

1

2

Spatial Capture-Recapture

3

The Four Horsemen (and women)

4

USGS Patuxent Wildlife Research Center
North Carolina State University

5

6

ACADEMIC PRESS

CONTENTS

9	1 Introduction	9
10	2 Introduction to Bayesian Analysis of GL(M)Ms Using R/WinBUGS	11
11		
12	2.1 Notation	12
13	2.2 GLMs and GLMMs	13
14	2.3 Bayesian Analysis	14
15	2.3.1 Bayes Rule	15
16	2.3.2 Bayesian Inference	16
17	2.3.3 Prior distributions	18
18	2.3.4 Posterior Inference	18
19	2.3.5 Small sample inference	19
20	2.4 Characterizing posterior distributions by MCMC simulation	20
21	2.5 What Goes on Under the MCMC Hood	21
22	2.5.1 Rules for constructing full conditional distributions	22
23	2.5.2 Metropolis-Hastings algorithm	23
24	2.6 Practical Bayesian Analysis and MCMC	24
25	2.6.1 Choice of prior distributions	24
26	2.6.2 Convergence and so-forth	25
27	2.6.3 Bayesian confidence intervals	26
28	2.6.4 Estimating functions of parameters	27
29	2.7 Bayesian Analysis using WinBUGS	27
30	2.7.1 Linear Regression in WinBUGS	27
31	2.7.2 Inference about functions of model parameters	30
32	2.8 Model Checking and Selection	30
33	2.8.1 Goodness-of-fit	31
34	2.8.2 Model Selection	32
35	2.9 Poisson GLMs	33
36	2.9.1 Important properties of the Poisson distribution	34
37	2.9.2 Example: Breeding Bird Survey Data	34
38	2.9.3 Doing it in WinBUGS	35
39	2.9.4 Constructing your own MCMC algorithm	38
40	2.10 Poisson GLM with Random Effects	43

41	2.11	Binomial GLMs	45
42	2.11.1	Binomial regression	45
43	2.11.2	Example: Waterfowl Banding Data	46
44	2.12	Summary and Outlook	48
45	3	Estimating the Size of a Closed Population	51
46	3.1	The Simplest Closed Population Model: Model M0	52
47	3.1.1	The Spatial Context of Capture-Recapture	54
48	3.1.2	Conditional likelihood	54
49	3.2	Data Augmentation	55
50	3.2.1	DA links occupancy models and closed population models . .	56
51	3.2.2	Model M_0 in BUGS	57
52	3.2.3	Formal development of data augmentation	59
53	3.2.4	Remarks on Data Augmentation	60
54	3.2.5	Example: Black Bear Study on Fort Drum	61
55	3.3	Temporally varying and behavioral effects	65
56	3.4	Models with individual heterogeneity	66
57	3.4.1	Analysis of Model Mh	69
58	3.4.2	Analysis of the Fort Drum data	70
59	3.4.3	Building your own MCMC algorithm	72
60	3.4.4	Exercises related to model Mh	76
61	3.5	Individual Covariate Models: Toward Spatial Capture-Recapture . .	76
62	3.5.1	Example: Location of capture as a covariate.	78
63	3.5.2	Extension of the Model	81
64	3.5.3	Invariance of density to maxD	84
65	3.5.4	Toward Fully Spatial Capture-recapture Models	84
66	3.6	DISTANCE SAMPLING: A primitive Spatial Capture-Recapture	
67		Model	85
68	3.6.1	Example: Muntjac deer survey from Nagarahole, India . . .	87
69	3.7	Summary and Outlook	88
70	4	Fully Spatial Capture-Recapture Models	91
71	4.1	Sampling Design and Data Structure	92
72	4.2	The binomial observation model	93
73	4.2.1	Distance as a latent variable	94
74	4.3	The Binomial Point-process Model	95
75	4.3.1	Definition of home range center	96
76	4.3.2	The state-space of the point process	97
77	4.3.3	Invariance and the State-space as a model assumption	98
78	4.3.4	Connection to Model Mh	99
79	4.3.5	Connection to Distance Sampling	100
80	4.4	Simulating SCR Data	100
81	4.4.1	Formatting and manipulating real data sets	102

82	4.5	Fitting an SCR Model in BUGS	102
83	4.6	Unknown N	105
84	4.6.1	Analysis using data augmentation in WinBUGS	106
85	4.6.2	Use of other BUGS engines: JAGS	109
86	4.7	Case Study: Wolverine Camera Trapping Study	110
87	4.7.1	Fitting the model in WinBUGS	113
88	4.7.2	Conclusion of Analysis	116
89	4.8	Constructing Density Maps	117
90	4.8.1	Example: Wolverine density map.	119
91	4.9	Discrete State-Space	120
92	4.9.1	Evaluation of Coarseness of Discrete Approximation	122
93	4.9.2	Analysis of the wolverine camera trapping data	123
94	4.9.3	SCR models as multi-state models	126
95	4.10	Summary and Outlook	127
96	5	Other observation models	131
97	6	Likelihood Analysis of Spatial Capture-Recapture Models	133
98	6.1	Likelihood analysis	134
99	6.1.1	Implementation (simulated data)	136
100	6.2	MLE when N is Unknown	139
101	6.2.1	Exercises	142
102	6.2.2	Integrated Likelihood using the model under data augmenta- tion	143
103	6.2.3	Extensions	143
104	6.3	Classical model selection and assessment	144
105	6.4	Likelihood analysis of the wolverine camera trapping data	144
106	6.4.1	Restricted state-space	147
107	6.4.2	Exercises	147
108	6.5	Program DENSITY and the R package secr	147
109	6.5.1	Analysis using the secr package	149
110	6.6	Summary and Outlook	151
111	7	MCMC for Spatial Capture-Recapture	153
112	7.1	Introduction	153
113	7.1.1	Why build your own MCMC algorithm?	153
114	7.2	MCMC and posterior distributions	154
115	7.3	Types of MCMC sampling	157
116	7.3.1	Gibbs sampling	157
117	7.3.2	Metropolis-Hastings sampling	160
118	7.3.3	Metropolis-within-Gibbs	164
119	7.4	GLMMs Poisson regression with a random effect	164
120	7.4.1	Rejection sampling and slice sampling	168
121			

