prior information in behavioral capture-recapture methods: Demographic influences on drug injectors' propensity to be listed in data sources and their drug-related mortality. *American Journal of Epidemiology*, 162(7), 694–703.
- King, R., Bird, S. M., Overstall, A. M., Hay, G., & Hutchinson, S. J. (2014). Estimating prevalence of injecting drug users and associated heroin-related death rates in England by using regional data and incorporating prior information. *Journal of the Royal Statistical Society Series A*, 177(1), 209–236.
- King, R., & Brooks, S. P. (2001). On the Bayesian analysis of population sizes. *Biometrika*, 88(2), 317–336.
- King, R., & Brooks, S. P. (2008). On the Bayesian estimation of a closed population size in the presence of heterogeneity and model uncertainty. *Biometrics*, 64(3), 816–824.
- King, R., McClintock, B. T., Kidney, D., & Borchers, D. (2016). Capture-recapture abundance estimation using a semi-complete data likelihood approach. *Annals of Applied Statistics*, 10(1), 264–285.
- King, R., & McCrea, R. S. (2019). Capture-recapture methods and models: Estimating population size. In *Handbook of Statistics* (Vol. 40, pp. 33–83). Elsevier.
- King, R., Morgan, B. J. T., Gimenez, O., & Brooks, S. P. (2010). *Bayesian analysis for population ecology*. Chapman and Hall/CRC.
- Kissling, M., & Garton, E. (2006). Estimating detection probability and density from point-count Surveys: Combination of distance and double-observer sampling. *Auk*, 123(3), 735–752.
- Kitzes, J., Harte, J., & Freckleton, R. (2014). Beyond the species-area relationship: Improving macroecological extinction estimates. *Methods in Ecology and Evolution*, 5(1), 1–8.
- Kleinn, C., & Vilčko, F. (2006a). A new empirical approach for estimation in k-tree sampling. *Forest Ecology and Management*, 237(1), 522–533.
- Kleinn, C., & Vilčko, F. (2006b). Design-unbiased estimation for point-to-tree distance sampling. *Canadian Journal of Forest Research*, 36(6), 1407–1414.
- Knape, J., Arlt, D., & et al. (2018) Sensitivity of binomial N-mixture models to overdispersion: The importance of assessing model fit. *Methods of Ecological Evolution*, 9(10), 2102–2114.
- Knape, J., & Korner-Nievergelt, F. (2016). On assumptions behind estimates of abundance from counts at multiple sites. *Methods in Ecology and Evolution*, 7(3), 206–209.
- Kohn, M. H., York, E., Kamradt, D. A., Haught, G., Sauvajot, R., & Wayne, R. K. (1999). Estimating population size by genotyping faeces. *Proceedings of the Royal Society of London, Series B*, 266, 657–663.
- Koneff, M. D., Royle, J. A., Otto, M. C., Wortham, J. S., & Bidwell, J. K. (2008). A double-observer method to estimate detection rate during aerial waterfowl surveys. *Journal of Wildlife Management*, 72(7), 1641–1649.
- Köse, T., Orman, M., Ikiz, F., Baksh, M. F., Gallagher, J., & Böhning, D. (2014). *Statistics in Medicine*, 33(24), 4237–4249. Correction: *Statistics in Medicine*, 36(9), 1519–1520.
- Kroll, A. J., Garcia, T. S., & et al. (2015). Evaluating multi-level models to test occupancy state responses of plethodontid salamanders. *PLoS One*, 10(11), e0142903.

- Laake, J. L., & Borchers, D. L. (2004). Methods for incomplete detection at distance zero. In S. T. Buckland, D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers, & L. Thomas (Eds.), *Advanced distance sampling* (pp. 108–189). Oxford: Oxford University Press.
- Laake, J., Borchers, D., Thomas, L., Miller, D., & Bishop, J. (2013). mrds: Mark—Recapture Distance Sampling (mrds). **R** package version 2.1.4. <http://CRAN.R-project.org/package=mrds>.
- Laake, J. L., Collier, B. A., Morrison, M. L., & Wilkins, R. N. (2011). Point-based mark-recapture distance sampling. *Journal of Agricultural, Biological, and Environmental Statistics*, 16(3), 389–408.
- Laake, J. L., Dawson, M. J., & Hone, J. (2008). Visibility bias in aerial survey: Mark—recapture, line-transect or both? *Wildlife Research*, 35(4), 299–309.
- Laake, J., Guenzel, R. J., Bengtson, J. L., Boveng, P., Cameron, M., & Hanson, M. B. (2008). Coping with variation in aerial survey protocol for line-transect sampling. *Wildlife Research*, 35(4), 289–299.
- Lachish, S., Gopalaswamy, A. M., Arjun, M., Knowles, S. C. L., & Sheldon, B. C. (2012). Site-occupancy modelling as a novel framework for assessing test sensitivity and estimating wildlife disease prevalence from imperfect diagnostic tests. *Methods in Ecology and Evolution*, 3(2), 339–348.
- Lachish, S., & Murray, K. A. (2018). The certainty of uncertainty: Potential sources of bias and imprecision in disease ecology studies. *Frontiers in Veterinary Science*, 5, 90.
- Lahoz-Monfort, J. J., Guillera-Arroita, G., & Tingley, R. (2016). Statistical approaches to account for false positive errors in environmental DNA samples. *Molecular Ecology Resource*, 16(3), 673–685.
- Lahoz-Monfort, J. J., Guillera-Arroita, G., & Wintle, B. A. (2014). Imperfect detection impacts the performance of species distribution models. *Global Ecology and Biogeography*, 23(4), 504–515.
- Lampa, S., Henle, K., Klenke, R., Hoehn, M., & Gruber, B. (2013). How to overcome genotyping errors in non-invasive genetic mark-recapture population size estimation: A review of available methods illustrated by a case study. *Journal of Wildlife Management*, 77(8), 1490–1511.
- Lancaster, T., & Imbens, G. (1996). Case-control studies with contaminated controls. *Journal of Econometrics*, 71(1), 145–160.
- Landis, J. R., & Koch, G. C. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174.
- Langrock, R., Borchers, D. L., & Skaug, H. J. (2013). Markov modulated nonhomogeneous poisson processes for modeling detections in surveys of marine mammal abundance. *Journal of the American Statistical Association*, 108(503), 840–851.
- Lany, N. K., Zarnetske, P. L., Gouhier, T. C., & Mengé, B. A. (2017). Incorporating context dependency of species interactions in species distribution models. *Integrative and Comparative Biology*, 57(1), 159–167.
- Laplanche, C. (2010). A hierarchical model to estimate fish abundance in alpine streams by using removal sampling data from multiple locations. *Biometrical Journal*, 52(2), 209–221.
- Latimer, A. M., Banerjee, S., Sang, Jr, H., Mosher, E. S., & Silander, Jr, J. A. (2009). Hierarchical models facilitate spatial analysis of large data sets: A case study on invasive plant species in the northeastern United States. *Ecology Letters*, 12(2), 144–154.
- Latimer, A. M., Wu, S., Gelfand, A. E., & Silander, Jr, J. A. (2006). Building statistical models to analyze species distributions. *Ecological Applications*, 16(1), 33–50.
- Lauritzen, S. L. (1996). *Graphical models* (Vol. 17). Clarendon Press.
- Lee, A. H., Wang, K., Scott, J. A., Yau, K. K. W., & McLachlan, G. J. (2006). Multi-level zero-inflated Poisson regression modelling of correlated count data with excess zeros. *Statistical Methods in Medical Research*, 15(1), 47–61.
- Lee, A. J. (2002). Effect of list errors on the estimation of population size. *Biometrics*, 58, 185–191.
- Lee, A. J., & Seber, G. A. F. (2001). Residuals. *Encyclopedia of Environmetrics*, 1770–1775. In A. El-Shaarawi (Ed.). Wiley.

- Lee, A. J., Seber, G. A. F., Holden, J. K., & Huakau, J. T. (2001). Capture-recapture, epidemiology, and list mismatches: Several Lists. *Biometrics*, 7(3) 707–713.
- Lee, D. C., & Marsden, S. J. (2008). Adjusting count period strategies to improve the accuracy of forest bird abundance estimates from point transect distance sampling surveys. *Ibis*, 150(2), 315–325.
- Lee, S.-M. (1996). Estimating population size for capture-recapture data when capture probabilities vary by time, behavior and individual animal. *Communications in Statistics—Simulation and Computation*, 25, 431–457.
- Lee, S.-M., & Chao, A. (1994). Estimating population size via sample coverage for closed capture-recapture models. *Biometrics*, 50(1), 88–97.
- Lee, S.-M., & Chen, C. W. S. (1998). Bayesian inference of population size for behavioral response models. *Statistica Sinica*, 8(4), 1233–1247.
- Lee, S.-M., Hwang, W.-H., & de Dieu Tapsoba, J. (2016). Estimation in closed capture-recapture models when covariates are missing at random. *Biometrics*, 72(4), 1294–1304.
- Lee, S.-M., Hwang, W.-H., & Huang, L.-H. (2003). Bayes estimation of population size from capture-recapture models with time variation and behavior response. *Statistica Sinica*, 13(2), 477–494.
- Lee, T. E. (2014). A simple numerical tool to infer whether a species is extinct. *Methods in Ecology and Evolution*, 5(8), 791–796.
- Lele, S. R. (2009). A new method for estimation of resource selection probability function. *Journal of Wildlife Management*, 73(1), 122–127.
- Lele, S. R., & Allen, K. L. (2006). On using expert opinion in ecological analyses: A frequentist approach. *Environmetrics*, 17(7), 683–704.
- Lele, S. R., & Keim, J. L. (2006). Weighted distributions and estimation of resource selection probability functions. *Ecology*, 87(12), 3021–3028.
- Lele, S. R., Merrill, E. H., Keim, J., & Boyce, M. S. (2013). Selection, use, choice and occupancy: Clarifying concepts in resource selection studies. *Journal of Animal Ecology*, 82(6), 1183–1191.
- Lele, S. R., Moreno, M., & Bayne, E. (2012). Dealing with detection error in site occupancy surveys: What can we do with a single survey? *Journal of Plant Ecology*, 5(1), 22–31.
- Lenoir, J., Bertrand, R., & et al. (2020). Species better track climate warming in the oceans than on land. *Nature Ecology and Evolution*, *Nature*, 4(8), 1044–1059.
- Leslie, P. H., & Davis, D. H. S. (1939). An attempt to determine the absolute of rats on a given area. *Journal of Animal Ecology*, 8(1), 94–1011.
- Lessard, V., Reed, D. D., & Monkevich, N. (1994). Comparing n-tree distance sampling with point and plot sampling in northern Michigan forest types. *Northern Journal of Applied Forestry*, 11(1), 12–16.
- Lessard, V. C., Drummer, T. D., & Reed, D. D. (2002). Precision of density estimates from fixed-radius plots compared to n-tree distance sampling. *Forest Science*, 48(1), 1–6.
- Leujak, W., & Ormond, R. F. G. (2007). Comparative accuracy and efficiency of six coral community survey methods. *Journal of Experimental Marine Biology and Ecology*, 351(1), 168–187.
- Levin, S. A. (Ed.) (2013). *Encyclopedia of biodiversity* (2nd ed.). Elsevier.
- Lewandowski, E., & Specht, H. (2015). Influence of volunteer and project characteristics on data quality of biological surveys. *Conservation Biology*, 29(3), 713–723.
- Lewins, W. A., & Joanes, D. N. (1984). Bayesian estimation of the number of species. *Biometrics*, 40(2), 323–328.
- Lewis, J. S., Bailey, L. L., VandeWoude, S., & Crooks, K. R. (2015). Interspecific interactions between wild felids vary across scales and levels of urbanization. *Ecology and Evolution*, 5(24), 5946–5961.
- Liang, K.-Y., & Zeger, S. L. (1995). Inference based on estimating functions in the presence of nuisance parameters (with discussion). *Statistical Science*, 10(2), 158–199.
- Lin, D. Y., & Yip, S. F. P. (1999). Parametric regression models for continuous time removal and recapture studies. *Journal of the Royal Statistical Society, Series B*, 61(2), 401–411.

- Linden, A., Mantyniemi, S., & Fortin, M. (2011). *Ecology*, 92(7), 1414–1421.
- Linden, D. W., Fuller, A. K., Royle, J. A., & Hare, M. P. (2017). Examining the occupancy-density relationship for a low-density carnivore. *Journal of Applied Ecology*, 54(6), 2043–2052.
- Linden, D. W., Sirén, A. P., & Pekins, P. J. (2018). Integrating telemetry data into spatial capture–recapture modifies inferences on multi-scale resource selection. *Ecosphere*, 9(4), e02203.
- Lindsay, B. G., & Roeder, K. (1987). A unified treatment of integer parameter models. *Journal of the American Statistical Association*, 82(399), 758–764.
- Link, W. A. (2003). Nonidentifiability of population size from capture-recapture data with heterogeneous detection probabilities. *Biometrics*, 59(4), 1123–1130.
- Link, W. A. (2013). A cautionary note on the discrete uniform prior for the binomial n . *Ecology*, 94(10), 2173–2179.
- Link, W. A. (2014). A cautionary note on the discrete uniform prior for the binomial n : Reply. *Ecology*, 95(9), 2677–2679.
- Link, W. A., & Barker, R. J. (1994). Density estimation using the trapping web design: A geometric analysis. *Biometrics*, 50(3), 733–745.
- Link, W. A., & Barker, R. J. (2006). Model weights and the foundations of multimodel inference. *Ecology*, 87(10), 2626–2635.
- Link, W. A., & Barker, R. J. (2010). *Bayesian inference: With ecological applications*. San Diego, USA: Academic Press.
- Link, W. A., & Sauer, J. (1999). Controlling for varying effort in count surveys-analysis of Christmas Bird Count data. *Journal of Agricultural, Biological and Environmental Statistics*, 4(2) 116–125.
- Link, W. A., Yoshizaki, J., Bailey, L. L., & Pollock, K. H. (2010). Uncovering a latent multinomial: Analysis of mark-recapture data with misidentification. *Biometrics*, 66(1), 178–185.
- Liow, L.H. (2013). Simultaneous estimation of occupancy and detection probabilities: An illustration using Cincinnati brachiopods. *Paleobiology*, 39, 193–213.
- Liu, L., Guo, Z., & Duan, X. (2016). Population size estimation with missing nonignorable. *Acta Mathematicae Applicatae Sinica, English Series*, 32(3), 659–668.
- Liu, L., Liu, Y., Li, P., & Qin, J. (2018). Full likelihood inference for abundance from continuous time capture–recapture data. *Journal of the Royal Statistical Society, Series B*, 80(5), 995–1014.
- Liu, L., & Yip, P. S. F. (2002). Estimating population size in proportional trapping-removal models. *Statistica Sinica*, 13(1), 243–254.
- Liu, L., Yip, P. S. F., & Watson, R. K. (2003). Removal process estimation of population size for a population with known sex ratio. *Environmental and Ecological Statistics*, 10(2), 281–292.
- Liu, S., Baret, F., & et al. (2017a). A method to estimate plant density and plant spacing heterogeneity: Application to wheat crops. *Plant Methods*, 13(1), 38–49.
- Liu, Y., Li, P., & Qin, J. (2017b). Maximum empirical likelihood estimation for abundance in a closed population from capture–recapture data. *Biometrika*, 104(3), 527–543.
- Liu, Y., Zhu, L., Liu, G., & Li, H. (2020). Abundance estimation with a categorical covariate subject to missing in continuous-time capture-recapture studies. *Communications in Statistics. Theory and Methods*, 49(20), 4919–4928.
- Liu, Z., Wang, X., Teng, L., Cui, D., & Li, X. (2008). Estimating seasonal density of blue sheep (*Pseudois nayaur*) in the Helan Mountain region using distance sampling methods. *Ecological Research*, 23(2), 393–400.
- Loehle, C. (2010). Population estimation using partial double sampling. (2010). *Forest Science*, 56(4), 417–420.
- Lomolino, M. V. (2000). Ecology's most general, yet protean pattern: The species-area relationship. *Journal of Biogeography*, 27(1), 17–26.
- Long, J. A., & Nelson, T. A. (2012). Time geography and wildlife home range delineation. *Journal of Wildlife Management*, 76(2), 407–413.
- Loonam, K. E., Lukacs, P. M., Ausband, D. E., Mitchell, M. S., & Robinson, H. S. (2021). Assessing the robustness of time-to-event models for estimating unmarked wildlife abundance using remote cameras. *Ecological Applications*, 31(6), e02388.

- Loya, Y. (1972). Community structure and species diversity of hermatypic corals at Eliat, RED Sea. *Marine biology*, 13(2), 100–123.
- Lubow, B. C., & Ransom, J. I. (2009). Validating aerial photographic mark-recapture for naturally marked feral horses. *Journal of Wildlife Management*, 73(8), 1420–1429.
- Lucas, H. A., & Seber, G. A. F. (1977). Estimating coverage and particle density using the line intercept method. *Biometrika*, 64(3), 618–622.
- Luikart, G., Ryman, N., Tallmon, D. A., Schwartz, M. K., & Allendorf, F. W. (2010). Estimation of census and effective population sizes: The increasing usefulness of DNA-based approaches. *Conservation Genetics*, 11(2), 355–373.
- Lukacs, P. M. (2002). WebSim: Simulation software to assist in trapping web design. *Wildlife Society Bulletin*, 30(4), 1259–1261.
- Lukacs, P. M., & Burnham, K. P. (2005). Review of capture–recapture methods applicable to noninvasive genetic sampling. *Molecular Ecology*, 30(15), 3909–3919.
- Lukacs, P. M., Franklin, A. B., & Anderson, D. R. (2004). Passive approaches to detection in distance sampling. In S. T. Buckland, D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers, & L. Thomas (Eds.), *Advanced distance sampling* (pp. 260–280). Oxford, UK: Oxford University Press.
- Lunn, D., Jackson, C., Best, N., Thomas, A., & Spiegelhalter, D. (2013). *The BUGS book—a practical introduction to Bayesian analysis*. CRC Press, Chapman and Hall.
- Lunn, D., Spiegelhalter, D., Thomas, A., & Best, N. (2009). The BUGS project: Evolution, critique and future directions. *Statistics in Medicine*, 28(25), 3049–3067.
- Lunn, D., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS—a Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, 10(4), 325–337.
- Lurdes, B. S., Alves, M., Rui, B. E., & Silva, L. (2017). Comparison of T -square, point centered quarter, and N -tree sampling methods in pittosporum undulatum invaded woodlands. *International Journal of Forestry Research*, 2017. Article ID2818132, 13pp.
- Lynch, T. B. (2012). A mirage boundary correction method for distance sampling. *Canadian Journal of Forest Research*, 42(2), 272–278.
- Lynch, T. B. (2017). Optimal sample size and plot size or point sampling factor based on cost-plus-loss using the Fairfield Smith relationship for plot size. *Forestry*, 90(5), 697–709.
- Lynch, T. B., & Gove, J. H. (2014). The unbiasedness of a generalized mirage boundary correction method for Monte Carlo integration estimators of volume. *Canadian Journal of Forest Research*, 44(7), 810–819.
- Lynch, T. B., & Rusydi, R. (1999). Distance sampling for forest inventory in Indonesian teak plantations. *Forest Ecology and Management*, 113(2), 215–221.
- MacKenzie, D. I., & Bailey, L. L. (2004). Assessing the fit of site-occupancy models. *Journal of Agricultural Biological and Environmental Statistics*, 9(3), 300–318.
- MacKenzie, D. I., Bailey, L. L., Hines, J. E., & Nichols, J. D. (2011). An integrated model of habitat and species occurrence dynamics. *Methods in Ecology and Evolution*, 2(6), 612–622.
- MacKenzie, D. I., Bailey, L. L., & Nichols, J. D. (2004). Investigating species co-occurrence patterns when species are detected imperfectly. *Journal of Animal Ecology*, 73(3), 546–555.
- MacKenzie, D., & Hines, J. (2015). RPresence: R Interface for Program PRESENCE. R package version 1.1.1. 2. <https://cran.r-project.org/package=detect>.
- MacKenzie, D. I., & Nichols, J. D. (2004). Occupancy as a surrogate for abundance estimation. *Animal Biodiversity and Conservation*, 27(1), 461–467.
- MacKenzie, D. I., Nichols, J. D., Hines, J. E., Knutson, M. G., & Franklin, A. B. (2003). Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. *Ecology*, 84(8), 2200–2207.
- MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Royle, J. A., & Langtimm, C. A. (2002). Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, 83(8), 2248–2255.
- MacKenzie, D. I., Nichols, J. D., Royle, J. A., Pollock, K. H., Bailey, L. L., & Hines, J. E. (2006). *Occupancy estimation and modeling: Inferring patterns and dynamics of species occurrence* (1st ed.). Academic Press.