122	7.5	MCMC for closed capture-recapture Model Mh	168
123	7.6	MCMC algorithm for the basic spatial capture-recapture model . . .	169
124	7.6.1	SCR model with binomial encounter process	172
125	7.6.2	Looking at model output	173
126	7.6.3	Posterior density plots	176
127	7.6.4	Serial autocorrelation and effective sample size	176
128	7.6.5	Summary results	178
129	7.6.6	Other useful commands	180
130	7.7	Manipulating the state-space	180
131	7.8	MCMC software packages	183
132	7.8.1	WinBUGS	183
133	7.8.2	OpenBUGS	183
134	7.8.3	JAGS Just Another Gibbs Sampler	184
135	7.9	Summary and Outlook	185
136	8	Goodness of Fit and stuff	187
137	9	Covariate models	189
138	10	State-space Covariates	191
139	10.1	Homogeneous point process revisited	192
140	10.2	Inhomogeneous binomial point process	193
141	10.3	Examples	195
142	10.3.1	Simulation and analysis of inhomogeneous point processes . .	195
143	10.3.2	Fitting inhomogeneous point process SCR models	198
144	10.3.3	The jaguar data	202
145	10.4	Summary	204
146	11	Open models	207
147	12	Spatial Capture-Recapture for Unmarked Populations	209
148	12.1	Existing Models for Inference About Density in Unmarked Populations	210
149	12.2	Spatial Correlation as Information	211
150	12.3	Data Requirements and Survey Designs	212
151	12.4	Encounter Histories as Latent Variables	213
152	12.5	Estimation by MCMC	213
153	12.6	Northern Parula Example	214
154	12.7	On (Im)precision	216
155	12.8	How Much Correlation Is Enough?	216
156	12.9	Mutants	216
157	12.9.1	Other observation models	216
158	12.9.2	Linear designs	216
159	12.10	Summary	216

		7
	<hr/>	
160	12.10.1 Alternative Observation Models	219
161	12.10.2 Spatial point process models	219
162	12.11 Conclusion	220
163	Bibliography	221

164
165
166

1

INTRODUCTION

2

INTRODUCTION TO BAYESIAN ANALYSIS OF GL(M)MS USING R/WINBUGS

A major theme of this book is that spatial capture-recapture models are, for the most part, just generalized linear models (GLMs) wherein the covariate, distance between trap and home range center, is partially or fully unobserved – and therefore regarded as a random effect. Such models are usually referred to as Generalized Linear Mixed Models (GLMMs) and, therefore, SCR models can be thought of as a specialized type of GLMM. Naturally then, we should consider analysis of these slightly simpler models in order to gain some experience and, hopefully, develop a better understanding of spatial capture-recapture models.

In this chapter, we consider classes of GLM models - Poisson and binomial (i.e., logistic regression) GLMs - that will prove to be enormously useful in the analysis of capture-recapture models of all kinds. Many readers are probably familiar with these models because they represent probably the most generally useful models in all of Ecology and, as such, have received considerable attention in many introductory and advanced texts. We focus on them here in order to introduce the readers to the analysis of such models in **R** and **WinBUGS**, which we will translate directly to the analysis of SCR models in subsequent chapters.

Bayesian analysis is convenient for analyzing GLMMs because it allows us to work directly with the conditional model – i.e., the model that is conditional on the random effects, using computational methods known as Markov chain Monte Carlo (MCMC). Learning how to do Bayesian analysis of GLMs and GLMMs in **WinBUGS** is, in part, the purpose of this chapter. While we use **WinBUGS** to do the Bayesian computations, we organize and summarize our data and execute

WinBUGS from within **R** using the useful package **R2WinBUGS** (Sturtz et al., 2005). Kéry (2010), and Kéry and Schaub (2011) provide excellent introductions to the basics of Bayesian analysis and GLMs at an accessible level. We don't want to be too redundant with those books and so we avoid a detailed treatment of Bayesian methodology - instead just providing a cursory overview so that we can move on and attack the problems we're most interested in related to spatial capture-recapture. In addition, there are a number of texts that provide general introductions to Bayesian analysis, MCMC, and their applications in Ecology including McCarthy (2007), Kéry (2010), Link and Barker (2009), and King (2009).

While this chapter is about Bayesian analysis of GLMMs, such models are routinely analyzed using likelihood methods too, as discussed by Royle and Dorazio (2008), and Kéry (2010). Indeed, likelihood analysis of such models is the primary focus of many applied statistics texts, a good one being Zuur et al. (2009). Later in this book, we will use likelihood methods to analyze SCR models but, for now, we concentrate on providing a basic introduction to Bayesian analysis because that is the approach we will use in a majority of cases in later chapters.