- MacKenzie, D. I., Nichols, J. D., Royle, J. A., Pollock, K. H., Bailey, L. L., & Hines, J. E. (2018). *Occupancy estimation and modeling: Inferring patterns and dynamics of species occurrence* (2nd ed.). Elsevier, Academic Press.
- MacKenzie, D. I., Nichols, J. D., Seamans, M. E., & Gutiérrez, R. J. (2009). Modeling species occurrence dynamics with multiple states and imperfect detection. *Ecology*, 90(3), 823–835.
- MacKenzie, D. I., & Royle, J. A. (2005). Designing efficient occupancy studies: General advice and tips on allocation of survey effort. *Journal of Applied Ecology*, 42(6), 1105–1114.
- MacKenzie, D. I., Royle, J. A., Brown, J. A., & Nichols, J. D. (2004). Occupancy estimation and modeling for rare and elusive populations. In W. L. Thompson (Ed.), *Sampling rare or elusive species* (pp. 149–171). Washington, DC: Island Press.
- Madigan, D., & York, J. (1997). Bayesian methods for estimation of the size of a closed population. *Biometrika*, 84(1), 19–31.
- Magnussen, S. (2014). Robust fixed-count density estimation with virtual plots. *Canadian Journal of Forest Research*, 44(4), 377–382.
- Magnussen, S., Fehrmann, L., & Platt, W. J. (2012). An adaptive composite density estimator for k-tree sampling. *European Journal of Forest Research*, 131(2), 307–320.
- Magnussen, S., Kleinn, C., & Picard, N. (2008). Two new density estimators for distance sampling. *European Journal of Forest Research*, 127(3), 213–224.
- Magurran, A. E., & McGill, B. J. (Eds.). (2011). *Biological diversity: Frontiers in measurement and assessment*. Oxford: Oxford University Press.
- Mair, L., & Ruete, A. (2016). Explaining spatial variation in the recording effort of citizen science data across multiple taxa. *PLoS One*, 11(1), e0147796.
- Manly, B. F. J. (1997). *Randomization, bootstrap and Monte Carlo methods in biology* (2nd ed.). London: Chapman & Hall.
- Manly, B. F. J. (2004). Two-phase adaptive stratified sampling. In W. L. Thompson (Ed.), *Sampling rare or elusive species: Concepts, designs, and techniques for estimating population parameters* (pp. 123–133). Washington, D.C.: Island Press.
- Manly, B. F. J., McDonald, L. L., & Garner, G. W. (1996). Maximum likelihood estimation for the double-count method with Independent observers. *Journal of Agricultural, Biological, and Environmental Statistics*, 1(2), 170–189.
- Manly, B. F. J., McDonald, L. L., Thomas, D. L., McDonald, T. L., & Erickson, W. P. (2002). *Resource selection by animals: Statistical analysis and design for field studies* (2nd ed.). Boston, Massachusetts, USA: Kluwer Press.
- Manning, J. A., & Goldberg, C. S. (2010). Estimating population size using capture–recapture encounter histories created from point-coordinate locations of animals. *Methods in Ecology and Evolution*, 1(4), 389–397.
- Manrique-Vallier, D. (2016). Bayesian population size estimation using Dirichlet process mixtures. *Biometrics*, 72(4), 1246–1254.
- Mäntyniemi, S., Romakkaniemi, A., & Arjas, E. (2005). Bayesian removal estimation of a population size under unequal catchability. *Canadian Journal of Fisheries and Aquatic Sciences*, 62(2), 291–300.
- Mao, X. C., & You, N. (2009). On comparison of mixture models for closed population capture–recapture studies. *Biometrics*, 65(2), 547–553.
- Marquardt, T., Temesgen, H., & Anderson, P. D. (2010). Accuracy and suitability of selected sampling methods within conifer dominated riparian zones. *Forest Ecology and Management*, 260(3), 313–320.
- Marques, F. F. C., & Buckland, S. T. (2003). Incorporating covariates into standard line transect analyses. *Biometrics*, 59(4), 924–935.
- Marques, F. F. C., & Buckland, S. T. (2004). Covariate models for the detection function. In S. T. Buckland, D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers, & L. Thomas (Eds.), *Advanced distance sampling* (pp. 31–47). Oxford: Oxford University Press.
- Marques, F. F. C., Buckland, S. T., & et al. (2001). Estimating deer abundance from line transect surveys of dung: Sika deer in southern Scotland. *Journal of Applied Ecology*, 38(2), 349–363.

- Marques, T. A. (2004). Predicting and correcting bias caused by measurement error in line transect sampling using multiplicative error models. *Biometrics*, 60(3), 757–763.
- Marques, T. A., Buckland, S. T., Borchers, D. L., Tosh, D., & McDonald, R. A. (2010). Point transect sampling along linear features. *Biometrics*, 66(4), 1247–1255.
- Marques, T. A., Thomas, L., Fancy, S. G., & Buckland, S. T. (2007). Improving estimates of bird density using multiple covariate distance sampling. *Auk*, 124(4), 1229–1243.
- Marques, T. A., Thomas, L., Martin, S. W., & et al. (2012). Spatially explicit capture-recapture methods to estimate minke whale abundance from data collected at bottom mounted hydrophones. *Journal of Ornithology*, 152(Sup2), 445–455.
- Marques, T. A., Thomas, L., Martin, S. W., & et al. (2013). Estimating animal population density using passive acoustics. *Biological Reviews*, 88(2), 287–309.
- Marques, T. A., Thomas, L., Munger, L., Wiggins, S., & Hildebrand, J. A. (2011). Estimating North Pacific right whale (*Eubalaena japonica*) density using passive acoustic cue counting. *Endangered Species Research*, 13(3), 163–172.
- Marques, T. A., Thomas, L., Ward, J., DiMarzio, N., & Tyack, P. L. (2009). Estimating cetacean population density using fixed passive acoustic sensors: An example with beaked whales. *Journal of the Acoustical Society of America*, 125(4), 1982–1994.
- Marsh, H., & Sinclair, D. F. (1989). Correcting for visibility bias in transect aerial surveys of aquatic fauna. *Journal of Wildlife Management*, 53(4), 1017–1024.
- Marten, G. G. (1970). A regression method for mark-recapture estimates with unequal catchability. *Ecology*, 51(2), 150–152.
- Martin, J., Chamaillé-Jammes, S., & et al. (2010). Simultaneous modeling of habitat suitability, occupancy, and relative abundance: African elephants in Zimbabwe. *Ecological Applications*, 20(4), 1173–1182.
- Martin-Schwarze, A., Niemi, J., & Dixon, P. (2017). Assessing the impacts of time-to-detection distribution assumptions on detection probability estimation. *Journal of Agricultural, Biological, and Environmental Statistics*, 22(4), 465–480.
- Martinez-Minaya, J., Cameletti, M., Conesa, D., & Pennino, M. G. (2018). Species distribution modeling: A statistical review with focus in spatio-temporal issues. *Stochastic Environmental Research and Risk Assessment*, 32(11), 3227–3244.
- Mason, N. W. H., Holdaway, R. J., & Richardson, S. J. (2018). Incorporating measurement error in testing for changes in biodiversity. *Methods in Ecology and Evolution*, 9(5), 1296–1307.
- Matechou, E., Dennis, E. B., Freeman, S. N., & Brereton, T. (2014). Monitoring abundance and phenology in (multivoltine) butterfly species: A novel mixture model. *Journal of Applied Ecology*, 51(3), 766–775.
- Matechou, E., McCrea, R. S., Morgan, B. J. T., Nash, D. J., & Griffiths, R. A. (2016). Open models for removal data. *Annals of Applied Statistics*, 10(3), 1572–1589.
- Matias, M. G., Gravel, D., & et al. (2014). Estimates of species extinctions from species-area relationships strongly depend on ecological context. *Ecography*, 37(5), 431–442.
- Mattsson, B. J., Zipkin, E. F., Gardner, B., Blank, P. J., Sauer, J. R., & Royle, J. A. (2013). Explaining local-scale species distributions: Relative contributions of spatial autocorrelation and landscape heterogeneity for an avian assemblage. *PLoS One*, 8(2), e55097.
- Mayfield, M. M., & D. B. Stouffer. (2017). Higher-order interactions capture unexplained complexity in diverse communities. *Nature Ecology and Evolution*, 1(3), 62–62.
- Mazerolle, M. J. (2015). Estimating detectability and biological parameters of interest with the use of the R Environment. *Journal of Herpetology*, 49(4), 541–559.
- Mazerolle, M. J., Bailey, L. L., Kendall, W. L., Royle, J. A., Converse, S. J., & Nichols, J. D. (2007). Making great leaps forward: Accounting for detectability in herpetological field studies. *Journal of Herpetology*, 41(4), 672–689.
- McCallum, D. A. (2005). A conceptual guide to detection probability for point counts and other count-based survey methods. In *USDA Forest Service General Technical Report PSW-GTR-191*.

- McClintock, B. T. (2012). Mark-resight models. In E. G. Cooch & G. C. White (Eds.), *A gentle introduction to program MARK*, chapter 18. Ithaca, New York. York. <http://www.phidot.org/software/mark/docs/book>.
- McClintock, B. T., Bailey, L. L., Pollock, K. H., & Simons, T. R. (2010). Unmodeled observation error induces bias when inferring patterns and dynamics of species occurrence via aural detections. *Ecology*, 91(8), 2446–2454.
- McClintock, B. T., Conn, P. B., Alonso, R. S., & Crooks, K. R. (2013). Integrated modeling of bilateral photo-identification data in mark-recapture analyses. *Ecology*, 94(7), 1464–1471.
- McClintock, B. T., Nichols, J. D., Bailey, L. L., MacKenzie, D. I., Kendall, W. L., & Franklin, A. B. (2010). Seeking a second opinion: Uncertainty in disease ecology. *Ecology Letters*, 13(6), 659–674.
- McClintock, B. T., White, G. C., Antolin, M. F., & Tripp, D. W. (2009a). Estimating abundance using mark-resight when sampling is with replacement or the number of marked individuals is unknown. *Biometrics*, 65, 236–246.
- McClintock, B. T., White, G. C., Burnham, K. P., & Pryde, M. A.. (2009b). A generalized mixed effects model of abundance for mark-resight data when sampling is without replacement. In D. L. Thomson, E. G. Cooch, & M. J. Conroy (Eds.), *Modeling demographic processes in marked populations* (pp. 271–289). New York, U.S.: Springer.
- McDonald, L. L. (1980). Line-intercept sampling for attributes other than coverage and density. *Journal of Wildlife Management*, 44(2), 530–533.
- McDonald, L. L. (2004). Sampling rare populations. In W. L. Thompson (Ed.), *Sampling rare or elusive species: Concepts, designs, and techniques for estimating population parameters* (pp. 11–42). Washington, D.C.: Island Press.
- McDonald, T. (2010). MRA: Analysis of mark-recapture data. R package version 2.7. <http://CRAN.R-project.org/package=mra>.
- McDonald, T., R. Nielson, J. Griswold, & McCann, P. (2012). *Rdistance: Distance sampling analyses*. R package version 1.1.
- McGarvey, R., Byth, K., Dixon, C. D. Day, R. W., & Feenstra J. E. (2005). Field trials and simulations of point-nearest-neighbor distance methods for estimating abalone density. *Journal of Shellfish Research*, 24(2), 393–399.
- McGregor H. W., Legge, S., Potts, J., Jones, M. E., & Johnson, C. N. (2015). Density and home range of feral cats in north-western Australia. *Wildlife Research*, 42(3), 223–231.
- McKelvey, K. S., Aubry, K. B., & Schwartz, M. K. (2008). Using anecdotal occurrence data for rare or elusive species: The Illusion of reality and a call for evidentiary standards. *BioScience*, 58(6), 549–555.
- McLaughlin, P., & Bar, H. (2021). A spatial capture–recapture model with attractions between individuals. *Environmetrics*, 32(1), e2653.
- Meador, M. R., McIntyre, J. P., & Pollock, K. H. (2011). Assessing the efficacy of single-pass backpack electrofishing to characterize fish community structure. *Transactions of the American Fisheries Society*, 132(1), 39–46.
- Meek, P. D., Ballard, G.-A., Verne, K., & Fleming, P. J. S. (2015). The history of wildlife camera trapping as a survey tool in Australia. *Australian Mammalogy*, 37(1), 1–12.
- Mellinger, D. K., Stafford, K. M., Moore, S. E., Dziak, R. P., & Matsumoto, H. (2007). An overview of fixed passive acoustic observation methods for cetaceans. *Oceanography*, 20(4), 36–45.
- Melville, G., & Stone, C. (2016). Optimising nearest neighbour information—a simple, efficient sampling strategy for forestry plot imputation using remotely sensed data. *Australian Forestry*, 79(3), 217–228.
- Melville, G. J., Welsh, A. H., & Stone, C. (2015). Improving the efficiency and precision of tree counts in pine plantations using airborne LiDAR Data and flexible-radius plots: Model-based and design-based approaches. *Journal of Agricultural, Biological, and Environmental Statistics*, 20(2), 229–257.
- Mendes, R. S., Evangelista, L. R., Thomaz, S. M., Agostinho, A. A., & Gomes, L. C. (2008). A unified index to measure ecological diversity and species rarity. *Ecography*, 31(4), 450–456.

- Metcalf, O. C., Barlow, J., & et al. (2022). Optimizing tropical forest bird surveys using passive acoustic monitoring and high temporal resolution sampling. *Remote Sensing in Ecology and Conservation*, 8(1), 45–56.
- Metcalf, K., Delavenne, J., & et al. (2013). Impacts of data quality on the size of conservation planning targets: the species-area relationship. *Biodiversity and Distributions*, 19(1), 1–13.
- Meyer, K. A., & High, B. (2011). Accuracy of removal electrofishing estimates of trout abundance in Rocky Mountain streams. *North American Journal of Fisheries Management*, 31(5), 923–933.
- Millar, R. B. (2009). Comparison of hierarchical Bayesian models for overdispersed count data using DIC and Bayes' factors. *Biometrics*, 65(3), 962–969.
- Miller, C., Joyce, P., & Waits, L. (2005). A new method for estimating the size of small populations from genetic mark-recapture data. *Molecular Ecology*, 14(7), 1991–2005.
- Miller, D. A. W., Bailey, L. L., Grant, E. H. C., McClintock, B. C., Weir, L. A., & Simon, T. R. (2015). Performance of species occurrence estimators when basic assumptions are not met: A test using field data where true occupancy status is known. *Methods in Ecology and Evolution*, 6(5), 557–565.
- Miller, D. A. W., Brehme, C. S., Hines, J. E., Nichols, J. D., & Fisher, R. N. (2012). Joint estimation of habitat dynamics and species interactions: Disturbance reduces co-occurrence of non-native predators with an endangered toad. *Journal of Animal Ecology*, 81(6), 1288–1297.
- Miller, D. A., & Grant, E. H. C. (2015). Estimating occupancy dynamics for large-scale monitoring networks: Amphibian breeding occupancy across protected areas in the northeast United States. *Ecology and Evolution*, 5(21), 4735–4746.
- Miller, D. A., Nichols, J. D., McClintock, B. T., Campbell Grant, E. H., Bailey, L. L., & Weir, L. A. (2011). Improving occupancy estimation when two types of observational error occur: Non-detection and species misidentification. *Ecology*, 92(7), 1422–1428.
- Miller, D. A., Nichols, J. D., & et al. (2013). Determining occurrence dynamics when false positives occur: Estimating the range dynamics of wolves from public survey data. *PLoS One*, 8(6), e65808.
- Miller, D. A. W., Pacifici, K., Sanderlin, J. S., & Reich, B. J. (2019). The recent past and promising future for data integration methods to estimate species' distributions. *Methods in Ecology and Evolution*, 10(1), 22–37.
- Miller, D. A. W., Talley, B. L., Lips, K. R., & Grant, E. H. C. (2012c). Estimating patterns and drivers of infection prevalence and intensity when detection is imperfect and sampling errors occurs. *Methods in Ecology and Evolution*, 3(5), 805–859.
- Miller, D. A., Weir, L. A., McClintock, B. T., Grant, E. H. C., Bailey, L. L., & Simons, T. R. (2012). Experimental investigation of false positive errors in auditory species occurrence surveys. *Ecological Applications*, 22(5), 1665–1674.
- Miller, D. L., Burt, M. L., Rexstad, E. A., & Thomas, L. (2013). Spatial models for distance sampling data: Recent developments and future directions. *Methods in Ecology and Evolution*, 4(11), 1001–1010.
- Miller, D. L., Rexstad, E., Burt, L., Bravington, M. V., & Hedley, S. (2014). *DSM: Density surface modelling of distance sampling data*. R package version 2.2.3.
- Miller, D. L., Rexstad, E., Thomas, L., Marshall, L., & Laake, J. L. (2019). Distance Sampling in R. *Journal of Statistical Software*, 89(1), 1–28.
- Miller, D. L., & Thomas, L. (2015). Mixture models for distance sampling detection functions. *PLoS One*, 10(3), e0118726.
- Mizel, J. D., McIntyre, C. L., Lewis, S. B., Lindberg, M. S., & Schmidt, J. H. (2018). A multi-state, time-removal model for population dynamics of cliff nesting raptors. *Journal of Wildlife Management*, 82(8), 1701–1710.
- Molinari-Jobin, A., Kéry, M., & et al. (2012). Monitoring in the presence of species misidentification: The case of the Eurasian lynx in the Alps. *Animal Conservation*, 15(3), 266–273.
- Mollet, P., Kéry, M., Gardner, B., Pasinelli, G., & Royle, J. A. (2015). Estimating population size for capercaillie (*Tetrao urogallus* l.) with spatial capture-recapture models based on genotypes from one field sample. *PLoS One*, 10(6), e0129020.