2.1 NOTATION

We will sometimes use conventional “bracket notation” to refer to probability distributions. If y is a random variable the $[y]$ indicates its distribution or its probability density/mass function (pdf, pmf) depending on context. If x is another random variable then $[y|x]$ is the conditional distribution of y given x , and $[y, x]$ is the joint distribution of y and x . To differentiate specific distributions in some contexts we might label them $g(y)$, $g(y|\theta)$, $f(x)$, or similar. We will also write $y \sim \text{Normal}(\mu, \sigma^2)$ to indicate that y “is distributed as” a normal random variable with parameters μ and σ^2 . The expected value or mean of a random variable is $E[y] = \mu$, and $\text{Var}[y] = \sigma^2$ is the variance of y . To indicate specific observations we'll use an index such as “ i ”. So, y_i for $i = 1, 2, \dots, n$ indicates observations for n individuals. Finally, we write $\text{Pr}(y)$ to indicate specific probabilities, i.e., of events “ y ” or similar.

To illustrate these concepts and notation, suppose z is a binary outcome (e.g., species occurrence) and we might assume the model: $z \sim \text{Bern}(p)$ for observations. Under this model $\text{Pr}(z = 1) = \psi$, which is also the expected value $E[z] = \psi$. The variance is $\text{Var}[z] = \psi * (1 - \psi)$ and the probability mass function (pmf) is $[z] = \psi^z (1 - \psi)^{1-z}$. Sometimes we write $[z|\psi]$ when it is important to emphasize the conditional dependence of z on ψ . As another example, suppose y is a random variable denoting whether or not a species is detected if an occupied site is surveyed. In this case it might be natural to express the pmf of the observations y *conditional* on z . That is, $[y|z]$. In this case, $[y|z = 1]$ is the conditional pmf of y given that a site is occupied, and it is natural to assume that $[y|z = 1] = \text{Bern}(p)$ where p is the “detection probability” - the probability that we detect the species, given that it is present. The model for the observations y is completely specified once we describe

the other conditional pmf $[y|z = 0]$. For this conditional distribution it is sometimes reasonable to assume $\Pr(y = 1|z = 0) = 0$ (MacKenzie et al. (2002); see also Royle and Link (2006)). That is, if the species is absent, the probability of detection is 0. This implies that $\Pr(y = 0|z = 0) = 1$. To allow for situations in which the true state z is unobserved, we assume that $[z]$ is Bernoulli with parameter ψ . In this case, the marginal distribution of y is

$$[y] = [y|z = 1]Pr(z = 1) + [y|z = 0]Pr(z = 0)$$

because $[y|z = 0]$ is a point mass at $y = 0$, by assumption, then

$$\Pr(y = 1) = p\psi$$

And

$$\Pr(y = 0) = (1 - p) * \psi + (1 - \psi)$$

2.2 GLMS AND GLMMS

We have asserted already that SCR models work out most of the time to be variations of GLMs and GLMMS. Some of you might therefore ask: What are GLMs and GLMMS, anyhow? These models are covered extensively in many very good applied statistics books and we refer the reader elsewhere for a detailed introduction. We think Kéry (2010), Kéry and Schaub (2011), and Zuur et al. (2009) are all accessible treatments of considerable merit. Here, we'll give the 1 minute treatment of GLMMS, not trying to be complete but rather only to preserve a coherent organization to the book.

The generalized linear model (GLM) is an extension of standard linear models by allowing the response variable to have some distribution from the exponential family of distributions (i.e., not just normal). This includes the normal distribution but also dozens of others such as the Poisson, binomial, gamma, exponential, and many more. In addition, GLMS allow the response variable to be related to the predictor variables (i.e., covariates) using a link function, which is usually nonlinear. Finally, GLMs typically accommodate a relationship between the mean and variance. The classical reference for GLMs is Nelder and Wedderburn (1972) and also McCullagh and Nelder (1989). The GLM consists of three components:

1. A probability distribution for the dependent variable y , from a class of probability distributions known as the exponential family.
2. A "linear predictor" $\eta = \mathbf{X}\beta$.
3. A link function g that relates $E[y]$ to the linear predictor, $E[y] = \mu = g^{-1}(\eta)$. Therefore $g(E[y]) = \eta$.

The dependent variable y is assumed to be an outcome from a distribution of the exponential family which includes many common distributions including the

normal, gamma, Poisson, binomial, and many others. The mean of the distribution of y is assumed to depend on predictor variables x according to

$$g(E[y]) = \mathbf{x}'\beta$$

where $E[y]$ is the expected value of y , and $\mathbf{x}'\beta$ is termed the *linear predictor*, i.e., a linear function of the predictor variables with unknown parameters β to be estimated. The function g is the link function. In standard GLMs, the variance of y is a function V of the mean of y : $Var(y) = V(\mu)$ (see below for examples).

A Poisson GLM posits that $y \sim \text{Poisson}(\lambda)$ with $E[y] = \lambda$ and usually the model for the mean is specified using the *log link function* by

$$\log(\lambda_i) = \beta_0 + \beta_1 * x_i$$

The variance function is $V(y_i) = \lambda_i$. The binomial GLM posits that $y_i \sim \text{Binomial}(K, p)$ where K is the fixed sample size parameter and $E[y_i] = K * p_i$. Usually the model for the mean is specified using the *logit link function* according to

$$\text{logit}(p_i) = \beta_0 + \beta_1 * x_i$$

Where $\text{logit}(u) = \log(u/(1-u))$. The inverse-logit function, g^{-1} , is a function we will refer to as “expit”, so that $\text{expit}(u) = \exp(u)/(1 + \exp(u))$.