- Moore, J. E., Scheiman, D. M., & Swihart, R. K. (2004). Field comparison of removal and modified double-observer modeling for estimating detectability and abundance of birds. *Auk*, 121(3), 865–876.
- Moore, P. G. (1954). Spacing in plant populations. *Ecology*, 35(2), 222–227.
- Moradi, M. (2019). Using adaptive line-transect sampling in airborne geophysics studies. *Communications in Statistics. Simulation and Computation*, 48(3), 675–683.
- Moran, P. A. P. (1951). A mathematical theory of animal trapping. *Biometrika*, 38(3–4), 307–311.
- Mordecai, R. S., Mattsson, B. J., Tzilkowski, C. J., & Cooper, R. J. (2011). Addressing challenges when studying mobile or episodic species: Hierarchical Bayes estimation of occupancy and use. *Journal of Applied Ecology*, 48(1), 56–66.
- Moreno, M., & Lele, S. R. (2010). Improved estimation of site occupancy using penalized likelihood. *Ecology*, 91(2), 341–346.
- Morgan, B. J. T., Palmer, K. J., & Ridout, M. (2007). Negative Score Test statistic. *American Statistician*, 61(4), 285–288.
- Morin, D. J., Waits, L. P., McNitt, D. C., & Kelly, M. J. (2018). Efficient single survey estimation of carnivore density using fecal DNA and spatial capture-recapture: A bobcat case study. *Population Ecology*, 60(3), 197–209.
- Morisita, M. (1954). Estimation of population density by spacing method. *Memoirs of Faculty of Science, Kyushu University, EI*, 187–197.
- Morisita, M. (1957). A new method for the estimation of density by the spacing method applicable to non-randomly distributed populations. *Physiology and Ecology*, 7, 134–144. In Japanese, but translated into English by the United States Department of Agriculture, Division of Range Management in 1960.
- Moqanaki, E. M., Milleret, C., Tourani, M., Dupont, P., & Bischof, R. (2021). Consequences of ignoring variable and spatially autocorrelated detection probability in spatial capture-recapture. *Landscape Ecology*, 36(10), 2879–2895.
- Mosher, B. A., Bailey, L. L., Hubbard, B. A., & Huyvaert, K. P. (2018). Inferential biases linked to unobservable states in complex occupancy models. *Ecography*, 41(1), 32–39.
- Muñoz, D. J., Miller, D. A. W., Sutherland, C., & Grant, E. H. C. (2016). Using spatial capture-recapture to elucidate population processes and space-use in herpetological studies. *Journal of Herpetology*, 50(4), 570–581.
- Murphy, G. I. (1960). Estimating abundance from longline catches. *Journal of the Fisheries Research Board of Canada*, 17(1), 33–40.
- Murthy, M. N., & Rao, T. J. (1988). Systematic sampling with illustrative examples. In P. R. Krishnaiah & C. R. Rao (Eds.), *Handbook of Statistics, Vol. 6, Sampling* (pp. 147–185). Amsterdam: Elsevier Science Publishers.
- Murray, K. A., Skerratt, L. F., Spear, R., & McCallum, H. (2009). Impact and dynamics of disease in species threatened by the amphibian chytrid fungus, *Batrachochytrium dendrobatidis*. *Conservation Biology*, 23(5), 1242–1252.
- Murray, L. D., Gates, R. J., & Spinola, R. M. (2011). Evaluation of three methods to estimate density and detectability from roadside point counts. *Journal of Wildlife Management*, 75(5), 1072–1081.
- Naccarato, A., & Benassi, F. (2018). On the relationship between mean and variance of world's human population density: A study using Taylor's power law. *Letters in Spatial and Resource Sciences*, 11(3), 307–314.
- Nakashima, Y., Fukasawa, K., Samejima, H., & Stephens, P. (2018). Estimating animal density without individual recognition using information derivable exclusively from camera traps. *Journal of Applied Ecology*, 55(2), 735–744.
- Narayana Reddy, M., & Ramanatha Chetty, C. K. (1982). Effect of plot shape on variability in Smith's variance law. *Experimental Agriculture*, 18(4), 333–338.
- Neelon, B., Ghosh, P., & Loewe, P. F. (2013). A spatial Poisson hurdle model for exploring geographic variation in emergency department visits. *Journal of the Royal Statistical Society, Series A, Statistics in Society*, 176(2), 389–413.

- Neilson, E. W., Avgar, T., Burton, A. C., Broadley, K., & Boutin, S. (2018). Animal movement affects interpretation of occupancy models from camera-trap surveys of unmarked animals. *Ecosphere*, 9(1), ecs2.2092.
- Nelson, J. T., & Fancy, S. G. (1999). A test of the variable circular-plot method where exact density of a bird population was Known. *Pacific Conservation Biology*, 5(2), 139–143.
- Newbold, T., Gilbert, F., Zalat, S., El-Gabbas, A., & Reader, T. (2009). Climate-based models of spatial patterns of species richness in Egypt's butterfly and mammal fauna. *Journal of Biogeography*, 36(11), 2085–2095.
- Nichols, J. D., Bailey, L. L., & et al. (2008). Multi-scale occupancy estimation and modelling using multiple detection methods. *Journal of Applied Ecology*, 45(5), 1321–1329.
- Nichols, J. D., Hines, J. E., Sauer, J. R., Fallon, F.W., Fallon, J. E., & Heglund, P. J. (2000). A double-observer approach for estimating detection probability and abundance from point counts. *Auk*, 117(2), 393–408.
- Nichols, J. D., Hines, J. E., MacKenzie, D. I., Seamans, M. E., & Guitierrez, R. J. (2007). Occupancy estimation and modeling with multiple states and state uncertainty. *Ecology*, 88(6), 1395–1400.
- Nichols, J. D., & Karanth, K. U. (2002). Statistical concepts: Assessing spatial distributions. In K. U. Karanth, & J. D. Nichols (Eds.), *Monitoring tigers and their prey: A manual for wildlife managers, researchers, and conservationists* (pp. 29–38). Bangalore, India: Centre for Wildlife Studies.
- Nichols, J. D., Thomas, L., & Conn, P. B. (2009). Inferences about landbird abundance from count data: Recent advances and future directions. In D. L. Thompson, E. G. Cooch, & M. J. Conroy (Eds.), *Modeling demographic processes in marked populations* (pp. 201–235). New York, NY: Springer.
- Niedballa, J., Sollmann, R., Courtiol, A., & Wilting, A. (2016). cam-trapR: An R package for efficient camera trap data management. *Methods in Ecology and Evolution*, 7(12), 1457–1462.
- Nielson, R., Gray, B. R., McDonald, L. L., & Heglund, P. J. (2011). Estimating site occupancy rates for aquatic plants using spatial sub-sampling designs when detection probabilities are less than one. *Aquatic Botany*, 95(3), 221–225.
- Norberg, A., Abrego, N., & et al. (2019). A comprehensive evaluation of predictive performance of 33 species distribution models at species and community levels. *Ecological Monographs*, 89(3), e01370.
- Norris, J. L., & Pollock, K. H. (1995). A capture- recapture model with heterogeneity and behavioural response. *Environmental and Ecological Statistics*, 2(4), 305–313.
- Norris, J. L., & Pollock, K. H. (1996a). Nonparametric MLE under two closed capture-recapture models with heterogeneity. *Biometrics*, 52(2), 639–649.
- Norris, J. L., & Pollock, K. H. (1996b). Including model uncertainty in estimating variances in multiple capture studies. *Journal of Ecological and Environmental Statistics*, 3(3), 235–244.
- Norris, J. L., & Pollock, K. H. (1998). Non-parametric MLE for Poisson species abundance models allowing for heterogeneity between species. *Environmental and Ecological Statistics*, 5(4), 391–402.
- Norris, J. L., & Pollock, K. H. (2001). Nonparametric MLE incorporation of heterogeneity and model testing into premarked cohort studies. *Journal of Ecological and Environmental Statistics*, 8(1), 21–32.
- Northrup, J. M., & Gerber, B. D. (2018). A comment on priors for Bayesian occupancy models. *PLoS One*, 13(2), e0192819.
- Northrup, A., Saborowski, J., Nuske, R. S., & Stoyan, D. (2010). Density estimation based on k -tree sampling and point pattern reconstruction. *Canadian Journal of Forest Research*, 40(5), 953–967.
- O'Connell, F., Karanth, K. U., & Nichols, J. D. (Eds.). (2011). *Camera traps in animal ecology: Methods and analyses*. Heidelberg: Springer.
- O'Donnell, K. M., Thompson, 3rd, F. R., & Semlitsch, R. D. (2015). Partitioning detectability components in populations subject to within-season temporary emigration using binomial mixture models. *PLoS One*, 10(3), e0117216.

- Oedekoven, C. S., Buckland, S. T., Mackenzie, M. L., King, R., Evans, K. O., & Burger, L. W. (2014). Bayesian methods for hierarchical distance sampling models. *Journal of Agricultural, Biological, and Environmental Statistics*, 19(2), 219–239.
- Oedekoven, C. S., King, R., Buckland, S. T., Mackenzie, M. L., Evans, K. O., & Burger, Jr., L.W. (2016). Using hierarchical centering to facilitate a reversible jump MCMC algorithm for random effects models. *Computational Statistics and Data Analysis*, 98, 79–90.
- Oedekoven, C. S., Marques, T. A., & et al. (2021). A comparison of three methods for estimating call densities of migrating bowhead whales using passive acoustic monitoring. *Environmental and Ecological Statistics*, 29(1), 101–125.
- O’Hara, R. B., & Sillanpää, M. J. (2009). A review of Bayesian variable selection methods: What, how and which. *Bayesian Analysis*, 4(1), 85–117.
- Olds, B. P. Jerde, C. L., & et al. (2016). Estimating species richness using environmental DNA. *Ecology and Evolution*, 6(12), 4214–4226.
- Otis, D. L. (1980). An extension of the change-in-ratio method. *Biometrics*, 36(1), 141–147.
- Otis, D. L., Burnham, K. P., White, G. C., & Anderson, D. R. (1978). Statistical inference from capture data on closed animal populations. *Wildlife Monographs*, 62, 1–135.
- Otto, C. R. V., Bailey, L. L., & Roloff, G. J. (2013). Improving species occupancy estimation when sampling violates the closure assumption. *Ecography*, 36(12), 1299–1309.
- Ovaskainen, O., Abrego, N., Halme, P., & Dunson, D. (2016a). Using latent variable models to identify large networks of species-to-species associations at different spatial scales. *Methods in Ecology and Evolution*, 7(5), 549–555.
- Ovaskainen, O., Hottola, J., & Siitonen, J. (2010). Modeling species co-occurrence by multivariate logistic regression generates new hypotheses on fungal interactions. *Ecology*, 91(9), 2514–2521.
- Ovaskainen, O., Roy, D. B., Fox, R., & Anderson, B. J. (2016b). Uncovering hidden spatial structure in species communities with spatially explicit joint species distribution models. *Methods in Ecology and Evolution*, 7(4), 428–436.
- Ovaskainen, O., & Soininen, J. (2011). Making more out of sparse data: Hierarchical modeling of species communities. *Ecology*, 92(2), 289–295.
- Ovaskainen O., Tikhonov G., & et al. (2017a). How are species interactions structured in species-rich communities? A new method for analysing time-series data. *Proceedings of the Royal Society, Series B*, 284(1855), 20170768.
- Ovaskainen, O., Tikhonov, G., & et al. (2017b). How to make more out of community data? A conceptual framework and its implementation as models and software. *Ecology Letters*, 20(5) 561–576.
- Overton, W. S. (1969). Estimating the numbers of animals in wildlife populations. In R. H. Giles (Ed.), *Wildlife management techniques* (pp. 403–456). Washington, D.C.: The Wildlife Society.
- Oyster, J. H., Keren, I. N., Hansen, S. J. K., & Harris, R. B. (2018). Hierarchical mark-recapture distance sampling to estimate moose abundance. *Journal of Wildlife Management*, 82(8), 1668–1679.
- Ozdemir, I., Mert, A., Ozkan, U. Y., Aksan, S., & Unal, Y. (2018). Predicting bird species richness and micro-habitat diversity using satellite data. *Forest Ecology and Management*, 424, 483–493.
- Pacifici, K., Dorazio, R. M., & Conroy, M. J. (2012). A two-phase sampling design for increasing detections of rare species in occupancy surveys. *Methods in Ecology and Evolution*, 3(4), 721–730.
- Pacifici, K. P., Reich, B. J., Miller, D. A., & et al. (2017). Integrating multiple data sources in species distribution modeling: A framework for data fusion. *Ecology*, 98(3), 840–850.
- Packard, K. C., & Radtke, P. J. (2007). Forest sampling combining fixed- and variable-radius sample plots. *Canadian Journal of Forest Research*, 37(8), 1460–1471.
- Palencia, P., Vicente, J., Barroso, P., Barasona, J.Á., Sorriquer, R. C., & Acevedo, P. (2019). Estimating day range from camera-trap data: The animals’ behaviour as a key parameter. *Journal of Zoology*, 309(3), 182–190.

- Palka, D. L., & Hammond, P. S. (2001). Accounting for responsive movement in line transect estimates of abundance. *Canadian Journal of Fisheries and Aquatic Sciences*, 58(4), 777–787.
- Palmer, M., Balle, S., March, D., Alos, J., & Linde, M. (2011). Size estimation of circular home range from fish mark-release-(single)-recapture data: Case study of a small labrid targeted by recreational fishing. *Marine Ecology Progress Series*, 430, 87–97.
- Paloheimo, J. E. (1961). Studies on estimation of mortalities. 1: Comparison of a method described by Beverton and Holt and a new linear formula. *Journal of the Fisheries Research Board of Canada*, 18, 645–662.
- Parsons, A. W., Simons, T. R., Pollock, K. H., Stoskopf, M. K., Stocking, J. J., & O'Connell, Jr., A. F. (2015). Camera traps and mark-resight models: The value of ancillary data for evaluating assumptions. *Journal of Wildlife Management*, 79(7), 1163–1172.
- Patil, G. P., & Rao, C.R. (1978). Weighted distributions and size-biased sampling with applications to wildlife populations and human families. *Biometrics*, 34(2), 179–189.
- Patil, S. A., Burnham, K. P., & Kovner, J. L. (1979). Nonparametric estimation of plant density by the distance method. *Biometrics*, 35(3), 597–604.
- Paulik, G. J., & Robson, D. S. (1969). Statistical calculations for change-in-ratio estimators of population parameters. *Journal of Wildlife Management*, 33(1), 1–27.
- Pavlacky, Jr., D. C., Blakesley, J. A., White, G. C., Hanni, D. J., & Lukacs, P. M. (2012). Hierarchical multi-scale occupancy estimation from monitoring wildlife populations. *Journal of Wildlife Management*, 76(1), 154–162.
- Peak, R. G. (2011). A field test of the distance sampling method using Golden-cheeked Warblers. *Journal of Field Ornithology*, 82(3), 311–319.
- Pearce, J. L., & Boyce, M. S. (2006). Modelling distribution and abundance with presence-only data. *Journal of Applied Ecology*, 43(3), 405–412.
- Pérez-Granados, C., Bota, G., & et al. (2019). Vocal activity rate index: A useful method to infer terrestrial bird abundance with acoustic monitoring. *Ibis*, 161(4), 901–907.
- Pérez-Granados, C., & Traba, J. (2021). Estimating bird density using passive acoustic monitoring: A review of methods and suggestions for further research. *Ibis*, 163(3), 765–778.
- Péron, G., & Garel, M. (2018). Analysing patterns in population dynamics using repeated population surveys with three types of detection data. *Ecological Indicators*, 106, 105546.
- Peterka, T., Syrovátká, V., & et al. (2020). Is variable plot size a serious constraint in broad-scale vegetation studies? A case study on fens. *Journal of Vegetation Science*, 31(4), 594–605.
- Peterson, J. T., Thurow, R. F., & Guzevich, J. W. (2004). An evaluation of multi-pass electrofishing for estimating the abundance of stream-dwelling salmonids. *Transactions of the American Fisheries Society*, 133(2), 462–475.
- Petit, E., & Valiere, N. (2006). Estimating population size with noninvasive capture-mark-recapture data. *Conservation Biology*, 20(4), 1062–1073.
- Petrides, G. A. (1949). View points on the analysis of open season sex and age ratios. *Transactions of the North American Wildlife Conference*, 14, 391–410.
- Phillips, S., Anderson, R., & Schapire, R. (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190(3), 231–259.
- Phillips, S. J., & Dudík, M. (2008). Modeling of species distributions with Maxent: New extensions and a comprehensive evaluation. *Ecography*, 31(2), 161–175.
- Phillips, S. J., Dudík, M., & et al. (2009). Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data. *Ecological Applications*, 19(1), 181–197.
- Phillips, S., & Elith, J. (2011). Logistic methods for resource selection functions and presence-only species distribution models. In *Proceedings of the 25th AAAI Conference on Artificial Intelligence* (pp. 1384–1389). San Francisco, California, USA.
- Phillips, S. J., & Elith, J. (2013). On estimating probability of presence from use-availability or presence-background data. *Ecology*, 94(6), 1409–1419.
- Pillay, R., Miller, D. A., Hines, J. E., Joshi, A. A., & Madhusudan, M. (2014). Accounting for false positives improves estimates of occupancy from key informant interviews. *Diversity and Distributions*, 20(2), 223–235.

- Pledger, S. (2000). Unified maximum likelihood estimates for closed capture–recapture models using mixtures. *Biometrics*, 56(2), 434–442.
- Pledger, S. (2005). The performance of mixture models in heterogeneous closed population capture–recapture. *Biometrics*, 61(3), 868–873.
- Plhal, R., Kamler, J., & Homolka, M. (2014). Faecal pellet group counting as a promising method of wild boar population density estimation. *Acta Theriologica*, 59(4), 561–569.
- Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd International Workshop on Distributed Computing* (pp. 1–8). Vienna, Austria, March 20–22.
- Plummer, M. (2016). Rjags: Bayesian graphical models using MCMC. In *Interface to the JAGS MCMC library*.
- Pollard, J. H. (1971). On distance estimators of density in randomly distributed forests. *Biometrics*, 27(4), 991–1002.
- Pollard, J. H., & Buckland, S. T. (1997). A strategy for adaptive sampling in shipboard line transect surveys. *Report of the International Whaling Commission*, 47, 921–931.
- Pollard, J. H., Palka, D., & Buckland, S. T. (2002). Adaptive line transect sampling. *Biometrics*, 58(4), 862–870.
- Pollock, K. H. (1978). A family of density estimators for line-transect sampling. *Biometrics*, 34(3), 475–478.
- Pollock, K. H. (1982). A capture-recapture design robust to unequal probability of capture. *Journal of Wildlife Management*, 46(3), 757–760.
- Pollock, K. H., Hines, J. E., & Nichols, J. D. (1984). The use of auxiliary variables in capture-recapture and removal experiments. *Biometrics*, 40(2), 329–340.
- Pollock, K. H., Marsh, H., Bailey, L. L., Farnsworth, G. L., Simons, T. R., & Alldredge, M. W. (2004). Separating components of detection probability in abundance estimation: An overview with diverse examples. In W. L. Thompson (Ed.), *Sampling rare or elusive species: Concepts, designs, and techniques for estimating population parameters* (pp. 43–58). Washington, D.C.: Island Press.
- Pollock, K. H., Marsh, H. D., Lawler, I. R., & Alldredge, M. T. (2006). Estimating animal abundance in heterogeneous environments: An application to aerial surveys for dugongs. *Journal of Wildlife Management*, 70(1), 255–262.
- Pollock, K. H., & Otto, M. C. (1983). Robust estimation of population size in a closed animal populations from capture-recapture experiments. *Biometrics*, 39(4), 1035–1049.
- Pollock, L. J., Tingley, R., & et al. (2014). Understanding co-occurrence by modelling species simultaneously with a joint species distribution model (JSDM). *Methods in Ecology and Evolution*, 5(5), 397–406.
- Popescu, V. D., de Valpine, P., & Sweitzer, R. A. (2014). Testing the consistency of wildlife data types before combining them: The case of camera traps and telemetry. *Ecology and Evolution*, 4(7), 933–943.
- Popovic, G. C., Hui, F. K. C., & Warton, D. I. (2018). A general algorithm for covariance modeling of discrete data. *Journal of Multivariate Analysis*, 165, 86–100.
- Popovic, G. C., Warton, D. I., Thomson, F. J., & Hui, F. K. C. (2019). Untangling direct species associations from indirect mediator species effects with graphical models. *Methods in Ecology and Evolution*, 10(9) 1571–1583.
- Potts, J. M., & Elith, J. (2006). Comparing species abundance models. *Ecological Modeling*, 199(2), 153–163.
- Powell, R. A., & Mitchell, M. S. (2012). What is a home range? *Journal of Mammalogy*, 93(4), 948–958.
- Premaratna, W. A. L., Schwarz, C. J., & Jones, T. S. (2018). Partial stratification in two-sample capture–recapture experiments. *Environmetrics*, 29(4), e2498.
- Prentius, W., & Anton Grafström, A. (2022). Two-phase adaptive cluster sampling with circular field plots. *Environmetrics*, 33(5), e2729.

- Prevost, E., Cuzol, A., Bagliniere, J.-L., Parent, E., & Rivot, E. (2008). Hierarchical Bayesian modelling with habitat and time covariates for estimating riverine fish population size by successive removal method. *Canadian Journal of Fisheries and Aquatic Sciences*, 65(1), 117–133.
- Prieto Gonzalez, R., Thomas, L., & Marques, T. A. (2017). Estimation bias under model selection for distance sampling detection functions. *Environmental and Ecological Statistics*, 24(3), 399–414.
- Prodan, M. (1968). Punkstichprobe für die forsteinrichtung (A point sample for forest management planning). *Forst und Holzwirt*, 23(11), 225–226.
- Puerta, P., Hunsicker, M. E., Quetglas, A., Alvarez-Berastegui, D., Esteban, A., & Gonzalez, M. M. (2015). Spatially explicit modeling reveals cephalopod distributions match contrasting trophic pathways in the Western Mediterranean Sea. *PLoS One*, 10(7), e0133439.
- Purcell, J. E. (2009). Extension of methods for jellyfish and ctenophore trophic ecology to large-scale research. *Hydrobiologia*, 616(1), 23–50.
- Pyšek, P., Jarošk, V., & et al. (2012). A global assessment of invasive plant impacts on resident species, communities and ecosystems: The interaction of impact measures, invading species' traits and environment. *Global Change Biology*, 18(5), 1725–1737.
- Quang, P. X. (1993). Nonparametric estimators for variable circular plot surveys. *Biometrics*, 49(3), 837–852.
- Quang, P. X., & Becker, E. F. (1996). Line transect sampling under varying conditions with application to aerial surveys. *Ecology*, 77(4), 1297–1302.
- Quang, P. X., & Becker, E. F. (1997). Combining line transect and double count sampling techniques for aerial surveys. *Journal of Agricultural, Biological*, 2(2), 230–242.
- R Development Core Team. (2011). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.
- R Development Core Team. (2016). *R, version 3.3.1*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.
- Ralph, C. J., Droege, S., & Sauer, J. R. (1995a). Managing and monitoring birds using point counts: Standards and applications. In *Monitoring bird populations by point counts*. USDA Forest Service General Technical Report PSW-GTR-149 (pp. 161–175).
- Ralph, C. J., Geupel, G. R., Pyle, P., Martin, T. E., & DeSante, D. F. (1993). Handbook of field methods for monitoring landbirds. In P. Pyle, T. E. Martin, & D. F. DeSante (Eds.), *USDA Forest Service General Technical Report PSW-GTR-144*.
- Ralph, C. J., Sauer, J. R., & Droege, S. (Eds.). (1995b). *Monitoring bird populations by point counts*. USDA Forest Service General Technical Report PSW-GTR-149.
- Ramezani, H., Grafström, A., Naghavi, H., Fallah, A., Shataee, S. H., & Soosani, J. (2016). Evaluation of K-tree distance and fixed-sized plot sampling in zagros forests of western Iran. *Journal of Agricultural Science and Technology*, 18(1), 155–170.
- Ramos, P. L., Sousa, I., & et al. (2020). A review of capture-recapture methods and its possibilities in ophthalmology and vision sciences. *Ophthalmic Epidemiology*, 27(4), 310–324.
- Ramsey, D. S. L., Caley, P. A., & Robley, A. (2015). Estimating population density from presence-absence data using a spatially explicit model. *Journal of Wildlife Management*, 79(3), 491–499.
- Ramsey, D. S. L., Parkes, J., & Morrison, S. A. (2009). Quantifying eradication success: The removal of feral pigs from Santa Cruz Island, California. *Conservation Biology*, 23(2), 449–459.
- Ramsey, F. L., & Scott, J. M. (1979). Estimating population densities from variable circular plot surveys. In R. M. Cormack, G. P. Patil, & D. S. Robson (Eds.), *Sampling biological populations* (pp. 155–181). Fairland, Maryland: International Publishing House.
- Rankin, S., Oedekoven, C., & Archer, F. (2020). Mark recapture distance sampling: Using acoustics to estimate the fraction of dolphins missed by observers during shipboard line-transect surveys. *Environmental and Ecological Statistics*, 27(2), 233–251.
- Rasch, G. (1961). On general laws and the meaning of measurement in psychology. In J. Neyman (Ed.), *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* (pp. 321–333). Berkeley, CA: University of California Press.

- Rathbun, S. L., & Fei, S. (2006). A spatial zero-inflated Poisson regression model for Oak regeneration. *Environmental and Ecological Statistics*, 13(4), 409–426.
- Rees, H. C., Maddison, B. C., Middleditch, D. J., Patmore, J. R. M., & Gough, K. C. (2014). The detection of aquatic animal species using environmental DNA—a review of eDNA as a survey tool in ecology. *Journal of Applied Ecology*, 51(5), 1450–1459.
- Rehman, Z., Toms, C. N., & Finch, C. (2016). Estimating abundance: A non parametric mark recapture approach for open and closed systems. *Environmental and Ecological Statistics*, 23(4), 623–638.
- Rehman, Z., & Xie, Y. (2017). Reply to “Comment: Estimating abundance-a non parametric mark recapture approach for open and closed systems”. *Environmental and Ecological Statistics*, 24(4), 595–598.
- Reich, B. J., & Gardner, B. (2014). A spatial capture-recapture model for territorial species. *Environmetrics*, 25(8), 630–637.
- Reich, B. J., Pacifici, K., Stallings, J. W., & Kriticos, D. (2018). Integrating auxiliary data in optimal spatial design for species distribution modelling. *Methods in Ecology and Evolution*, 9(6), 1626–1637.
- Renner, I. W., & Warton, D. I. (2013). Equivalence of MAXENT and Poisson point process models for species distribution modeling in ecology. *Biometrics*, 69(1), 274–281.
- Reynolds, R. T., Scott, J. M., & Nussbaum, R. A. (1980). A variable circular-plot method for estimating bird numbers. *Condor*, 82(3), 309–313.
- Rexstad, E., & Burnham K. (1991). *User's guide for interactive program CAPTURE*. Colorado Cooperative Fish and Wildlife Research Unit, Fort Collins.
- Rhodes, J. R., Lunney, D., Moon, C., Matthews, A., & McAlpine, C. A. (2011). The consequences of using indirect signs that decay to determine species' occupancy. *Ecography*, 34(1), 141–150.
- Rich, L. N., Kelly, M. J., & et al. (2014). Comparing capture–recapture, mark–resight, and spatial mark–resight models for estimating puma densities via camera traps. *Journal of Mammalogy*, 95(2), 382–391.
- Rich, L. N., Russell, R. E., & et al. (2013). Estimating occupancy and predicting numbers of gray wolf packs in Montana using hunter surveys. *Journal of Wildlife Management*, 77(6), 1280–1289.
- Richmond, O. M., Hines, J. E., & Beissinger, S. R. (2010). Two-species occupancy models: A new parameterization applied to co-occurrence of secretive rails. *Ecological Applications*, 20(7), 2036–2046.
- Ricker, W. E. (1958). Handbook of computations for biological statistics of fish populations. *Fisheries Research Board of Canada, Bulletin*, 119, 1–300.
- Riddle, J. D., Moorman, C. E., & Pollock, K. H. (2008). A comparison of methods for estimating Northern Bobwhite covey detection probabilities. *Journal of Wildlife Management*, 72(6), 1437–1442.
- Riddle, J. D., Mordecai, R. S., Pollock, K. H., & Simons, T. R. (2010). Effects of prior detections on estimates of detection probability, abundance, and occupancy. *Auk*, 127(1), 94–99.
- Riddle, J. D., Pollock, K. H., & Simons, T. R. (2010). An unreconciled double-observer method for estimating detection probability and abundance. *Auk*, 127(4), 841–849.
- Riddle, J. S., Stanislav, S. J., Pollock, K. H., Moorman, C. E., & Perkins, F. S. (2010). Separating components of the detection process with combined methods: An example with northern bobwhite. *Journal of Wildlife Management*, 74(6), 1319–1325.
- Ridgway, M. S. (2010). Line transect distance sampling in aerial surveys for double-crested cormorants in coastal regions of Lake Huron. *Journal of Great Lakes research*, 36(3), 403–410.
- Ridout, M. S., & Linkie, M. (2009). Estimating overlap of daily activity patterns from camera trap data. *Journal of Agricultural, Biological, and Environmental Statistics*, 14(3), 322–337.
- Rissanen, J. (1983). A universal prior for integers and estimation by minimum description length. *Annals of Statistics*, 11(2), 416–431.
- Rivest, L. P., & Lévesque, T. (2001). Improved log-linear model estimators of abundance in capture-recapture experiments. *Canadian Journal of Statistics*, 29(4), 555–572.

- Robbins, J. R., Babey, L., & Embling, C. B. (2020). Citizen science in the marine environment: Estimating common dolphin densities in the north-east Atlantic. *Peer Journal*, 2020(2), e8335.
- Robert, C. P. (2014). On the Jeffreys-Lindley paradox. *Philosophy of Science*, 81, 216–232.
- Roberts, H. V. (1967). Informative stopping rules and inferences about population size. *Journal of the American Statistical Association*, 62(319), 763–775.
- Robinson, J. G., Ahmed, B., Das Gupta, P., & Woodrow, K. A. (1993). Estimation of population coverage in the 1990 United States Census based on demographic analysis. *Journal of the American Statistical Association*, 88(423), 1061–1079.
- Rocchini, D., Boyd, D. S., & et al. (2016). Satellite remote sensing to monitor species diversity: Potential and pitfalls. *Remote Sensing in Ecology and Conservation*, 2(1), 25–36.
- Romairone, J., Jiménez, J., Luque-Larena, J. J., & Mougeot, F. (2018). Spatial capture-recapture design and modelling for the study of small mammals. *PLoS One*, 13(6), e0198766.
- Ronconi, R. A., & Burger, A. E. (2009). Estimating seabird densities from vessel transects: Distance sampling and implications for strip transects. *Aquatic Biology*, 4, 297–309.
- Rondinini, C., & Chiozza, F. (2010). Quantitative methods for defining percentage area targets for habitat types in conservation planning. *Biological Conservation*, 143(7), 1646–1653.
- Rosenberger, A. E., & Dunham, J. B. (2005). Validation of abundance estimates from mark-recapture and removal techniques for rainbow trout captured by electrofishing in small streams. *North American Journal of Fisheries Management*, 25(4), 1395–1410.
- Rosenstock, S. S., Anderson, D. R., Giesen, K. M., Leukering, T., & Carter, M. F. (2002). Landbird counting techniques: Current practices and an alternative. *Auk*, 119(1), 46–53.
- Ross, S. M. (2010). *Introduction to probability models* (10th ed.). Burlington, MA: Academic Press.
- Rossi, J.-P. (2011). Rich: An R package to analyse species richness. *Diversity*, 3(1), 112–120.
- Rossman, S., Yackulic, C. B., Saunders, S. P., Reid, J., Davis, R., & Zipkin, E. F. (2016). Dynamic N-occupancy models: Estimating demographic rates and local abundance from detection-nondetection data. *Ecology*, 97(12), 3300–3307.
- Rota, C. T., Ferreira, M. A. R., & et al. (2016). A multispecies occupancy model for two or more interacting species. *Methods in Ecology and Evolution*, 7(10), 1164–1173.
- Rota, C. T., Fletcher, R. J., Dorazio, R. M., & Betts, M. G. (2009). Occupancy estimation and the closure assumption. *Journal of Applied Ecology*, 46(6), 1173–1181.
- Rota, C. T., Fletcher, R. J., Jr, Evans, J. M., & Hutto, R. L. (2011). Does accounting for imperfect detection improve species distribution models. *Ecography*, 34(4), 659–670.
- Rota, C. T., Wikle, C. K., Kays, R. W., & et al. (2016). A two-species occupancy model accommodating simultaneous spatial and interspecific dependence. *Ecology*, 97(1), 48–53.
- Rout, T. M., Kirkwood, R., Sutherland, D. R., Murphy, S., & McCarthy, M. A. (2014). When to declare successful eradication of an invasive predator? *Animal Conservation*, 17(2), 125–132.
- Roux, F., Steyn, G., Hay, C., & Wagenaar, I. (2018). Movement patterns and home range size of tigerfish (*Hydrocynus vittatus*) in the Incomati River system, South Africa. *Koedoe—African Protected Area Conservation and Science*, 60(1), 1–13.
- Rowcliffe, J. M., Carbone, C., Jansen, P. A., Kays, R., & Kranstauber, B. (2011). Quantifying the sensitivity of camera traps: An adapted distance sampling approach. *Methods in Ecology and Evolution*, 2(5), 464–476.
- Rowcliffe, J. M., Field, J., Turvey, S. T., & Carbone, C. (2008). Estimating animal density using camera traps without the need for individual recognition. *Journal of Applied Ecology*, 45(4), 1228–1236.
- Rowcliffe, J. M., Kays, R., Kranstauber, B., Carbone, C., & Jansen, P. A. (2014). Quantifying levels of animal activity using camera trap. *Methods in Ecology and Evolution*, 5(11), 1170–1179.
- Roy, C., Gilliland, S. G., & Reed, E. T. (2021). A hierarchical dependent double-observer method for estimating waterfowl breeding pairs abundance from helicopters. *Wildlife Biology*, 2022(1), e01003.
- Royle, J. A. (2004a). N-mixture models for estimating population size from spatially replicated counts. *Biometrics*, 60(1), 108–115.

- Royle, J. A. (2004b). Generalized estimators of avian abundance from count survey data. *Animal Biodiversity and Conservation*, 27(1), 375–386.
- Royle, J. A. (2006). Site occupancy models with heterogeneous detection probabilities. *Biometrics*, 62(1), 97–102.
- Royle, J. A., Chandler, R. B., Gazenski, K. D., & Graves., T. A. (2013a). Spatial capture-recapture models for jointly estimating population density and landscape connectivity. *Ecology*, 94(2), 287–294.
- Royle, J. A., Chandler, R. B., Sollmann, R., & Gardner, B. (2014). *Spatial capture-recapture* (1st ed.). Oxford: Elsevier.
- Royle, J. A., Chandler, R. B., Sollmann, R., & Gardner, B. (2014a). scrbook: Companion to the book: Spatial capture recapture. R package version 0.21-0.
- Royle, J. A., Chandler, R. B., Sun, C.C., & Fuller, A. K. (2013b). Integrating resource selection information with spatial capture–recapture. *Methods in Ecology and Evolution*, 4(6), 520–530.
- Royle, J. A., Chandler, R. B., Sun, C. C., Fuller, A. K., & O’Hara, R. B. (2014). Reply to Efford on “Integrating resource selection information with spatial capture–recapture”. *Methods in Ecology and Evolution*, 5(7), 603–605.
- Royle, J. A., Chandler, R. B., Yackulic, C., & Nichols, J. D. (2012). Likelihood analysis of species occurrence probability from presence-only data for modelling species distributions. *Methods in Ecology and Evolution*, 3(3), 545–554.
- Royle, J. A., & Converse, S. J. (2014). Hierarchical spatial capture–recapture models: Modelling population density in stratified populations. *Methods in Ecology and Evolution*, 5(1), 37–43.
- Royle, J. A., Dawson, D. K., & Bates, S. (2004). Modeling abundance effects in distance sampling. *Ecology*, 85(6), 1591–1597.
- Royle, J. A., & Dorazio, R. M. (2009). *Hierarchical modeling and inference in ecology: The analysis of data from populations, metapopulations and communities*. London: Academic Press.
- Royle, J. A., & Dorazio, R. M. (2012). Parameter-expanded data augmentation for Bayesian analysis of capture–recapture models. *Journal of Ornithology*, 152(Suppl 2), S521–S537.
- Royle, J. A., Dorazio, R. M., & Link, W. A. (2007a). Analysis of multinomial models with unknown index using data augmentation. *Journal of Computational and Graphical Statistics*, 16(1), 67–85.
- Royle, J. A., Fuller, A. K., & Sutherland, C. (2016). Spatial capture–recapture models allowing Markovian transience or dispersal. *Population Ecology*, 58(1), 53–62.
- Royle, J. A., Fuller, A. K., & Sutherland, C. (2018). Unifying population and landscape ecology with spatial capture–recapture. *Ecography*, 41(3), 444–456.
- Royle, J. A., & Gardner, B. (2011). Hierarchical spatial capture–recapture models for estimating density from trapping arrays. In F. O’Connell, K. U. Karanth, & J. D. Nichols, (Eds.), *Camera traps in animal ecology: Methods and analyses* (pp. 163–190). Heidelberg: Springer.
- Royle, J. A., Karanth, K. U., Gopalaswamy, A. M., & Kumar, N. S. (2009). Bayesian inference in camera trapping studies for a class of spatial capture-recapture models. *Ecology*, 90(11), 3233–3244.
- Royle, J. A.,& Kéry, M. (2007). A Bayesian state-space formulation of dynamic occupancy models. *Ecology*, 88(7), 1813–1823.
- Royle, J. A., Kéry, M., Gautier, M., & Schmid, H. (2007b). Hierarchical spatial models of abundance and occurrence from imperfect survey data. *Ecological Monographs*, 7(3), 465–481.
- Royle, J. A., Kéry, M., & Guélat, J. (2011). Spatial capture-recapture models for search-encounter data. *Methods in Ecology*, 2(6), 602–611.
- Royle, J. A., & Link, W. A. (2002). Random effects and shrinkage estimation in capture-recapture models. *Journal of Applied Statistics*, 29(1–4), 329–351.
- Royle, J. A., & Link, W. A. (2005). A general class of multinomial mixture models for anuran calling survey data. *Ecology*, 86(9), 2505–2512.
- Royle, J. A., & Link, W. A. (2006). Generalized site occupancy models allowing for false positive and false negative errors. *Ecology*, 87(4), 835–841.