A GLMM is the extension of GLMs to accommodate “random effects”. Often this involves adding a normal random effect to the linear predictor, and so a simple example is:

$$\log(\lambda_i) = \alpha_i + \beta_1 * x_i$$

where

$$\alpha_i \sim \text{Normal}(\mu, \sigma^2)$$

2.3 BAYESIAN ANALYSIS

Bayesian analysis is unfamiliar to many ecological researchers because older cohorts of ecologists were largely educated in the classical statistical paradigm of frequentist inference. But advances in technology and increasing exposure to benefits of Bayesian analysis are fast making Bayesians out of people or at least making Bayesian analysis an acceptable, general, alternative to classical, frequentist inference.

Conceptually, the main thing about Bayesian inference is that it uses probability directly to characterize uncertainty about things we don’t know. “Things”, in this case, are parameters of models and, just as it is natural to characterize uncertain outcomes of stochastic processes using probability, it seems natural also to characterize information about unknown “parameters” using probability. At least this seems natural to us and, we think, most ecologists either explicitly adopt that view or tend to fall into that point of view naturally. Conversely, frequentists use

probability in many different ways, but never to characterize uncertainty about parameters¹. Instead, frequentists use probability to characterize the behavior of *procedures* such as estimators or confidence intervals (see below), which can lead to some inelegant or unnatural interpretations of things. It is paradoxical that people readily adopt a philosophy of statistical inference in which the things you don't know (i.e., parameters) should *not* be regarded as random variables, so that, as a consequence, one cannot use probability to characterize one's state of knowledge about them.

2.3.1 Bayes Rule

As its name suggests, Bayesian analysis makes use of Bayes' rule in order to make direct probability statements about model parameters. Given two random variables z and y , Bayes rule relates the two conditional probability distributions $[z|y]$ and $[y|z]$ by the relationship:

$$[z|y] = [y|z][z]/[y]$$

Bayes' rule itself is a mathematical fact and there is no debate in the statistical community as to its validity and relevance to many problems. Generally speaking, these distributions are characterized as follows: $[y|z]$ is the conditional probability distribution of y given z , $[z]$ is the marginal distribution of z and $[y]$ is the marginal distribution of y . In the context of Bayesian inference we usually associate specific meanings in which $[y|z]$ is thought of as "the likelihood", $[z]$ as the "prior" and so on. We leave this for later because here the focus is on this expression of Bayes rule as a basic fact of probability.

As an example of a simple application of Bayes rule, consider the problem of determining species presence at a sample location based on imperfect survey information. Let z be a binary random variable that denotes species presence ($z = 1$) or absence ($z = 0$), let $\Pr(z = 1) = \psi$ where ψ is usually called occurrence probability, "occupancy" (MacKenzie et al., 2002) or "prevalence". Let y be the *observed* presence ($y = 1$) or absence ($y = 0$), and let p be the probability that a species is detected in a single survey at a site given that it is present. Thus, $\Pr(y = 1|z = 1) = p$. The interpretation of this is that, if the species is present, we will only observe presence with probability p . In addition, we assume here that $\Pr(y = 1|z = 0) = 0$. That is, the species cannot be detected if it is not present which is a conventional view adopted in most biological sampling problems (but see Royle and Link (2006)). If we survey a site T times but never detect the species, then this clearly does not imply that the species is not present ($z = 0$) at this site. Rather, our degree of belief in $z = 0$ should be made with a probabilistic statement $\Pr(z = 1|y_1 = 0, \dots, y_T = 0)$. If the T surveys are independent so that we might regard y_t as *iid* Bernoulli trials, then the total number of detections, say y , is Binomial with probability p then we can use Bayes rule to compute the

¹To hear this will be shocking to some readers perhaps.

probability that it is present given that it is not detected in T samples. In words, the expression we seek is:

$$\Pr(\text{present}|\text{not detected}) = \frac{\Pr(\text{not detected}|\text{present}) \Pr(\text{present})}{\Pr(\text{detected})}$$

Mathematically, this is

$$\begin{aligned} \Pr(z = 1|y = 0) &= \Pr(y = 0|z = 1) \Pr(z = 1) / \Pr(y = 0) \\ &= [(1 - p)^T \psi] / [(1 - p)^T \psi + (1 - \psi)]. \end{aligned}$$

To apply this, suppose that $T = 2$ surveys are done at a wetland for a species of frog, and the species is not detected there. Suppose further that $\psi = .8$ and $p = .5$ are obtained from a prior study. Then the probability that the species is present at this site is $.25 * .8 / (.25 * .8 + .2) = 0.50$. That is, there seems to be about a 50/50 chance that the site is occupied despite the fact that the species wasn't observed there.

In summary, Bayes' rule provides a simple linkage between the conditional probabilities $[y|z]$ and $[z|y]$ which is useful whenever one needs to deduce one from the other. Bayes' rule as a basic fact of probability is not disputed.

2.3.2 Bayesian Inference

What is controversial to some is the scope and manner in which Bayes rule is applied by Bayesian analysts. Bayesian analysts assert that Bayes rule is relevant, in general, to all statistical problems by regarding all unknown quantities of a model as realizations of random variables - this includes "data", latent variables, and also "parameters". Classical (non-Bayesian) analysts sometimes object to regarding "parameters" as outcomes of random variables. Classically, parameters are thought of as "fixed but unknown" (using the terminology of classical statistics). Of course, in Bayesian analysis they are also unknown and, in fact, there is a single data-generating value and so they are also fixed. The difference is that this fixed but unknown value is regarded as having been generated from some probability distribution. Specification of that probability distribution is necessary to carryout Bayesian analysis, but it is not required in classical frequentist inference.