- Royle, J. A., & Nichols, J. (2003). Estimating abundance from repeated presence-absence data or point counts. *Ecology*, 84(3), 777–790.
- Royle, J. A., Nichols, J. D., Karanth, K. U., & Gopalaswamy, A. (2009). A hierarchical model for estimating density in camera trap studies. *Journal of Applied Ecology*, 46(1), 118–127.
- Royle, J. A., Sutherland, C., Fuller, A. K., & Sun, C. C. (2015). Likelihood analysis of spatial capture-recapture models for stratified or class structured populations. *Ecosphere*, 6(2), 1–11.
- Royle, J. A., & Young, K. V. (2008). A hierarchical model for spatial capture-recapture data. *Ecology*, 89(8), 2281–2289.
- Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society, Series B*, 71(2), 319–392.
- Ruell, E. W., Riley, S. P. D., Douglas, M. R., Pollinger, J. P., & Crooks, K. R. (2009). Estimating bobcat population sizes and densities in a fragmented urban landscape using noninvasive capture–recapture sampling. *Journal of Mammalogy*, 90(1), 129–135.
- Ruete, A., Pärt, T., Berg, A., & Knape, J. (2017). Exploiting opportunistic observations to estimate changes in seasonal site use: An example with wetland birds. *Ecology and Evolution*, 7(15), 5632–5644.
- Ruiz, P., & Laplanche, C. (2010). A hierarchical model to estimate the abundance and biomass of salmonids by using removal sampling and biometric data from multiple locations. *Canadian Journal of Fisheries and Aquatic Sciences*, 67(12), 2032–2044.
- Ruiz-Gutiérrez, V., Hooten, M. B., & Grant, E. H. C. (2016). Uncertainty in biological monitoring: A framework for data collection and analysis to account for multiple sources of sampling bias. *Methods in Ecology and Evolution*, 7(8), 900–909.
- Ruiz-Gutiérrez, V., & Zipkin, E. F. (2011). Detection biases yield misleading patterns of species persistence and colonization in fragmented landscapes. *Ecosphere*, 2(5), 1–14.
- Russell, J. C., Stjernman, M., Lindström, A., & Smith, H. G. (2015). Community occupancy before-after-control-impact (CO-BACI) analysis of Hurricane Gudrun on Swedish forest birds. *Ecological Applications*, 25(3), 685–694.
- Russell, R. E., Royle, J. A., Saab, V. A., Lehmkuhl, J. F., Block, W. M., & Sauer, J. R. (2009). Modeling the effects of environmental disturbance on wildlife communities: Avian responses to prescribed fire. *Ecological Applications*, 19(5), 1253–1263.
- Russell, R. E., Royle, J. A., & et al. (2012). Estimating abundance of mountain lions from unstructured spatial sampling. *Journal of Wildlife Management*, 76(8), 1551–1561.
- Salasar, L. E. B., Leite, J. G., & Louzada, F. (2015). On the integrated maximum likelihood estimators for a closed population capture–recapture model with unequal capture probabilities. *Journal of Theoretical and Applied Statistics*, 49(6), 1204–1220.
- Salvatier, J., Wiecki, T. V., & Fonnesbeck, C. (2016). Probabilistic programming in Python using PyMC3. *PeerJ Computer Science*, 2, e55.
- Sanathanan, L. (1972a). Estimating the size of a multinomial population. *Annals of Mathematical Statistics*, 43(1), 142–152.
- Sanathanan, L. (1972b). Models and estimation methods in visual scanning experiments. *Techometrics*, 14(4), 813–829.
- Sanathanan, L. (1977). Estimating the size of a truncated sample. *Journal of the American Statistical Association*, 72(359), 669–672.
- Sanderlin, J. S., Block, W. M., & Ganey, J. L. (2014). Optimizing study design for multi-species avian monitoring programmes. *Journal of Applied Ecology*, 51(4), 860–870.
- Sauer, J. R., Blank, P. J., Zipkin, E. F., Fallon, J. E., & Fallon, F. W. (2013). Using multi-species occupancy models in structured decision making on managed lands. *Journal of Wildlife Management*, 77(1), 117–127.
- Sauer, J., Fallon, J., & Johnson, R. (2003). Use of North American Breeding Bird Survey data to estimate population change for bird conservation regions. *Journal of Wildlife Management*, 67(2), 372–389.
- Sauer, J. R., & Link, W. A. (2011). Analysis of the North American Breeding Bird Survey using hierarchical models. *Auk*, 128(1), 87–98.

- Schliep, E. M., Gelfand, A. E., Clark, J. S., & Kays, R. (2018.) Joint temporal point pattern models for proximate species occurrence in a fixed area using camera trap data. *Journal of Agricultural, Biological, and Environmental Statistics*, 23(3), 334–357.
- Schliep, E. M., Lany, N. K., & et al. (2018). Joint species distribution modelling for spatio-temporal occurrence and ordinal abundance data. *Global Ecology and Biogeography*, 27(1), 142–155.
- Schmid-Haas, P. (1969). Sichproben am waldrand. *Mitteilungen/Schweizerische Anstalt für das Forstliche Versuchswesen*, 45, 234–303.
- Schmidt, B. R., K(e)ry, M., Ursenbacher, S., Hyman, O. J., & Collins, J.P. (2013). Site occupancy models in the analysis of environmental DNA presence/absence surveys: A case study of an emerging amphibian pathogen. *Methods in Ecology and Evolution*, 4(7), 646–653.
- Schmidt, J. H., McIntyre, C. R., & MacCluskie, M. C. (2013). Accounting for incomplete detection: What are we estimating and how might it affect long-term passerine monitoring programs? *Biological Conservation*, 160, 130–139.
- Schnabel, Z. E. (1938). The estimation of the total fish population of a lake. *American Mathematical Monthly*, 45(6), 348–352.
- Schofield, M. R. (2017). Comment: Estimating abundance: A non parametric mark recapture approach for open and closed systems. *Environmental and Ecological Statistics*, 24(4), 587–594.
- Schofield, M. R., & Barker, R. J. (2008). A unified capture-recapture framework. *Journal of Agricultural, Biological and Environmental Statistics*, 13(4), 458–477.
- Schofield, M. R., & Barker, R. J. (2014). Hierarchical modeling of abundance in closed population capture-recapture models under heterogeneity. *Environmental and Ecological Statistics*, 21(3), 435–451.
- Schofield, M. R., & Barker, R. J. (2016). 50-year-old curiosities: Ancillarity and inference in capture-recapture models. *Statistical Science*, 31(2), 161–174.
- Schofield, M. R., & Barker, R. J. (2020). Rejoinder to “On continuous-time capture-recapture in closed populations”. *Biometrics*, 76(3), 1034–1035.
- Schofield, M. R., Barker, R. J., & Gelling, N. (2018). Continuous-time capture-recapture in closed populations. *Biometrics*, 74(2), 626–635.
- Schofield, M. R., Barker, R. J., Link, W. A., & Pavanato, H. (2023). Estimating population size: The importance of model and estimator choice. *Biometrics*, in press.
- Schofield, M. R., Barker, R. J., & MacKenzie, D. I. (2009). Flexible hierarchical mark-recapture modeling for open populations using WinBUGS. *Environmental and Ecological Statistics*, 16(3), 369–387.
- Schofield, M. R., & Bonner, S. J. (2015). Connecting the latent multinomial. *Biometrics*, 71(4), 1070–1080.
- Schroeder, N. M., Panebianco, A., Gonzalez, M. R., Carmanchahi, P. (2020). An experimental approach to evaluate the potential of drones in terrestrial mammal research: A gregarious ungulate as a study model. *Royal Society Open Science*, 7(1), 191482.
- Schwarz, C. J., & Seber, G. A. F. (1999). Estimating animal abundance: Review III. *Statistical Science*, 14(4), 427–456.
- Schweder, T. (1990). Independent observer experiments to estimate the detection function in line transect surveys of whales. *Report of the International Whaling Commission*, 40, 349–355.
- Seber, G. A. F. (1970). The effects of trap response on tag recapture estimates. *Biometrics*, 26(1), 13–22.
- Seber, G. A. F. (1979). Transects of random length. In R. M. Cormack, G. P. Patil, & D. S. Robson (Eds.), *Sampling biological populations, Statistical Ecology* (Vol. 5, pp. 183–192). Maryland: International Co-operative Publishing House.
- Seber, G. A. F. (1982). *The estimation of animal abundance* (2nd ed.). London: Griffin. Reprinted in paperback by the Blackburn press, Caldwell, N. J. (2002).
- Seber, G. A. F. (1984). *Multivariate observations*. New York: Wiley. Also available in paperback, 2004.
- Seber, G. A. F. (1986). A review of estimating animal abundance. *Biometrics*, 42(2), 267–292.

- Seber, G. A. F. (1992). A review of estimating animal abundance II. *International Statistical Review*, 60(2), 129–166.
- Seber, G. A. F. (2001). Some new directions in estimating animal population parameters. *Journal of Agricultural, Biological, and Environmental Statistics*, 6(2), 140–151.
- Seber, G. A. F. (2013). *Statistical models for proportions and probabilities*. Heidelberg: SpringerBriefs.
- Seber, G. A. F. (2015). *The linear model and hypothesis: A general unifying theory*. Heidelberg: Springer.
- Seber, G. A. F., & Felton, R. (1981). Tag loss and the Petersen mark-recapture experiment. *Biometrika*, 68(1), 211–219.
- Seber, G. A. F., & Lee, A. J. (2003). *Linear Regression*, 2nd edit., New York: Wiley.
- Seber, G. A. F., Huakau, J. T., & Simmons, D. (2000). Capture-recapture, epidemiology, and list mismatches: Two lists. *Biometrics*, 56(4), 1227–1232.
- Seber, G. A. F., & Le Cren, E. D. (1967). Estimating population parameters from catches large relative to the population. *Journal of Animal Ecology*, 36(3), 631–643.
- Seber, G. A. F., & Pemberton, J. R. (1979). The Ilne-intercept method for studying plant cuticles from rumen and fecal samples. *Journal of Wildlife Management*, 43(4), 916–925.
- Seber, G. A. F., & Salehi, M. M. (2013). *Adaptive sampling designs: Inference for sparse and clustered populations*. Berlin: Springer, SpringerBriefs in Statistics.
- Seber, G. A. F., & Schofield, M. R. (2019). *Capture-recapture: Parameter estimation for open animal populations*. Switzerland: Springer Nature.
- Seber, G. A. F., & Schwarz, C. J. (2002). Capture-recapture before and after EURING 2000. *Journal of Applied Statistics*, 29(1–4), 5–18.
- Seber, G. A. F., & Whale, J. F. (1970). The removal method for two and three samples. *Biometrics*, 26(3), 393–400.
- Seber, G. A. F., & Wild, C. J. (1989). *Nonlinear regression*. New York: Wiley. (Also in paperback, 2003).
- Self, S. G., & Liang, K. Y. (1987). Asymptotic properties of maximum-likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, 82(398), 605–610.
- Severini, T. A. (1995). Information and conditional inference. *Journal of the American Statistical Association*, 90(432), 1341–1346.
- Sewell, D., Beebee, T. J. C., & Griffiths, R. A. (2010). Optimising biodiversity assessments by volunteers: The application of occupancy modelling to large-scale amphibian surveys. *Biological Conservation*, 143(9), 2102–2110.
- Shaffer, C. A., Milstein, M. S., Yukuma, C., Marawanaru, E., & Suse, P. (2017). Sustainability and comanagement of subsistence hunting in an indigenous reserve in Guyana. *Conservation Biology*, 31(5), 1119–1131.
- Sharma, S., Sharma, H. P., Chaulagain, C., Katuwal, H. B., & Belant, J. L. (2020). Estimating occupancy of Chinese pangolin (*Manis pentadactyla*) in a protected and non-protected area of Nepal. *Ecology and Evolution*, 10(10), 4303–4313.
- Shea, C. P., Peterson, J. T., Wisniewski, J. M., & Johns, N. A. (2011). Misidentification of freshwater mussel species (*Bivalvia: Unionidae*): Contributing factors, management implications, and potential solutions. *Journal of the North American Benthological Society*, 30(2), 446–458.
- Shen, T.-J., Chao, A., & Lin, J.-F. (2003). Predicting the number of new species in further taxonomic sampling. *Ecology*, 84(3), 798–804.
- Shen, G., Wang, X., & He, F. (2020). Distance-based methods for estimating density of nonrandomly distributed populations. *Ecology*, 101(10), e03143.
- Sidle, J. G., Augustine, D. J., Johnson, D. H., Miller, S. D., Cully Jr., J. F., & Reading, R. P. (2012). Aerial surveys adjusted by ground surveys to estimate area occupied by black-tailed prairie dog colonies. *Wildlife Society Bulletin*, 36(2), 248–256.
- Sillett, T. S., Chandler, R. B., Royle, J. A., Kéry, M., & Morrison, S. A. (2012). Hierarchical distance-sampling models to estimate population size and habitat-specific abundance of an island endemic. *Ecological Applications*, 22(7), 1997–2006.

- Simons, T. R., Alldredge, M. W., Pollock, K. H., & Wetrot, J. M. (2007). Experimental analysis of the auditory detection process on avian point counts. *Auk*, 124(3), 986–999.
- Simons, T. R., Pollock, K. H., Wetrot, J. M., Alldredge, M. W., Pacifici, K., & Brewster, J. (2009). Sources of measurement error, misclassification error, and bias in auditory avian point count data. In D. L. Thomson, E. G. Cooch, & M. J. Conroy (Eds.), *Environmental and Ecological Statistics. Modeling demographic processes in marked populations* (Vol. 3). New York: Springer Science and Business Media.
- Singh, S. K., Ash, G. J., & Hodda, M. (2017). Keeping ‘one step ahead’ of invasive species: Using an integrated framework to screen and target species for detailed biosecurity risk assessment. *Biological Invasions*, 17(4), 1069–1086.
- Sirén, A. P. K., Pekins, P. J., Abdu, P. L., & Ducey, M. J. (2016). Identification and density estimation of American martens (*Martes americana*) using a novel camera-trap method. *Diversity*, 8(4), 3.
- Skalski, J., & Millspaugh, J. J. (2006). Application of multidimensional change-in-ratio methods using program USER. *Wildlife Society Bulletin*, 34(2), 433–439.
- Skalski, J. R., & Robson, D. S. (1982). A mark and removal field procedure for estimating population abundance. *Journal of Wildlife Management*, 46(3), 741–751.
- Skalski, J. R., Ryding, K. E., & Millspaugh, J. J. (2005). *Wildlife demography: Analysis of sex, age, and count data*. San Diego, California: Academic Press.
- Skaug, H. J. (2006). Markov modulated Poisson processes for clustered line transect data. *Environmental and Ecological Statistics*, 13(2), 199–211.
- Skaug, H. J., & Schweder, T. (1999). Hazard models for line transect surveys with independent observers. *Biometrics*, 55(1), 29–36.
- Sliwinski, M., Powell, L., Koper, N., Giovanni, M., & Schacht, W. (2016). Research design considerations to ensure detection of all species in an avian community. *Methods in Ecology and Evolution*, 7(4), 456–462.
- Smart, A., Tingley, R., Weeks, A., van Rooyen, A., & McCarthy, M. (2015) Environmental DNA sampling is more sensitive than a traditional survey technique for detecting an aquatic invader. *Ecological Applications*, 25(7), 1944–1952.
- Smith, D. R., Brown J. A., & Lo, N. C. H. (2004). Application of adaptive sampling to biological populations. In W. L. Thompson (Ed.), *Sampling rare or elusive species: Concepts, designs, and techniques for estimating population parameters* (pp. 77–122). Island Press: Washington, D.C.
- Smith, D. R., Gray, B. R., Newton, T. J., & Nichols, D. (2010). Effect of imperfect detectability on adaptive and conventional sampling: Simulated sampling of freshwater mussels in the upper Mississippi River. *Environmental Monitoring and Assessment*, 170(1–4), 499–507.
- Smith, M. M. & Goldberg, C. S. (2019). Data from: Occupancy in dynamic systems: accounting for multiple scales and false positives using environmental DNA to inform monitoring. Dryad Digital Repository, <https://doi.org/10.5061/dryad.bk3j9kd6q>.
- Smith, M. M. & Goldberg, C. S. (2020). Occupancy in dynamic systems: accounting for multiple scales and false positives using environmental DNA to inform monitoring. *Ecography*, 43(3), 376–386.
- Smith, G. W. (1995). *A critical review of the aerial and ground surveys of breeding waterfowl in North America*. National Biological Service Biological Science Report 5, Washington, D.C., USA.
- Smith, H. F. (1938). An empirical law describing heterogeneity in the yields of agricultural crops. *Journal of Agricultural Science*, 28(01), 1–23.
- Smith, P. J. (1988). Bayesian methods for multiple capture-recapture surveys. *Biometrics*, 44(4), 1177–1189.
- Smith, P. J. (1991). Bayesian analyses for a multiple capture-recapture model. *Biometrika*, 78(2), 399–407.
- Smolensky, N. L., & Fitzgerald, L. A. (2010). Distance sampling underestimates population densities of dune-dwelling lizards. *Journal of Herpetology*, 44(3), 372–381.

- Snäll, T., Forslund, P., Jeppsson, T., Lindhe, A., & O'Hara, R. B. (2013). Evaluating temporal variation in Citizen Science Data against temporal variation in the environment. *Ecography*, 37(3), 293–300.
- Snäll, T., Kindvall, O., Nilsson, J., & Pärt, T. (2010). Evaluating citizen-based presence data for bird monitoring. *Biological Conservation*, 144(2), 804–810.
- Solberg, E. J., Grøtan, V., Rolandsen, C. M., Brøseth, H., & Brainerd, S. (2005). Change-in-sex ratio as an estimator of population size for Norwegian moose *Alces alces*. *Wildlife Biology*, 11(2), 163–172.
- Sollmann, R., Gardner, B., & Belant, J. L. (2012). How does spatial study design influence density estimates from spatial capture–recapture models? *PLOS One*, 7(4), e34575.
- Sollmann, R., Gardner, B., & et al. (2013a). Using multiple data sources provides density estimates for endangered Florida panther. *Journal of Applied Ecology*, 50(4), 961–968.
- Sollmann, R., Gardner, B., & et al. (2013b). A spatial mark-resight model augmented with telemetry data. *Ecology*, 94(3), 553–559.
- Sollmann, R., Gardner, B., & et al. (2016). A hierarchical distance sampling model to estimate abundance and covariate associations of species and communities. *Methods in Ecology and Evolution*, 7(5), 529–537.
- Sollmann, R., Törres, N. M., & et al. (2013). Combining camera-trapping and noninvasive genetic data in a spatial capture–recapture framework improves density estimates for the jaguar. *Biological Conservation*, 167, 242–247.
- Sólymos, P., & Lele, S. (2016). Revisiting resource selection probability functions and single-visit methods: Clarification and extensions. *Methods in Ecology and Evolution*, 7(2), 196–205.
- Sólymos, P., Lele, S., & Bayne, E. (2012). Conditional likelihood approach for analyzing single visit abundance survey data in the presence of zero inflation and detection error. *Environmetrics*, 23(2), 197–205.
- Sólymos, P., Matsuoka, S. M., & et al. (2013). Calibrating indices of avian density from non-standardized survey data: Making the most of a messy situation. *Methods in Ecology and Evolution*, 4(11), 1047–1058.
- Sólymos, P., Matsuoka, S. M., & et al. (2018). Evaluating time-removal models for estimating availability of boreal birds during point count surveys: Sample size requirements and model complexity. *Condor*, 120(4), 765–786.
- Sólymos, P., Moreno, M., & Lele, S. R. (2020). Detect: Analyzing wildlife data with detection error. R package version 0.4-4. <https://cran.r-project.org/package=detect>.
- Song, W., Kim, E., Lee, D., Lee, M., & Jeon, S.-W. (2013). The sensitivity of species distribution modeling to scale differences. *Ecological Modelling*, 248(10), 113–118.
- Southgate, R., Lee, T. E., Markopoulos, N., Florance, B., & Masters, P. (2022). Eradication of feral goats *Capra hircus* from Kangaroo Island: Estimating the probability of persistence from multiple lines of evidence. *Restoration Ecology*, 30(4), e13553.
- Southwood, T. R. E. (1978). *Ecological methods, with particular reference to the study of insect populations*. The English Language Book Society and Chapman-Hall.
- Specht, H. M., Reich, H. T., & et al. (2017). Occupancy surveys with conditional replicates: An alternative sampling design for rare species. *Methods in Ecology and Evolution*, 8(12), 1725–1734.
- Spiegelhalter, D., Thomas, A., & Best, N. (2006). WinBUGS, version 1.4. MRC Biostatistics Unit, Cambridge, UK. <http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml>.
- St. Clair, K., Dunton, E., & Giudice, J. (2013). A comparison of models using removal effort to estimate animal abundance. *Journal of Applied Statistics*, 40(3), 527–545.
- Ståhl, G., Ekstöm, M., & et al. (2020). Presence–absence sampling for estimating plant density using survey data with variable plot size. *Methods in Ecology and Evolution*, 11(4), 580–590.
- Stanislav, S. J., Pollock, K. H., Simons, T. R., & Alldredge, M. W. (2010). Separation of availability and perception processes for aural detection in avian point counts: A combined multiple-observer and time-of-detection approach. *Avian Conservation and Ecology*, 5(1), article 3.
- Stanley, T. R., & Burnham, K. P. (1999). A closure test for time-specific capture-recapture data. *Environmental and Ecological Statistics*, 6(2), 197–209.

- Stanley, T. R., & Richards, J. D. (2005). A program for testing capture-recapture data for closure. *Wildlife Society Bulletin*, 33(2), 782–785.
- Stanley, T. R., & Royle, J. A. (2005). Estimating site occupancy and abundance using indirect detection indices. *Journal of Wildlife Management*, 69(3), 874–883.
- Stauffer, S., Jucker, M., & et al. (2021). How many replicates to accurately estimate fish biodiversity using environmental DNA on coral reefs? *Ecology and Evolution*, 11(21), 14630–14643.
- Steen, V. A., Elphick, C. S., Tingley, M. W., & Albright, T. (2019). An evaluation of stringent filtering to improve species distribution models from citizen science data. *Diversity and distributions*, 25(12), 1857–1869.
- Steenweg, R., Hebblewhite, M., Whittington, J., Lukacs, P., & McKelvey, K. (2018). Sampling scales define occupancy and underlying occupancy–abundance relationships in animals. *Ecology*, 99(1), 172–183.
- Steinberg, D., & Cardell, N. S. (1992). Estimating logistic regression models when the dependent variable has no variance. *Communications in Statistics-Theory and Methods*, 21(2), 423–450.
- Steiniger, S., & Hunter, A. J. S. (2012). A scaled line-based kernel density estimator for the retrieval of utilization distributions and home ranges from GPS movement tracks. *Ecological Informatics*, 13, 1–8.
- Stephens, M. (2000). Dealing with label switching in mixture models. *Journal of the Royal Statistical Society, Series B*, 62(4), 795–809.
- Stevens, B. S., & Conway, C. J. (2019). Predicting species distributions: Unifying model selection and scale optimization for multi-scale occupancy models. *Ecosphere*, 10(5), e02748.
- Stevenson, B. C., Borchers, D. L., Altweig, R., Swift, R. J., Gillespie, D. M., & Measey, G. J. (2015). A general framework for animal density estimation from acoustic detections across a fixed microphone array. *Methods in Ecology and Evolution*, 6(1), 38–48.
- Stevenson, B. C., Borchers, D. L., & Fewster, R. M. (2019). Cluster capture-recapture to account for identification uncertainty on aerial surveys of animal populations. *Biometrics*, 75(1), 326–336.
- Stevenson, B. C., van Dam-Bates, P., Young, C. K. Y., & Measey, J. (2021). A spatial capture–recapture model to estimate call rate and population density from passive acoustic surveys. *Methods in Ecology and Evolution*, 12(3), 432–442.
- Stoklosa, J., Daly, C., Foster, S. D., Ashcroft, M. B., & Warton D. I. (2015). A climate of uncertainty: Accounting for error in climate variables for species distribution models. *Methods in Ecology and Evolution*, 6(4), 412–423.
- Stoklosa, J., & Huggins, R. M. (2012). A robust P-spline approach to closed population capture–recapture models with time dependence and heterogeneity. *Computational Statistics and Data Analysis*, 56(2), 408–417.
- Stoklosa, J., Hwang, W.-H., Wu, S.-H., & Huggins, R. (2011). Heterogeneous capture-recapture models with covariates: A partial likelihood approach for closed populations. *Biometrics*, 67(4), 1659–1665.
- Strickfaden, K. M., Fagre, D. A., & et al. (2020). Dependent double-observer method reduces false-positive errors in auditory avian survey data. *Ecological Applications*, 30(2), e02026.
- Strindberg, S., & Buckland, S. T. (2004). Zigzag survey designs in line transect sampling. *Journal of Agricultural, Biological, and Environmental Statistics*, 9(4), 443–461.
- Strindberg, S., Buckland, S. T., & Thomas, L. (2004). Design of distance sampling surveys and geographic information systems. In S. T. Buckland, D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers, & L. Thomas (Eds.), *Advanced distance sampling* (pp. 190–228). Oxford, UK: Oxford University Press.
- Su, Y.-S., & Yajima, M. (2013) R2jags: A package for running jags from R. <http://CRAN.R-project.org/package=R2jags>.
- Sugai, L. S. M., Silva, T. S. F., Ribeiro, J. W., Jr., & Llusia, D. (2019). Terrestrial passive acoustic monitoring: Review and perspectives. *Bioscience*, 69(1), 15–25.

- Sullivan, B. L., Wood, C. L., Iliff, M. J., Bonney, R. E., Fink, D., & Kelling, S. (2009). eBird: A citizen-based bird observation network in the biological sciences. *Biological Conservation*, 142(10), 2282–2292.
- Sun, C. C., Fuller, A. K., & Royle, J. A. (2014). Trap configuration and spacing influences parameter estimates in spatial capture-recapture models. *PLoS One*, 9(2), e88025.
- Sun, Y., & Fu, L. (2018). Coarse-fine-stitched: A robust maritime horizon line detection method for unmanned surface vehicle applications. *Sensors*, 18(9), 2825.
- Sutherland, C., Royle, J. A., & Linden, D. W. (2019). oSCR: A spatial capture–recapture R package for inference about spatial ecological processes. *Ecography*, 42(9), 1459–1469.
- Sutherland, C. S., Elston, D. A., & Lambin, X. (2014). A demographic, spatially explicit patch occupancy model of metapopulation dynamics and persistence. *Ecology*, 95(11), 3149–3160.
- Sutherland, J., & Schwarz, C. (2005). Multi-list methods using incomplete lists in closed populations. *Biometrics*, 61(1), 134–140.
- Takahara, T., Minamoto, T., & Doi, H. (2013). Using environmental DNA to estimate the distribution of an invasive fish species in ponds. *PLoS One*, 8(2), e56584.
- Tanadini, L. G., & Schmidt, B. R. (2011). Population size influences amphibian detection probability: Implications for biodiversity monitoring programs. *PLoS One*, 6(12), e28244.
- Tanaka, U., Ogata, Y., & Stoyan, D. (2008). Parameter estimation and model selection for Neyman–Scott point processes. *Biometrical Journal*, 50(1), 43–57.
- Tancredi, A., Auger-Méthé, M., Marcoux, M., & Liseo, B. (2013). Accounting for matching uncertainty in two stage capture-recapture experiments using photographic measurements of natural marks. *Environmental and Ecological Statistics*, 20(4), 647–665.
- Tanner, M. A., & Wong, W. H. (1987). The calculation of posterior distributions by data augmentation (with discussion). *Journal of the American Statistical Association*, 82(398), 529–550.
- Tardella, L. (2002). A new Bayesian method for nonparametric capture-recapture models in presence of heterogeneity. *Biometrika*, 89, 807–817.
- Taylor, L. R. (1961). Aggregation, variance and the mean. *Nature*, 189(4766), 732–735.
- Taylor, R. A. J. (2018). Spatial distribution, sampling efficiency and Taylor's power law. *Ecological Entomology*, 43(2), 215–225.
- Taylor, R. A. J. (2021). Spatial distribution, sampling efficiency and Taylor's power law. 2. Interpreting density-dependent sampling efficiency. *Agricultural and Forest Entomology*, 23(2), 173–188.
- Tenan, S., O'Hara, R. B., Hendriks, I., & Tavecchia, G. (2014). Bayesian model selection: The steepest mountain to climb. *Ecological Modelling*, 283, 62–69.
- Thomas, J. F., Engeman, R. M., & et al. (2013). Optimizing line intercept sampling and estimation for feral swine damage levels in ecologically sensitive wetland plant communities. *Environmental Science Pollution Research and Management*, 20(3), 1503–1510.
- Thomas, L., Buckland, S. T., & et al. (2010). Distance software: Design and analysis of distance sampling surveys for estimating population size. *Journal of Applied Ecology*, 47(1), 5–14.
- Thompson, C. J., Lee, T. E., Stone, L. M., McCarthy, M. A., & Burgman, M. A. (2013). Inferring extinction risks from sighting records. *Journal of Theoretical Biology*, 338, 16–22.
- Thompson, C. M., Royle, J. A., & Garner, J. D. (2012). A framework for inference about carnivore density from unstructured spatial sampling of scat using detector dogs. *Journal of Wildlife Management*, 76(4), 863–871.
- Thompson, F. R., III, & La Sorte, F. A. (2008). Comparison of methods for estimating bird abundance and trends from historical count data. *Journal of Wildlife Management*, 72(8), 1674–1682.
- Thompson, G. G., Withers, P. C., Pianka, E. R., & Thompson, S. A. (2003). Assessing biodiversity with species accumulation curves; inventories of small reptiles by pit-trapping in Western Australia. *Austral Ecology*, 28(4), 361–383.
- Thompson, H. R. (1956). Distribution of distance to the nth neighbor in a population of randomly distributed individuals. *Ecology*, 37(2), 391–394.

- Thompson, I. D., Ortiz, D. A., Jastrebski, C., & Corbett, D. (2006). A comparison of prism plots and modified point-distance sampling to calculate tree stem density and basal area. *Northern Journal of Applied Forestry*, 23(3), 218–222.
- Thompson, S. K. (1990). Adaptive cluster sampling. *Journal of the American Statistical Association*, 85, 1050–1059.
- Thompson, S. K. (2012). *Sampling* (3rd ed.). Hoboken, New Jersey: Wiley.
- Thompson, S. K., & Seber, G. A. F. (1994). Detectability in conventional and adaptive sampling. *Biometrics*, 50(3), 712–724.
- Thompson, S. K., & Seber, G. A. F. (1996). *Adaptive sampling*. New York: Wiley.
- Thompson, W. L. (2002). Towards reliable bird surveys: Accounting for individuals present but not detected. *Auk*, 119(1), 18–25.
- Thompson, W. L. (Ed.). (2004). *Sampling rare or elusive species: Concepts, designs, and techniques for estimating population parameters*. Washington, D. C.: Island Press.
- Thomsen, F., Laczny, M., & Piper, W. (2006). A recovery of harbour porpoises (*Phocoena phocoena*) in the southern North Sea? A case study off Eastern Frisia, Germany. *Helgoland Marine Research*, 60(3), 189–195.
- Thomsen, P. F., Kielgast, F., & et al. (2012). Monitoring endangered freshwater biodiversity using environmental DNA. *Molecular Ecology*, 21(11), 2565–2573.
- Thorson, J. T. (2019). Guidance for decisions using the Vector Autoregressive Spatio-Temporal (VAST) package in stock, ecosystem, habitat and climate assessments. *Fisheries Research*, 210, 143–161.
- Thorson, J. T., & Barnett, L. A. K. (2017). Comparing estimates of abundance trends and distribution shifts using single- and multispecies models of fishes and biogenic habitat. *ICES Journal of Marine Science*, 74(5), 1311–1321.
- Thorson, J. T., & Berkson, J. (2010). Evaluating single- and multi-species procedures to estimate time-varying catchability functional parameters. *Fisheries Research*, 101(1), 38–49.
- Thorson, J. T., Ianelli, J. N., & et al. (2016). Joint dynamic species distribution models: A tool for community ordination and spatio-temporal monitoring. *Global Ecology and Biogeography*, 25(9), 1144–1158.
- Thorson, J. T., Scheuerell, M. D., Shelton, A. O., See, K. E., Skaug, H. J., & Kristensen, K. (2015). Spatial factor analysis: A new tool for estimating joint species distributions and correlations in species range. *Methods in Ecology and Evolution*, 6(6), 627–637.
- Thorson, J. T., Shelton, A. O., Ward, E. J., & Skaug, H. J. (2015). Geostatistical delta-generalized linear mixed models improve precision for estimated abundance indices for West Coast groundfishes. *ICES Journal of Marine Science*, 72(5), 1297–1310.
- Thorson, J. T., Skaug, H. J., & et al. (2015). The importance of spatial models for estimating the strength of density dependence. *Ecology*, 96(5), 1202–1212.
- Thorson, J. T., & Ward, E. J. (2013). Accounting for space–time interactions in index standardization models. *Fisheries Research*, 147, 426–433.
- Thorson, J. T., & Ward, E. J. (2014). *Fisheries Research*, 155, 168–176.
- Thurman, L. L., Barner, A. K., Garcia, T. S., & Chestnut, T. (2019). Testing the link between species interactions and species co-occurrence in a trophic network. *Ecography*, 42(10), 1658–1670.
- Tiago, P., Ceia-Hasse, A., Marques, T. A., Capinha, C., & Pereira, H. M. (2017). Spatial distribution of citizen science casuistic observations for different taxonomic groups. *Scientific Reports*, 7(1), 12832.
- Tikhonov, G., Duan, L., & et al. (2020). Computationally efficient joint species distribution modeling of big spatial data. *Ecology*, 101(2), 1–8.
- Tingley, M. W., Nadeau, C. P., Sandor, M. E., & Baselga, A. (2020). Multi-species occupancy models as robust estimators of community richness. *Methods in Ecology and Evolution*, 11(5), 633–642.
- Tioli, S., Cagnacci, F., Stradiotto, A., Rizzoli, A., & Rizzoli, A. (2009). Edge effect on density estimates of a radiotracked population of yellow-necked mice. *Journal of Wildlife Management*, 73(2), 184–190.

- Tobler, M. W., Kéry, M., Hui, F. K. C., Guillera-Arroita, G., Knaus, P., & Sattler, T. (2019). Joint species distribution models with species correlations and imperfect detection. *Ecology*, 100(8), e02754.
- Tobler, M. W., & Powell, G. V. N. (2013). Estimating jaguar densities with camera traps: Problems with current designs and recommendations for future studies. *Biological Conservation*, 159, 109–118.
- Tobler, M. W., Zúñiga Hartley, A., Carrillo-Percastegui, S. E., Powell, G. V. N., & Lukacs, P. (2015). Spatiotemporal hierarchical modelling of species richness and occupancy using camera trap data. *Journal of Applied Ecology*, 52(2), 413–421.
- Torres-Vila, L. M., Sanchez-González, Á., Ponce-Escudero, F., Martín-Vertedor, D., & Ferrer-García, J. J. (2012). Assessing mass trapping efficiency and population density of *Cerambyx welensii* Küster by mark-recapture in dehesa open woodlands. *European Journal of Forest Research*, 131(4), 1103–1116.
- Tounkara, F., & Rivest, L.-P. (2015). Mixture regression models for closed population capture–recapture data. *Biometrics*, 71(3), 721–730.
- Tourani, M. (2022). A review of spatial capture–recapture: Ecological insights, limitations, and prospects. *Ecology and Evolution*, 12(1), e8468.
- Tourani, M., Dupont, P., Nawaz, M. A., & Bischof, R. (2020). Multiple observation processes in spatial capture–recapture models: How much do we gain? *Ecology*, 101(7), e03030.
- Tsay, P. K., & Chao, A. (2001). Population size estimation for capture–recapture models with applications to epidemiological data. *Journal of Applied Statistics*, 28(1), 25–36.
- Turek, D., Milleret, C., & et al. (2021). Efficient estimation of large-scale spatial capture–recapture models. *Ecosphere*, 12(2), e03385.
- Tyre, A. J., Tenhumberg, B., Field, S. A., Niejalke, D., Parris, K., & Possingham, H. P. (2003). Improving precision and reducing bias in biological surveys: Estimating false-negative error rates. *Ecological Applications*, 13(6), 1790–1801.
- Udevitz, M. S., & Pollock, K. P. (1991). Change-in-ratio estimators for populations with more than two subclasses. *Biometrics*, 47(4), 1531–1546.
- Udevitz, M. S., & Pollock, K. P. (1995). Using effort information with change-in-ratio data for population estimation. *Biometrics*, 51(2), 471–481.
- Urbanek, R. E., Nielsen, C. K., Preuss, T. S., & Glowacki, G. A. (2012). Comparison of aerial surveys and pellet-based distance sampling methods for estimating deer density. *Wildlife Society Bulletin*, 36(1), 100–106.
- Urbina-Barreto, I., Garnier, R., & et al. (2021). Which method for which purpose? A comparison of line intercept transect and underwater photogrammetry methods for coral reef surveys. *Frontiers in Marine Science*, 8, 636902.
- Vajas, P., Calenge, C., & et al. (2021). Catch-effort model used as a management tool in exploited populations: Wild boar as a case study. *Ecological Indicators*, 124, 107442.
- Vale, R. T. R., Fewster, R. M., Carroll, E. L., & Patenaude, N. J. (2014). Maximum likelihood estimation for model $M_{t,\alpha}$ for capture–recapture data with misidentification. *Biometrics*, 70(4), 962–971.
- Van Deventer, J., & Platts, W. S. (1989). Microcomputer software system for generating population statistics from electrofishing data—user's guide for MicroFish 3.0. In *U.S. Forest Service General Technical Report INT-254*.
- Van Poortena, B. T., Barrett, B., Walters, C. J., & Ahrens, R. N. M. (2017). Are removal-based abundance models robust to fish behavior? *Fisheries Research*, 196, 160–169.
- Van Strien, A. J., van Swaay, C. A. M., & Termaat, T. (2013). Opportunistic citizen science data of animal species produce reliable estimates of distribution trends if analysed with occupancy models. *Journal of Applied Ecology*, 50(6), 1450–1458.
- Veloz, S. D. (2009). Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-only niche models. *Journal of Biogeography*, 36(12), 2290–2299.
- Venzon, D. J., & Moolgavkar, S. H. (1988). A method for computing profile-likelihood-based confidence intervals. *Journal of the Royal Statistical Society, Series C*, 37(1), 87–94.