To see the general relevance of Bayes rule in the context of statistical inference, let y denote observations - i.e., "data" - and let $[y|\theta]$ be the observation model (often colloquially referred to as the "likelihood"). Suppose θ is a parameter of interest having (prior) probability distribution $[\theta]$. These are combined to obtain the posterior distribution using Bayes' rule, which is:

$$[\theta|y] = [y|\theta][\theta]/[y]$$

Asserting the general relevance of Bayes rule to all statistical problems, we can conclude that the two main features of Bayesian inference are that: (1) "parameters"

θ are regarded as realizations of a random variable and, as a result, (2) inference is based on the probability distribution of the parameters given the data, $[\theta|y]$, which is called the posterior distribution. This is the result of using Bayes rule to combine “the likelihood” and the prior distribution. The key concept is regarding parameters as realizations of a random variable because, once you admit this conceptual view, this leads directly to the posterior distribution, a very natural quantity upon which to base inference about things we don’t know - including parameters of statistical models. In particular, $[\theta|y]$ is a probability distribution for θ and therefore we can make direct probability statements to characterize uncertainty about θ .

The denominator of our invocation of Bayes rule, $[y]$, is the marginal distribution of the data y . We note without further remark right now that, in many practical problems, this can be an enormous pain to compute. The main reason that the Bayesian paradigm has become so popular in the last 20 years or so is because methods exist for characterizing the posterior distribution that do not require that we possess a mathematical understanding of $[y]$, i.e., we never have to compute it or know what it looks like, or know anything specific about it.

A common misunderstanding on the distinction between Bayesian and frequentist inference goes something like this “in frequentist inference parameters are fixed but unknown but in a Bayesian analysis parameters are random.” At best this is a sad caricature of the distinction and at worst it is downright wrong. What is true is that, to a Bayesian, parameters are random variables. However, a Bayesian assumes, just like a frequentist, that there was a single data-generating value of that parameter - a fixed, and unknown value that produced the given data set. The distinction between Bayesian and frequentist approaches is that Bayesians regard the parameter as a random variable, and its value as the outcome of a random value, on par with the observations. This allows Bayesians to use probability to make direct probability statements about parameters. Frequentist inference procedures do not permit direct probability statements to be made about parameter values - because parameters are not random variables!

While we can understand the conceptual basis of Bayesian inference merely by understanding Bayes rule - that’s really all there is to it - it is not so easy to understand the basis of classical “frequentist” inference which is mostly like² a “basket of methods” with little coherent organization. What is mostly coherent in frequentist inference is the manner in which items in this basket of methods are evaluated - the performance of a given procedure is evaluated by “averaging over” hypothetical realizations of y , regarding the *estimator* as a random variable. For example, if $\hat{\theta}$ is an estimator of θ then the frequentist is interested in $E_y[\hat{\theta}|y]$ which is used to characterize bias. If the expected value of $\hat{\theta}$, when averaged over realizations of y , is equal to θ , then $\hat{\theta}$ is unbiased.

The view of parameters as fixed constants and estimators as random variables leads to interpretations that are not so straightforward. For example confidence

²Characterization from Sims REF XYZ

intervals having the interpretation “95% probability that the interval contains the true value” and p-values being “the probability of observing an outcome as extreme or more than the one observed.” These are far from intuitive interpretations to most people. Moreover, this is conceptually problematic to some because the hypothetical realizations that characterize the performance of our procedure we will never get to observe.

While we do tend to favor Bayesian inference for the conceptual simplicity (parameters are random, posterior inference), we mostly advocate for a pragmatic non-partisian approach to inference because, frankly, some of these “bucket of methods” are actually very convenient in certain situations as we will see in later chapters.

2.3.3 Prior distributions

The prior distribution $[\theta]$ is an important feature of Bayesian inference. As a conceptual matter, the prior distribution characterizes “prior beliefs” or “prior information” about a parameter. Indeed, an oft-touted benefit of Bayesian analysis is the ease with which prior information can be included in an analysis. However, more commonly, the prior is chosen to express a lack of prior information, even if previous studies have been done and even if the investigator does in fact know quite a bit about a parameter. This is because the manner in which prior information is embodied in a prior (and the amount of information) is usually very subjective and thus the result can wind up being very contentious, e.g., if different investigators might report different results based on subjective assessments of things. Thus it is usually better to “let the data speak” and use priors that reflect absence of information beyond the data set being analyzed.

But still the need occasionally arises to embody prior information or beliefs about a parameter formally into the estimation scheme. In SCR models we often have a parameter that is closely linked to “home range radius” and thus auxiliary information on the home range size of a species can be used as prior information (e.g., see Chandler and Royle (2012) ; also chapter XYZ).

XXXXXXXXXX

noninformative prior on one scale is informative on another scale. e.g., flat prior on $\text{logit}(p)$ is very different from $\text{uniform}(0,1)$ on p ... show graphic.....

reference to non-invariance of prior distributions to transformation.....

XXXXXXXXXX

2.3.4 Posterior Inference

In Bayesian inference, we are not focusing on estimating a single point or interval but rather on characterizing a whole distribution – the posterior distribution – from which one can report any summary of interest. A point estimate might be the posterior mean, median, mode, etc.. In many applications in this book, we will

compute 95% Bayesian intervals using the 2.5% and 97.5% quantiles of the posterior distribution. For such intervals, it is correct to say $\Pr(L < \theta < U) = 0.95$. That is, "the probability that θ is between L and U is 0.95". It is not a subtle thing that this cannot be said using frequentist methods - although people tend to say it anyway and not really understand why it is wrong or even that it is wrong. This is actually a failing of frequentist ideas and the inability of frequentists to get people to overcome their natural tendency to use probability - which is something that, as a frequentist, you simply cannot do in the manner that you would like to.