- Ver Hoef, J. M., & Jansen, J. K. (2007). Space-time zero-inflated count models of Harbor seals. *Environmetrics*, 18(7), 697–712.
- Vieilledent, G., Merow, C., & et al. (2019). Hierarchical Bayesian species distribution models with the hSDM R Package.
- Villa, C., & Walker, S. (2014). A cautionary note on the discrete uniform prior for the binomial N: comment. *Ecology*, 95(9), 2674–2677.
- Waddle, J. H., Dorazio, R. M., Walls, S. C., & et al. (2010). A new parameterization for estimating co-occurrence of interacting species. *Ecological Applications*, 20(5), 1467–1475.
- Wagner, T., Hansen G. J. A., & et al. (2020). Improved understanding and prediction of freshwater fish communities through the use of joint species distribution models. *Canadian Journal of Fisheries and Aquatic Sciences*, 77(9), 1540–1551.
- Wang, J.-P. Z., & Lindsay, B. G. (2005). A penalized nonparametric maximum likelihood approach to species richness estimation. *Journal of the American Statistical Association*, 100(471), 942–959.
- Wang, X., He, C. Z., & Sun, D. (2007). Bayesian population estimation for small sample capture-recapture data using noninformative priors. *Journal of Statistical Planning and Inference*, 137(4), 1099–1118.
- Wang, Y., & Yip, P. S. F. (2002). Estimation of population size for additive–multiplicative models based on continuous-time recapture experiments. *Environmetrics*, 13(8), 847–857.
- Ward, A. I., White, P. C. L., & Crutchley, C. H. (2004). Roe deer *Capreolus capreolus* behaviour affects density estimates from distance sampling surveys. *Mammal Review*, 34(4), 315–319.
- Ward, G., Hastie, T., Barry, S., Elith, J., & Leathwick, J. (2009). Presence-only data and the EM algorithm. *Biometrics*, 65(2), 554–563.
- Ward, H. G. M., Askew, P. J., Post, J. R., & Rose, K. (2013). A mechanistic understanding of hyperstability in catch per unit effort and density-dependent catchability in a multistock recreational fishery. *Canadian Journal of Fisheries and Aquatic Sciences*, 70(10), 1542–1550.
- Ward, P. (2008). Empirical estimates of historical variations in the catchability and fishing power of pelagic longline fishing gear. *Reviews in Fish Biology and Fisheries*, 18(4), 409–426.
- Warren, W. G., & Olsen, P. F. (1964). The line intercept method in forest fuel sampling. *Forest Science*, 14, 20–26.
- Warton, D. I., Blanchet, F. G., & et al. (2015). So many variables: Joint modeling in community. *Trends in Ecology and Evolution*, 30(12), 766–799.
- Warton, D. I., Renner, I. W., Ramp, D., & Roberts, D. L. (2013). Model-based control of observer bias for the analysis of presence-only data in ecology. *PLoS One*, 8(11), e79168.
- Warton, D. I., & Shepherd, L. C. (2010). Poisson point process models solve the “pseudo-absence problem” for presence-only data in ecology. *Annals of Applied Statistics*, 4(3), 1393–1402.
- Warton, D. I., Stoklosa, J., Guillera-Arroita, G., MacKenzie, D. I., Welsh, A. H., O’Hara, R. B. (2017). Graphical diagnostics for occupancy models with imperfect detection. *Methods in Ecology and Evolution*, 8(4), 408–419.
- Watanabe, S. (2013). A widely applicable Bayesian information criterion. (2013). *Journal of Machine Learning Research*, 14(1), 867–897.
- Watson, G. S. (1971). Estimating functionals of particle size distributions. *Biometrika*, 58(3), 483–490.
- Watson, R. A., & Quinn, T. J., II. (1997). Performance of transect and point count underwater visual census methods. *Ecological Modelling*, 104(1), 103–112.
- Watts, D. E., Parker, I. D., Lopez, R. R., Silvy, N. J., & Davis, D. S. (2008). Distribution and abundance of endangered Florida Key Deer on outer islands. *Journal of Wildlife Management*, 72(2), 360–366.
- Webb, M. H., Wotherspoon, S., & et al. (2014). Location matters: Using spatially explicit occupancy models to predict the distribution of the highly mobile, endangered swift parrot. *Biological Conservation*, 176, 99–108.
- Wegener, M. G., Schramm, H. L., Neal, J. W., & Gerard, P. D. (2018). Effect of fishing effort on catch rate and catchability of largemouth bass in small impoundments. *Fisheries Management and Ecology*, 25(1), 66–76.

- Wegge, P., Pokheral, C. P., & Jnawali, S. R. (2004). Effects of trapping effort and trap shyness on estimates of tiger abundance from camera trap studies. *Animal Conservation*, 7(3), 251–256.
- Wei, J. W., Lee, B. P.Y.-H., Wen, L. B., & Gofredo, S. (2016). Citizen science and the urban ecology of birds and butterflies—A systematic review. *PloS One*, 11(6), e0156425.
- Weir, L. A., & Mossman, M. J. (2005). North American Amphibian Monitoring Program (NAAMP). In M. J. Lannoo (Ed.), *Amphibian declines: Conservation status of United States species* (pp. 307–313). Berkeley: University of California Press.
- Weir, L. A., Royle, J. A., Nanjappa, P., & Jung, R. E. (2005). Modeling anuran detection and site occupancy on North American amphibian monitoring program (NAAMP) routes in Maryland. *Journal of Herpetology*, 39(4), 627–639.
- Wellendorf, S. D., Palmer, W. E., & Bromley, P. T. (2004). Estimating calling rates of northern bobwhite coveys and measuring abundance. *Journal of Wildlife Management*, 68(3), 672–682.
- Weller, T. J., Thomas, S. C., & Baldwin, J. A. (2014). Use of long-term opportunistic surveys to estimate trends in abundance of hibernating Townsend's big-eared bats. *Journal of Fish and Wildlife Management*, 5(1), 59–69.
- Welsh, A. H., Lindenmayer, D. B., & Donnelly, C. F. (2013). Fitting and interpreting occupancy models. *PLoS One*, 8(1), e52015.
- Welvaert, M., & Caley, P. (2016). Citizen surveillance for environmental monitoring: Combining the efforts of citizen science and crowdsourcing in a quantitative data framework. *SpringerPlus*, 5(1), 1–14.
- Wenger, S. J., & Freeman, M. C. (2008). Estimating species occurrence, abundance, and detection probability using zero-inflated distributions. *Ecology*, 89(10), 2953–2959.
- White, G. C., & Burnham, K. P. (1999). Program MARK: Survival estimation from populations of marked animals. *Bird Study*, 46(sup1), S120–S139.
- White, G. C., & Cooch, E. G. (2017). Population abundance estimation with heterogeneous encounter probabilities using numerical integration. *Journal of Wildlife Management*, 81(2), 368–377.
- White, J. C., Coops, N. C., Wulder, M. A., Vastaranta, M., Hilker, T., & Tompalski, P. (2016). Remote sensing technologies for enhancing forest inventories: A review. *Canadian Journal of Remote Sensing*, 42(5), 619–641.
- White, S., Faulk, E., Tzilkowski, C., Weber, A. S., Marshall, M., & Wagner, T. (2020). Predicting fish species richness and habitat relationships using Bayesian hierarchical multispecies occupancy models. *Canadian Journal of Fisheries and Aquatic Sciences*, 77(3), 602–610.
- Whitehead, H. (2009). Estimating abundance from one-dimensional passive acoustic surveys. *Journal of Wildlife Management*, 73(6), 1000–1009.
- Whittington, J., Hebblewhite, M., & Chandler, R. B. (2017). Generalized spatial mark–resight models with an application to grizzly bears. *Journal of Applied Ecology*, 55(1), 157–168.
- Whittington, J., Heuer, K., Hunt, B., Hebblewhite, M., & Lukacs, P. M. (2015). Estimating occupancy using spatially and temporally replicated snow surveys. *Animal Conservation*, 18(1), 92–101.
- Wilcox, T. M., McKelvey, K. S., & et al. (2013). Robust detection of rare species using environmental DNA: The importance of primer specificity. *PLoS One*, 8(3), e59520.
- Wileyto, E. P., Ewens, W. J., & Mullen, M. A. (1994). Markov-recapture population estimates: A tool for improving interpretation of trapping experiments. *Ecology*, 75(4), 1109–1117.
- Wileyto, E. P., Norris, J. L., Weaver, D. K., & Arbogast, T. (2000). Self-marking recapture models for estimating closed insect populations. *Journal of Agricultural, Biological, and Environmental Statistics*, 5(4), 456–474.
- Wilkinson, D. P., Golding, N., Guillera-Arroita, G., Tingley, R., McCarthy, M. A., & Freckleton, R. (2012). Defining and evaluating predictions of joint species distribution models. *Methods in Ecology and Evolution*, 12(3), 394–404.
- Wilkinson, D. P., Golding, N., Guillera-Arroita, G., Tingley, R., & McCarthy, M. A. (2019). A comparison of joint species distribution models for presence–absence data. *Methods in Ecology and Evolution*, 10(2), 198–211.

- Williams, P. J., Hooten, M. B., Womble, J. N., Bower, M. R., & Gaggiotti, O. (2017). Estimating occupancy and abundance using aerial images with imperfect detection. *Methods in Ecology and Evolution*, 8(12), 1679–1689.
- Williams, R., & Thomas, L. (2007). Distribution and abundance of marine mammals in the coastal waters of British Columbia, Canada. *Journal of Cetacean Research and Management*, 9(1), 15–28.
- Williams, R., & Thomas, L. (2009). Cost-effective abundance estimation of rare marine animals: Small-boat surveys for killer whales in British Columbia, Canada. *Biological Conservation*, 142(7), 1542–1547.
- Wilson, J. (2011). Cover plus: Ways of measuring plant canopies and the terms used for them. *Journal of Vegetation Science*, 22(2), 197–206.
- Wilson, K. R., & Anderson, D. R. (1995). Continuous-time capture-recapture population estimation when capture probabilities vary over time. *Environmental and Ecological Statistics*, 2(1), 55–69.
- Wimmer, J., Towsey, M., Roe, P., & Williamson, I. (2013). Sampling environmental acoustic recordings to determine bird species richness. *Ecological Applications*, 23(6), 1419–1428.
- Winfree, R., Fox, J. W., Williams, N. M., Reilly, J. R., & Cariveau, D. P. (2015). Abundance of common species, not species richness, drives delivery of a real-world ecosystem service. *Ecology Letters*, 18(7), 626–635.
- Wintle, B. A., Kavanagh, R. P., McCarthy, M. A., & Burgman, M. A. (2005). Estimating and dealing with detectability in occupancy surveys for forest owls and arboreal marsupials. *Journal of Wildlife Management*, 69(3), 905–917.
- Wintle, B. A., McCarthy, M. A., Parris, K. M., & Burgman, M. A. (2004). Precision and bias of methods for estimating point survey detection probabilities. *Ecological Applications*, 14(3), 703–712.
- Wintle, B. A., Walshe, T. V., Parris, K. M., & McCarthy, M. A. (2012). Designing occupancy surveys and interpreting non-detection when observations are imperfect. *Diversity and Distributions*, 18(4), 417–424.
- Wisz, M. S., Hijmans, R. J., Li, J., Peterson, A. T., Graham, C., & Guisan, A. (2008). Effects of sample size on the performance of species distribution models. *Diversity and Distributions*, 14(5), 763–773.
- Witt, M. J., Baert, B., & et al. (2009). Aerial surveying of the world's largest leatherback turtle rookery: A more effective methodology for large-scale monitoring. *Biological Conservation*, 142(8), 1719–1727.
- Wolter, K. M. (1984). An investigation of some estimators of variance for systematic sampling. *Journal of the American Statistical Association*, 79(388), 781–790.
- Wolter, K. M. (2007). *Introduction to variance estimation* (2nd ed.) New York: Springer.
- Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society, Series B*, 73(1), 3–36.
- Woodall, C. W., Walters, B. F., & Westfall, J. A. (2012). Tracking downed dead wood in forests over time: Development of a piece matching algorithm for line intercept sampling. *Forest Ecology and Management*, 277, 196–204.
- Worton, B. J. (1995). Using Monte Carlo simulation to evaluate kernel-based home range estimators. *Journal of Wildlife Management*, 59(4), 794–800.
- Wrege, P. H., Rowland, E. D., Keen, S., Shiu, Y., & Matthiopoulos, J. (2017). Acoustic monitoring for conservation in tropical forests: Examples from forest elephants. *Methods in Ecology and Evolution*, 8(10), 1292–1301.
- Wright, J. A., Barker, R. J., Schofield, M. R., Frantz, A. C., Byrom, A. E., & Gleeson, D. M. (2009). Incorporating genotype uncertainty into mark-recapture-type model for estimating animal abundance. *Biometrics*, 65(3), 833–840.
- Wright, W. J., Irvine, K. M., & Higgs, M. D. (2019). Identifying occupancy model inadequacies: Can residuals separately assess detection and presence? *Ecology*, 100(6), 1–9.

- Wyatt, R. (2002). Estimating riverine fish population size from single- and multiple-pass removal sampling using a hierarchical model. *Canadian Journal of Fisheries and Aquatic Sciences*, 59(4), 695–706.
- Xi, L., Watson, R., Wang, J.-P., & Yip, P. S. F. (2009). Estimation in capture-recapture models when covariates are subject to measurement errors and missing data. *Canadian Journal of Statistics*, 37(4), 645–658.
- Xi, L., Watson, R., & Yip, P. S. F. (2008). The minimum capture proportion for reliable estimation in capture-recapture models. *Biometrics*, 64(1), 242–249.
- Xi, L., Yip, P. S. F., & Watson, R. (2007). A unified likelihood-based approach for estimating population size in continuous-time capture–recapture experiments with frailty. *Biometrics*, 63(1), 228–236.
- Xu, C., Sun, D., & He, C. (2014). Objective Bayesian analysis for a capture-recapture model. *Annals of the Institute of Statistical Mathematics*, 66(2), 245–278.
- Xu, Y., Liu, L., You, N., Pan, H., & Yip, P. (2007). Estimating population size for a continuous time frailty model with covariates in a capture-recapture study. *Biometrics*, 63(3), 917–921.
- Yackulic, C. B., Chandler, R., & et al. (2013). Presence-only modelling using MAXENT: When can we trust the inferences? *Methods in Ecology and Evolution*, 4(3), 236–243.
- Yackulic, C. B., Reid, J., Davis, R., Hines, J. E., Nichols, J. D., & Forsman, E. (2012). Neighborhood and habitat effects on vital rates: Expansion of the Barred Owl in the Oregon Coast Ranges. *Ecology*, 93(8), 1953–1966.
- Yamaura, Y., Connor, E. F., & et al. (2016). Estimating species—area relationships by modeling abundance and frequency subject to incomplete sampling. *Ecology and Evolution*, 6(14), 4836–4848.
- Yamaura, Y., Kéry, M., & Royle, A. J. (2016). Study of biological communities subject to imperfect detection: Bias and precision of community N-mixture abundance models in small-sample situations. *Ecological Research*, 31(3), 289–305.
- Yamaura, Y., Royle, J. A., Kuboi, K., Tada, T., Ikeno, S., & Makino, S. (2011). Modelling community dynamics based on species-level abundance models from detection/nondetection data. *Journal of Applied Ecology*, 48(1), 67–75.
- Yang, D., Yang, A., Yang, J., Xu, R., & Qiu, H. (2021). Unprecedented migratory bird die-off: A citizen-based analysis on the spatiotemporal patterns of mass mortality events in the Western United States. *Geohealth*, 5(4), e2021GH000395.
- Yang, H.-C., & Chao, A. (2005). Modeling animals' behavioral response by Markov chain models for capture-recapture experiments. *Biometrics*, 61(4), 1010–1017.
- Yang, H., Magnussen, S., Fehrmann, L., Mundhenk, P., & Kleinn, C. (2016). Two neighborhood-free plot designs for adaptive sampling of forests. *Environmental and Ecological Statistics*, 23(2), 279–299.
- Ye, Y., & Darren, D. (2009). How reliable are the abundance indices derived from commercial catch-effort standardization? *Canadian Journal of Fisheries and Aquatic Sciences*, 66(7), 1169–1178.
- Yee, T., Stoklosa, J., & Huggins, M. (2015). The VGAM package for capture-recapture data using the conditional likelihood. *Journal of Statistical Software*, 65(5), 1–33.
- Yip, P. (1989). An inference procedure for a capture and recapture experiment with time-dependent capture probabilities. *Biometrics*, 45(2), 471–479.
- Yip, P. (1991). A martingale estimating equation for a capture-recapture experiment in discrete time. *Biometrics*, 47(3), 1081–1088.
- Yip, P. S. F., & Chao, A. (1996). Estimating population size from capture-recapture studies via sample coverage and estimating functions. *Communications in Statistics, Stochastic Models*, 12, 17–35.
- Yip, P. S. F., & Fong, D. Y. T. (1993). Estimating population size from a removal experiment. *Statistics and Probability Letters*, 16(2), 129–135.
- Yip, P. S. F., Lin, H.-Z., & Xi, L. (2005). A semiparametric method for estimating population size for capture-recapture experiments with random covariates in continuous time. *Biometrics*, 61(4), 1085–1092.

- Yip, P. S. F., & Wang, Y. (2002). A unified parametric regression model for recapture studies with random removals in continuous time. *Biometrics*, 58(1), 192–199.
- Yoccoz, N. G., Brathen, K. A., & et al. (2012). DNA from soil mirrors plant taxonomic and growth form diversity. *Molecular Ecology*, 21(15), 3647–3655.
- Yoshizaki, J., Brownie, C., & et al. (2011). Modeling misidentification errors that result from use of genetic tags in capture–recapture studies. *Environmental and Ecological Statistics*, 18 (1), 27–55.
- Yoshizaki, J., Pollock, K. H., Brownie, C., & Webster, R. A. (2009). Modeling misidentification errors in capture-recapture studies using photographic identification of evolving marks. *Ecology*, 90(1), 3–9.
- You, N., & Liu, L. (2006). Proportional trapping-removal model with a known ratio. *Communications in Statistics—Theory and Methods*, 35(1), 33–41.
- Yuan, Y., Bachl, F. E., & et al. (2017). Point process models for spatio-temporal distance sampling data from a large-scale survey of blue whales. *Annals of Applied Statistics*, 11(4), 2270–2297.
- Zaslavsky, A. M., & Wolfgang, G. (1993). Triple-system modeling of census, postenumeration survey, and administrative-list data. *Journal of Business and Economic Statistics*, 11(3), 279–288.
- Zeller, K. A., Nijhawan, S., Salom-Perez, R., Potosme, S. H., & Hines, J. E. (2011). Integrating occupancy modeling and interview data for corridor identification: A case study for jaguars in Nicaragua. *Biological Conservation*, 144(2), 892–901.
- Zemanova, M. A. (2020). Towards more compassionate wildlife research through the 3Rs principles: Moving from invasive to non-invasive methods. *Wildlife Biology*, 2021(1), 1–17.
- Zhang, C., Chen, Y., Xu, B., Xue, Y., & Ren, Y. (2018). Comparing the prediction of joint species distribution models with respect to characteristics of sampling data. *Ecography*, 41(11), 1876–1887.
- Zhang, W., & Bonner, S. J. (2020). On continuous-time capture-recapture in closed populations. *Biometrics*, 76(3), 1028–1033.
- Zhang, W., Bonner, S. J., & McCrea, R. S. (2022). Latent multinomial models for extended batch-mark data. *Biometrics*, 22. <https://doi.org/10.1111/biom.13789>.
- Zhang, W., Bravington, M. V., & Fewster, R. M. (2019). Fast likelihood-based inference for latent count models using the saddlepoint approximation. *Biometrics*, 75(3), 723–733.
- Zhang, W., Price, S. J., & Bonner, S. J. (2021). Maximum likelihood inference for the band-read error model for capture-recapture data with misidentification. *Environmental and Ecological Statistics*, 28(2), 405–422.
- Zhou, M., McCrea, R. S., Matechou, E., Cole, D. J., & Griffiths, R. A. (2018). Removal models accounting for temporary emigration. *Biometrics*, 75(1), 24–35.
- Zhou, S., & Griffiths, S. P. (2007). Estimating abundance from detection-nondetection data for randomly distributed or aggregated elusive populations. *Ecography*, 30(4), 537–549.
- Zhou, S., Vance, D. J., Dichmont, C. M., Burridge, C. Y., & Toscas, P. J. (2008). Estimating prawn abundance and catchability from catch-effort data: Comparison of fixed and random effects models using maximum likelihood and hierarchical Bayesian methods. *Marine and Freshwater Research*, 59(1), 1–9.
- Zhu, X. Z., Gao, T., & Zhang, J. T. (2014). Point-centred quadrangle method: A novel distance method for density estimation. *Advanced Materials Research*, 1073–1076, 479–483.
- Zipkin, E. F., Dewan, A., & Royle, J. A. (2009). Impact of forest fragmentation on species richness: A hierarchical approach to community modeling. *Journal of Applied Ecology*, 46(4), 815–822.
- Zipkin, E. F., Rossman, S., & et al. (2017). Integrating count and detection-nondetection data to model population dynamics. *Ecology*, 98(6), 1640–1650.
- Zipkin, E. F., Royle, J. A., Dawson, D. K., & Bates, S. (2010). Multi-species occurrence models to evaluate the effects of conservation and management actions. *Biological Conservation*, 143(2), 479–484.
- Zipkin, E. F., & Saunders, S. P. (2018). Synthesizing multiple data types for biological conservation using integrated population models. *Biological Conservation*, 217, 240–250.