Posterior inference is the main practical element of Bayesian analysis. We get to make an inference conditional on the data that we actually observed - i.e., what we actually know. To us, this seems logical - to condition on what we know. Conversely, frequentist inference is based on considering average performance over hypothetical unobserved data sets (i.e., the "relative frequency" interpretation of probability). Frequentists know that their procedures work well when averaged over all hypothetical, unobserved, data sets but no one ever really knows how well they work for the specific data set analyzed. That seems like a relevant question to biologists who oftentimes only have their one, extremely valuable, data set. This distinction comes into play a lot in exposing philosophical biases in the peer review of statistical analyses in ecology in the sense that, despite these opposing conceptual views to inference (i.e. conditional on the data you have, or averaged over hypothetical realizations), those who conduct a Bayesian analysis are often (in ecology, almost always) required to provide a frequentist evaluation of their Bayesian procedure.

2.3.5 Small sample inference

Using Bayesian inference, we obtain an estimate of the posterior distribution which is an exhaustive summary of the state-of-knowledge about an unknown quantity. It is the posterior distribution - not an estimate of that thing. It is also not, usually, an approximation except to within Monte Carlo error (in cases where we use simulation to calculate it). One of the great virtues of Bayesian analysis which is not really appreciated is that it is completely valid for any particular sample size. i.e., it is $[\theta|y]$, as precise as we claim it to be based on our ability to do calculations, for the particular sample size and observations that we have even if we have only a single datum y . The same cannot be said for almost all frequentist procedures in which estimates or variances are very often (almost always in practice) based on "asymptotic approximations" to the procedure which is actually being employed.

There seems to be a prevailing view in statistical ecology that classical likelihood-based procedures are virtuous because of the availability of simple formulas and procedures for carrying out inference, such as calculating standard errors, doing model selection by AIC, and assessing goodness-of-fit. In large samples, this may be an important practical benefit, but the theoretical validity of these procedures cannot be asserted in most situations involving small samples. This is not a minor

issue because it is typical in many wildlife sampling problems - especially in surveys of carnivores or rare/endangered species - to wind up with a small, sometimes extremely small, data set. For example, a recent paper on the fossa (*Cryptoprocta ferox*), an endangered carnivore in Madagascar, estimated an adult density of 0.18 adults / km sq based on 20 animals captured over 3 years (Hawkins and Racey, 2005). A similar paper on the endangered southern river otter (*Lontra provocax*) estimated a density of 0.25 animals per river km based on 12 individuals captured over 3 years (Sepúlveda et al., 2007). Gardner et al. (2010) analyzed data from a study of the Pampas cat, a species for which very little is known, wherein only 22 individual cats were captured during the two year period. Trolle and Kéry (2005) reported only 9 individual ocelots captured and Jackson et al. (2006) captured 6 individual snow leopards using camera trapping. Thus, studies of rare and/or secretive carnivores necessarily and flagrantly violate one of Le Cam's Basic Principles, that of "If you need to use asymptotic arguments, do not forget to let your number of observations tend to infinity." (Le Cam, 1990).

The biologist thus faces a dilemma with such data. On one hand, these datasets, and the resulting inference, are often criticized as being poor and unreliable. Or, even worse³, "the data set is so small, this is a poor analysis." On the other hand, such data may be all that is available for species that are extraordinarily important for conservation and management. The Bayesian framework for inference provides a valid, rigorous, and flexible framework that is theoretically justifiable in arbitrary sample sizes. This is not to say that one will obtain precise estimates of density or other parameters, just that your inference is coherent and justifiable from a conceptual and technical statistical point of view. That is, we report the posterior probability $\Pr(D|data)$ which is easily interpretable and just what it is advertised to be and we don't need to do a simulation study to evaluate how well some approximate $\Pr(D|data)$ deviates from the actual $\Pr(D|data)$ because they are precisely the same quantity.

2.4 CHARACTERIZING POSTERIOR DISTRIBUTIONS BY MCMC SIMULATION

In practice, it is not really feasible to ever compute the marginal probability distribution $\Pr(y)$, the denominator resulting from application of Bayes' rule. For decades this impeded the adoption of Bayesian methods by practitioners. Or, the few Bayesian analyses done were based on asymptotic normal approximations to the posterior distribution. While this was useful stuff from a theoretical and technical standpoint and, practically, it allowed people to make the probability statements that they naturally would like to make, it was kind of a bad joke around the Bayesian water-cooler to, on one hand, criticize classical statistics for being, essentially, completely ad hoc in their approach to things but then, on the other hand,

³Actual quote from a referee

524 have to devise various approximations to what they were trying to characterize.
 525 The advent of Markov chain Monte Carlo (MCMC) methods has made it easier to
 526 calculate posterior distributions for just about any problem to arbitrary levels of
 527 precision.

528 Broadly speaking, MCMC is a class of methods for drawing random numbers
 529 (sampling or simulating) from the target posterior distribution. Thus, even though
 530 we might not recognize the posterior as a named distribution or be able to ana-
 531 lyze its features analytically, e.g., devise mathematical expressions for the mean
 532 and variance, we can use these MCMC methods to obtain a large sample from the
 533 posterior and then use that sample to characterize features of the posterior. What
 534 we do with the sample depends on our intentions – typically we obtain the mean or
 535 median for use as a point estimate, and take a confidence interval based on Monte
 536 Carlo estimates of the quantiles. These are estimates, but not like frequentist es-
 537 timates. Rather, they are Monte Carlo estimates with an associated Monte Carlo
 538 error which is largely determined arbitrarily by the analyst. They are not estimates
 539 qualified by a sampling distribution as in classical statistics. If we run our MCMC
 540 long enough then our reported value of $E[\theta|y]$ or any feature of the posterior dis-
 541 tribution is precisely what we say it is. There is no “sampling variation” in the
 542 frequentist sense of the word. In summary, the MCMC samples provide a Monte
 543 Carlo characterization of *the* posterior distribution.