- Zippin, C. (1956). An evaluation of the removal method of estimating animal populations. *Biometrics*, 12(2), 163–189.
- Zippin, C. (1958). The removal method of population estimation. *Journal of Wildlife Management*, 22(1), 82–90.
- Zucchini, W., & Channing, A. (1986). Bayesian estimation of animal abundance in small populations using capture-recapture information. *South African Journal of Science*, 8, 137–140.
- Zulian, V., Miller, D. A. W., Ferraz, G., & Jung, M. (2021). Integrating citizen-science and planned-survey data improves species distribution estimates. *Diversity & Distributions*, 27(12), 2498–2509.
- Zurell, D., Franklin, J., & et al. (2020a). A standard protocol for reporting species distribution models. *Ecography*, 43(9), 1261–1277.
- Zurell, D., Pollock, L. J., & Thuiller, W. (2018). Do joint species distribution models reliably detect interspecific interactions from co-occurrence data in homogenous environments? *Ecography*, 41(11), 1812–1819.
- Zurell, D., Zimmermann, N. E., Gross, H., Baltensweiler, A., Sattler, T., & Wiëst, R. O. (2020b). Testing species assemblage predictions from stacked and joint species distribution models. *Journal of Biogeography*, 47(1), 101–113.

Index

A

- Acoustic methods, 325
 - spatial capture-recapture, 597
- Adaptive sampling, 36
 - shipboard method, 365
- Aerial censusing, 346
 - double observers, 282
 - examples with multiple observers, 349
 - removal data and, 401
- Animal movement, 360
- Animal signs, 61, 84
 - line transects and, 318
- Apparent contagion model
 - line transect, 302
- Autocorrelation, 85, 142
 - spatial, 163, 176
- Autonomous recording unit (ARU), 325

B

- Bayesian capture-recapture
 - M_h , 564
 - M_{tb} , 555
 - M_{thb} , 557
 - M_t , Poisson model, 549
 - two-sample optimal stopping rule, 547
 - two samples, 545
- Bayesian estimation of abundance, 563, 565
- Bayesian graphical models, 554
- Bayesian heterogeneity, 556
- Bayesian history, 543
- Bayesian inference
 - advantages of, 4
 - credible intervals, 4

Bayesian list dependence, 553

- Bayesian methods
 - distance sampling, 366
- Bayesian mixture model, 558
- Bayesian nonparametric methods, 562
- Bayesian sequential updating, 551
- Bayesian species distribution models, 551
- Bayesian species richness, 551
- Bayes rule, 544
- Binomial catchability coefficient, 439
- Biodiversity, 227
- Biodiversity measures, 228
- Bootstrap methods, 13
- Buffer zone, 261

C

- Capture-recapture
 - assumptions, 457
 - coverage models, 509
 - dependence and, 344
 - Dorazio and Royle's model, 504
 - epidemiological populations, 463
 - general multinomial model, 476
 - groups, 476
 - Huggins' models, 496
 - hypergeometric distribution, 456
 - identifiability, 504
 - list of models, 475
 - log-linear models, 506, 508
 - M_0 , 483
 - M_{bh} , 493
 - M_b , 484
 - M_h , 490

- M_{tbh}, 495
- M_{tb}, 485
- M_{th}, 492
 - martingales for, 515
 - misidentification, 521
 - mixture models, 497
 - Norris and Pollock's models, 501
 - one at a time, 523
 - Pledger's models, 497
 - Pollock and Otto's models, 501
 - Pollock's models, 478
 - spatial acoustic arrays, 597
 - stratified sampling, 470
 - test of model assumptions, 497
 - time-to-detection models, 518
 - Tounkar and Rivest's model, 505
 - triple catch method, 472
 - two-sample incomplete detection, 471
 - two samples, 456
- Catchability coefficient defined, 439
- Catch-effort methods, 439
 - estimating functions and, 441
 - fisheries, 452
 - regression models, 445
 - underlying assumptions, 448
- Catch-effort models
 - Bayesian, 449
 - extinction, 450
- Catch-effort models listed, 440
- Catch per unit effort (CPUE), 401, 430, 440
- Center of activity, 28
- Change in ratio (CIR), 418
- Circular plots, 45
- CIR methods
 - binomial assumptions, 420
 - continuous sampling, 426
 - exploitation rate, 425
 - known sampling effort, 430
 - multiple removals and samples, 427
 - removals estimated, 423
 - subclasses, removals and samples, 430
 - subsampling, 426
 - survival rates, 422
 - two classes, 418
- Citizen science, 144
 - disasters, 147
 - multiple species and, 235
 - presence-only data, 149
 - pros and cons of, 145
- Closed population, 1
- Closest individual distance, 237
 - nonrandom distributions and, 242
- Closure test, 13
 - between primary units, 95
- occupancy, 94
- Royle's model, 55
- Cluster sampling, 268, 311
 - aerial censusing, 346
 - capture-recapture, 497
 - capture-recapture and, 476
 - line transects and, 328
 - spatial models, 328
- Combined methods
 - point-distance sampling, 289
- Combine proposal, 561
- Combining detection types, 118
- Common protocols
 - distance sampling, 256
- Community species, 177
- Components of detectability, 283
- Computer programs
 - AUTOMATCH, 464
 - CAPWIRE, 523
 - CloseTest, 497
 - closure.test (secr), 497
 - CRAN, 191
 - DISTANCE, 277
 - E-SURGE, 157
 - home range, 30
 - JAGS, 109, 191, 199
 - joint species distributions, 190
 - MAXENT, 77, 115, 127, 193, 231
 - multivariate abundance data, 191
 - occupancy, 73
 - Presence, 121
 - presence-absence data, 135
 - R package mra, 478
 - runjags, 13
 - selection of, 12
 - SESAM, 7
 - Spade* R, species richness, 223
 - spatial capture-recapture, 593
 - species distributions, 173
 - for species models, 176
 - species richness, 203, 208
 - SURVIV, 291
 - VAST for fisheries, 168
 - vector generalized additive model, 497
 - web simulation, 271
- Computing, 11
 - distance methods, 345
 - F.huggins.estim, 478
 - WINBUGS, 155
- Conditional adaptive design, 105
- Conditionally autoregressive (CAR) model, 81, 85, 87, 167
- Conditional multinomial, 477
- Conditional multinomial model, 507

- Confusion matrix, 201
Consumer resource model, 199
Continuous sampling
 change in ratio methods, 426
Continuous time
 submodel list, 527
Continuous-time models, 525
 frailty method, 526
 Markov method, 526
 random removals, 535
 Schofield et al., 529
 self-correcting processes, 535
Continuous time-removal model, 277
Coral reef surveys, 373
Correlation, 181
 track surveys, 80
Cost function, 100
Covariates, 76
Coverage, 321, 375
 catch-effort and, 444
 species richness and, 214
Coverage and Pollock's models, 511
Coverage models
 capture-recapture, 509
Covered area, 258, 295
CPUE methods, 439
Credible intervals, 544
- D**
- Data augmentation, 86, 160, 172, 188, 207, 401, 476, 491, 564, 605, 607, 611, 617, 625
Dependence
 double observers, 281
 dual record and, 513
 Markovian, 156
 measure of, 460
 nearby, 87
 sample, 469
 spatial, 140, 188
 species and, 179, 191
 species misidentification, 235
Dependence and capture-recapture, 344
Dependent double observers, 343
Dependent replicates, 65
Design-based methods, 2
Design considerations
 general, 13
 line transects, 313
 occupancy, 99
 plot sampling, 18
 point distance sampling, 261
 removal method, 409
- Detection function
 definition, 256
 properties of, 306
 types of, 264, 306, 307
Detection methods, 55
Detection probability, 39
Diagnostic checks
 line transects, 343
Diagnostics, 6–8, 10, 32, 72, 73, 238, 310, 340, 391, 394
Disease mapping, 161
Disease modeling
 invasive spread dynamics, 164
 multiple data sources, 163
 multiple tissues, 163
 occupancy and, 161
 underlying problems, 161
DISTANCE
 goodness-of-fit tests, 310
DNA
 one at a time sampling, 523
DNA methods, 109, 123, 520
 camera methods and, 635
Double observers
 dependent, 179
 theory for, 278
 unreconciled method, 281
 using distances, 334
Double sampling, 47, 99
Dual platform model, 337
Dual record system
 coverage and, 513
 description, 468
- E**
- Ecological distance, 600
Ecological niche factor analysis, 200
Edge effect, 3, 27
 line intercept method, 381, 386
 nearest object, 238
 point-distance sampling, 273
Effective detection area, 259
Effective strip width, 298
Electrofishing, 406
 removal method, 402
Elusive aggregated populations, 111
Empirical Bayes, 4
Encounter rate, 301
Environment
 multiple species and, 174
 species richness and, 231
Environmental conditions, 130
Environmental DNA (eDNA), 109, 124

- species richness, 226
- Environmental effects, 195
- Epidemiological populations
 - assumptions, 464
 - diabetes, 465
 - drug users, 463, 468
 - lists used, 464
 - several lists, 468
 - two lists, 463
 - vision impairment, 463
- Eradication programs, 450
- Error detection, 558
- Error detection, Bayesian, 549
- Estimating equations, 515

- F**
- Fairfield Smith's Variance Law, 35
- False negative
 - definition, 57
- False positives, 114
 - definition, 57
 - species misidentification, 234
- Fisheries
 - catch-per-unit effort, 452
- Fisheries and marine environments, 167
- Flushing animals
 - line intercept method, 379
- Flushing distance, 296
- Flushing region, 320
- Frailty models
 - continuous time, 527

- G**
- Genotyping error, 523
- Geographic distribution, 129
- Gibbs sampler, 552, 555, 558, 559
- Graphical models, 554
- Group sampling, 311

- H**
- Hazard rate
 - detection function, 308
- Hierarchical modeling, 5
- Highest posterior density region, 544
- Home range, 28
- Horvitz-Thompson-like estimator, 274, 287, 340, 536
 - abundance, 323
 - line transect, 314, 348
 - spatial telemetry, 600
 - tree sampling, 46

- I**
- Independence
 - levels of, 340
- Index-removal method, 429
- Invasive species, 67, 124, 193

- J**
- Joint species' distribution model (JSDM), 179
 - comparison of models, 189

- K**
- Key function, 264
 - series use, 310

- L**
- Latent class models, 561
- LiDAR scanner, 46, 245
- Limiting independence, 342
- Lincoln-Petersen estimate, 456
- Line point intercept, 387
- Line transects
 - assumptions, 296
 - distance theory, 298
 - mark-recapture and, 335
 - moving observer, 309
 - one- and two-sided surveys, 337
 - radial distances, 319
 - stratification and, 297
 - three dimensional, 368
- List dependence, 515
 - Bayesian, 553
- Lists, 463
- Log-linear model
 - line transect and, 330

- M**
- Marine mammals
 - line transects and, 299
- Markov chain Monte Carlo, 552
- Markov network, 194
- Mark-recapture distance sampling (MRDS), 335, 338
 - heterogeneity and, 338
- Martingales, 515
 - capture-recapture models, 516
 - using coverage, 516

- Matching methods, 464
Measurement errors, 262
Minimum capture proportion, 478, 491
Mirage point, 240
Misidentification, 521
Mixture models, 5
 Bimix, 5
Model averaging, 10
 Bayesian, 11
Model-based methods, 2
Model selection criteria, 8
 AIC, 8
 AICc, 9
 BIC, 9
 DIC, 9
 focussed information, FIC, 495
 WAIC, 9
Moving observer, 309
Multi-catch traps, 595
Multipass electrofishing, 404
Multiple data methods, 165
Multiple detection devices, 122
Multi-scale models, 137
Multi-species models, 141
Multi-state methods, 152
Multivariate species models, 179
- N**
Nearest neighbor distance, 237
Negative-binomial distribution, 33, 414
 closest distance model, 241
n-tree distance-sampling (NTDS), 246
- O**
Observation errors, 112
Occupancy, 57
 abundance and, 107
 accounting for false positives, 146
 assumptions, 92
 description, 57
 disease modeling, 161
 fossils and, 57
 interacting species, 184
 model checking, 71, 93
 model for, 59
 multiple species, 184, 229
 multi-state methods, 152
 related to habitat, 157
 two-stage design, 101
 undetected states, 154
Occupancy differences, 89
Occupancydesign considerations, 99
- Omnibus removal methods, 434
One at a time sampling, 523
One removal, 408
Open population, 1
Opportunistic data
 optimal design, 151
Opportunistic surveys, 144
 multiple species and, 235
Optimal design, 100
 Bayesian, 101
Optimal estimating equations, 516
Optimum allocation
 two-class change in ratio, 424
Ordinal data, 183, 193
 examples, 175
Overview, Data, Model, Assessment, and
 Prediction (ODMAP), 173
- P**
Partial repeat sampling, 50
Particle density, 378
Passive acoustic arrays, 326, 356
Passive detector array, 595
Pearson goodness-of-fit test, 72
Penalized likelihood, 88
Petersen estimate, 456
Plants, 67, 91, 201
 aquatic, 79
 coverage, 321, 375
 crossed design, 321
 edge effect, 18
 particle density, 378
 variable plot size, 77
Plot sampling, 17
 design of, 18
 optimal design, 25
Point counts
 definition, 255
 removal method combined with, 268
Point-distance sampling, 255, 256
 combined methods, 289
 design of, 261, 293
 heterogeneity and, 259
 traps or lures, 272
 validation of, 288
Point independence, 280, 340
Point or line transects, 332
Point sampling along lines, 265
Poisson catchability coefficient, 439
Poisson dispersion test, 32
Poisson point process, 128
 inhomogeneous, 135
Pollock's models

loglinear form, 508
 Pooling robustness, 305
 Population distributions, 32
 Post enumeration survey, 469
 Posterior distribution, 4, 544
 Presence/absence data
 occupancy and, 108
 and presence only data, 135
 transects and, 327
 Presence only data, 127, 135
 Primary and secondary units, 22, 82
 theory for, 22
 Prior detection effects, 519
 Pseudo-absences, 127
 species and, 200

R

Radial distance
 line transect and, 319
 Random catchability, 397
 Rarefaction, 203, 216
 completeness, 216
 Rare species, 63
 adaptive methods for, 103
 Rasch model, 555, 556
 Recolonization, 164
 Regression methods, 6
 Regression models
 catch-effort and, 445
 Relative density, 30
 index methods, 30
 Removal methods, 389
 aerial sightings and, 401
 Bayesian model, 397
 change in ratio, 418
 constant probability, 390
 design of, 409
 effort information and, 399
 electrofishing and, 402
 with an initial mark release, 393
 one removal, 408
 point counts and, 268
 point-count surveys, 414
 random catchability, 397
 regression models, 395
 subpopulations, 410
 three removals, 404
 trap reduction, 392
 two removals, 405
 variable catchability and, 396
 Removals estimated
 change in ratio methods, 423
 Residuals, 6, 7, 73

Resource selection, 44
 Reversible jump MCMC (RJMCMC), 555
 Reviews, 62
 acoustic methods, 325
 citizen science programs, 148
 closest objects, 238
 drug users, 463
 joint species methods, 180
 marine populations, 168
 occupancy, 62
 spatial capture-recapture, 592
 species distributions, 173
 species richness, 203, 208
 vision impairment, 463
 RJMCMC algorithm, 554
 RN model, 63, 107
 Robust design, 82
 temporary immigration and, 232

S

SCR model, 591
 Self-correcting processes, 535
 Shipboard acoustic surveys, 356
 Shipboard and aerial surveys combined, 358
 Simulation methods, 7
 Spatial capture-recapture (SCR), 591
 basic model, 593
 description, 592
 DNA and camera methods, 635
 telemetry and, 600
 Spatially replicated counts, 53
 general model for, 277
 Spatial mark-recapture with unmarked, 617
 Spatial mark-resight models, 614
 Spatial models
 distance sampling and, 362
 Dorazio's method, 143
 for multiple species, 190
 occupancy, 84
 Spatial partial identity model, 625
 Species accumulation curve, 203
 Species accumulation curve, 216
 Species-area relationship (SAR), 232
 Species correlations, 191
 Species' counts, 197
 Species distribution models (SDM), 173
 Species extinction, 450
 Species misidentification, 234
 Species ranking, 220
 Species richness, 171, 203
 for abundance data, 209
 adding new species, 221
 Bayesian, 551

- Chao1 estimators, 212
Chao2 estimators, 214
coverage and, 214
coverage estimators, 217
eDNA and, 226
for incidence data, 212, 214
nonparametric methods, 212
parametric methods, 212
sample size, 223
- Split proposal, 561
- Staggered arrival and departure times, 96
- State-space model
 dynamic, 156
 occupancy and, 115, 156
 species interactions, 199
- Stratified sampling, 23
 capture-recapture, 470
 theory for, 23
- Striplet estimator, 303
- Study design, 13
- Subpopulations
 removal methods for, 410
- Survival rates
 change in ratio methods, 422
- Systematic design, 303
- Systematic sampling, 22, 303
- two observers and, 462
- Track surveys
 with correlation, 80
- Transects in plots, 313
- Transects of random length, 316, 382
- Trapping web, 271
- Tree sampling, 21
 closest distance problems, 252
 circular plot, 45
 distance estimators, 248
 distance methods, 244, 246
 horizontal point sampling, HPS, 245
 Mirage point, 252
 plots and, 244
 prism use, 245
 sector method, 246
- Triple-catch method, 472
- Triple dependent record system, 473
- Triple record system, 473
- Two interacting species, 184
- Two interactive species, 193
- Two-phase adaptive design, 103
- Two removals, 405
- Two-stage model
 opportunistic data, 144
- Two-step method, 98
- T**
- Tag loss, 465
- Taylor's Power law, 35
- Telemetry
 distance sampling, 361
 integrated spatial data and, 601
 mark-resight and, 626
 occupancy and, 62, 166
 point transect and, 288
 spatial capture-recapture, 634
 spatial capture-recapture and, 600
 spatial mark-recapture, 615
- Temporary emigration
 species richness and, 231
- Terrestrial laser scanner, 45
- Three levels of independence, 342
- Three removals, 404
- Time-to-detection (TTD) models, 91, 201
 capture-recapture, 518
 multiple species and, 201
 removal methods, 416
- U**
- Underwater visual census, 360
- V**
- Variable catchability, 396, 458
Variable catchability models, 449
- Vocal activity rate index, 326
- W**
- Wildfire
 aftereffects, 82
- Z**
- Zero-inflated, 172
Zero-inflated models, 141
Zero-inflated Poisson (ZIP), 6, 142
Zigzag designs, 313, 315