2.5 WHAT GOES ON UNDER THE MCMC HOOD

544 We will develop and apply MCMC methods in some detail for spatial capture-
 545 recapture models in chapter 7. Here we provide a simple illustration of some basic
 546 ideas related to the practice of MCMC.

547 A type of MCMC method relevant to most problems is Gibbs sampling (REF
 548 XYZ XYZ), which is based on the idea of iterative simulation from the “full con-
 549 ditional” distributions (also called conditional posterior distributions). The full
 550 conditional distribution for an unknown quantity is the conditional distribution of
 551 that quantity given every other random variable in the model - the data and all other
 552 parameters. For example, for a normal regression model with $y \sim \text{Normal}(\alpha + \beta x, 1)$
 553 then the two full conditionals are, in symbolic terms,

$$[\alpha|y, \beta]$$

554 and

$$[\beta|y, \alpha].$$

555 We might use our knowledge of probability to identify these mathematically. In
 556 particular, by Bayes’ Rule, $[\alpha|y, \beta] = [y|\alpha, \beta][\alpha|\beta]/[y|\beta]$ and similarly for $[\beta|y, \alpha]$.
 557 For example, if we have priors for $[\alpha]$ and $[\beta]$ which are also normal distributions,
 558 some algebra reveals that XXXX COPY NOTATION FFROM CH. 6 XXXXX

$$[\alpha|y, \beta] = \text{Normal}(\text{ybar}, \dots \text{weightedvariancehere} \dots).$$

559 Similarly,

$$[\beta|y, \alpha] \text{isnormal}(\dots\dots)$$

560 The MCMC algorithm for this model has us simulate in succession, repeat-
 561 edly, from those two distributions. See Gelman et al. (2004) for more examples of
 562 Gibbs sampling for the normal model. A conceptual representation of the MCMC
 563 algorithm for this simple model is therefore: XXXX Check out ALGORITHM en-
 564 vironment XXXXX

565 **Algorithm**

```
566
567     0. Initialize  $\alpha$  and  $\beta$ 
568
569     Repeat{
570         1. Draw a new value of  $\alpha$  from Eq. \ref{xyz}
571
572         2. Draw a new value of  $\beta$  from Eq. \ref{xyz}
573     }
```

574 As we just saw for this simple “normal-normal” model it is sometimes possible
 575 to specify the full conditional distributions analytically. In general, when certain
 576 so-called conjugate prior distributions are chosen, the form of full conditional distri-
 577 butions is similar to that of the observation model. In this normal-normal case, the
 578 normal distribution for the mean parameters is the conjugate prior under the normal
 579 model, and thus the full-conditional distributions are also normal. This is conven-
 580 nient because, in such cases, we can simulate directly from them using standard
 581 methods (or **R** functions). But, in practice, we don’t really ever need to know such
 582 things because most of the time we can get by using a simple algorithm, called the
 583 Metropolis-Hastings (henceforth “MH”) algorithm, to obtain samples from these
 584 full conditional distributions without having to recognize them as specific, named,
 585 distributions. This gives us enormous freedom in developing models and analyzing
 586 them without having to resolve them mathematically because to implement the MH
 587 algorithm we need only identify the full conditional distribution up to a constant
 588 of proportionality, that being the marginal distribution in the denominator (e.g.,
 589 $[y|\beta]$ above).

590 We will talk about the Metropolis-Hastings algorithm shortly, and we will use
 591 it extensively in the analysis of SCR models (e.g., chapter 7).

592 2.5.1 Rules for constructing full conditional distributions

593 The basic strategy for constructing full-conditional distributions for devising MCMC
 594 algorithms can be reduced conceptually to a couple of basic steps summarized as
 595 follows:

596 (step 1) Collect all stochastic components of the model;

- 597 (step 2) Recognize and express the full conditional in question as proportional to
 598 the product of all components;
 599 (step 3) Remove the ones that don't have the focal parameter in them.
 600 (step 4) Do some algebra on the result in order to identify the resulting pdf or pmf.

601 Of the 4 steps, the last of those is the main step that requires quite a bit of statistical
 602 experience and intuition because various algebraic tricks can be used to reshape the
 603 mess into something noticeable - i.e., a standard, named distribution. But step 4
 604 is not necessary if we decide instead to use the Metropolis-Hastings algorithm as
 605 described below.

606 To illustrate for computing $[\alpha|y, \beta]$ we first apply step 1 and identify the model
 607 components as: $[y|\alpha, \beta]$, $[\alpha]$ and $[\beta]$. Step 2 has us write $[\alpha|y, \beta] \propto [y|\alpha, \beta][\alpha][\beta]$.
 608 Step 3: We note that $[\beta]$ is not a function of alpha and therefore we remove it to
 609 obtain $[\alpha|y, \beta] \propto [y|\alpha, \beta][\alpha]$. Similarly we obtain $[\beta|y, \alpha] \propto [y|\alpha, \beta][\beta]$. We apply
 610 step 4 and manipulate these algebraically to arrive at the result or, alternatively, we
 611 can sample them indirectly using the Metropolis-Hastings algorithm (see below).

612 2.5.2 Metropolis-Hastings algorithm

613 The Metropolis-Hastings (MH) algorithm is a completely generic method for sam-
 614 pling from any distribution, say $f(\theta)$. In our applications, $f(\theta)$ will typically be
 615 the full conditional distribution of θ . While we sometimes use Gibbs sampling,
 616 we seldom use “pure” Gibbs sampling because we might use MH to sample from
 617 one or more of the full conditional distributions. When the MH algorithm is used
 618 to sample from full conditional distributions of a Gibbs sampler the resulting hy-
 619 brid algorithm is called *Metropolized Gibbs sampling* or more commonly *Metropolis-*
 620 *within-Gibbs*. Shortly we will actually construct such an algorithm for a simple
 621 class of models.

622 The MH algorithm generates candidates from some proposal or candidate-
 623 generating distribution, that may be conditional on the current value of the pa-
 624 rameter, denoted by $h(\theta^*|\theta^t)$. Here, θ^* is the *candidate* or proposed value and θ^t is
 625 the current value, i.e., at iteration t of the MCMC algorithm. The proposed value
 626 is accepted with probability XXXX check notation with Rahel XXXXXX

$$r = \frac{f(\theta^*)h(\theta^t|\theta^*)}{f(\theta^t)h(\theta^*|\theta^t)}$$

627 which we call the MH acceptance probability. This ratio can sometimes be > 1 in
 628 which case we set it equal to 1. It is useful to note that $h()$ can be anything at all.
 629 Absolutely anything! You can generate candidate values from a *normal*(0,1) dis-
 630 tribution, from a *uniform*(-3455,3455) distribution, or anything of proper support.
 631 Note, however, that good choices of $h()$ are those that approximate the posterior
 632 distribution. Obviously if $h() = f(\theta|y)$ (i.e., the posterior) then you always accept
 633 the draw, and it stands to reason that proposals that are more similar to $f(\theta|y)$

will lead to higher acceptance probabilities. No matter the choice of $h()$, we can evaluate this ratio numerically because the marginal $f(y)$ cancels from both the numerator and denominator, which is the magic of the MH algorithm.

A special kind of $h()$ are those that are symmetric, which means that $h(a|b) = h(b|a)$ in which case $h(a|b)$ and $h(b|a)$ just cancel out from the MH acceptance probability and r is then just the ratio of the target density evaluated at the candidate value to that evaluated at the current value. A type of symmetric proposal useful in many situations is the so-called *random-walk* proposal distribution where candidate values are drawn from a normal distribution with mean equal to the current value and some standard deviation, say δ , which is prescribed by the user. For parameters that have support on the real line, e.g., α in our example above, the random walk proposal generator has us generate $\alpha^* \sim \text{Normal}(\alpha^t, \delta)$. If we set δ very small we have a high probability of accepting the proposal and vice versa. In practice, we “tune” delta to achieve a compromise between acceptance rate and efficient mixing of the Markov chains (see below for an example) normally assessed by autocorrelation. Low δ increases the acceptance rate but will tend to produce Markov chains with high autocorrelation, and vice versa.

Parameters with bounded support: Many models contain parameters that have bounded support. E.g., variance parameters live on $[0, \infty]$, parameters that represent probabilities live on $[0, 1]$, etc.. In that case it is sometimes convenient to use a random walk proposal distribution that can generate any real number (e.g., a normal random walk proposal). In that case, we can just reject parameters that are outside of the parameter space (XXXX REF FOR THIS XXXX).

2.6 PRACTICAL BAYESIAN ANALYSIS AND MCMC

There are a number of really important practical issues to be considered in any Bayesian analysis and we cover some of these briefly here.

2.6.1 Choice of prior distributions

**XXX integrate this material with previous section on prior distributions
XXXXXX**

Bayesian analysis requires that we choose prior distributions for all of the structural parameters of the model (we use the term structural parameter to mean all parameters that aren’t customary thought of as latent variables). We will strive to use priors that are meant to express little or no prior information - default or customary “non-informative” or diffuse priors. This will be $\text{Unif}(a, b)$ priors for parameters that have a natural bounded support and, for parameters that live on the real line we use either (1) diffuse normal priors; (2) “improper” uniform priors or (3) sometimes even a bounded $\text{Unif}(a, b)$ prior if that greatly improves the performance of **WinBUGS** or other software doing the MCMC for us. In **WinBUGS** a prior with low “precision”, τ , where $\tau = 1/\sigma^2$, such as $\text{Norm}(0, .01)$ will typically

be used. Of course $\tau = 0.01$ ($\sigma^2 = 100$) might be very informative for a regression parameter that has a high variance. Therefore, we recommend that predictor variables *always* be standardized. Clearly there are a lot of choices for ostensibly non-informative priors, and the degree of non-informativeness depends on the parameterization. For example, a natural non-informative prior for the intercept of a logistic regression

$$\text{logit}(p_i) = \alpha + \beta x_i$$

Would be $[\alpha] = \text{const}$ which is the same as saying $a \sim \text{Unif}(\infty, \text{infy})$, the customary improper uniform prior. However, we might also use a prior on the parameter $p_0 = \text{logit}^{-1}(a)$, which is $\text{Pr}(y = 1)$ for the value $x = 0$. Since p_0 is a probability a natural choice is $p_0 \sim \text{Unif}(0, 1)$. These two priors can affect results (see Chapter 3.XYZ), yet they are both sensible non-informative priors. Choice of priors and parameterization is very much problem-specific and often largely subjective. Moreover, it also affects the behavior of MCMC algorithms and therefore the analyst needs to pay some attention to this issue and possibly try different things out. XXX REFS on prior distributions XXXXX

2.6.2 Convergence and so-forth

Once we have carried-out an analysis by MCMC, there are many other practical issues that we have to confront. One of the most important is “have the chains converged?” Most MCMC algorithms only guarantee that, eventually, the samples being generated will be from the target posterior distribution. So-called “convergence” of the Markov chain is achieved when that happens. Typically a period of transience is observed in the early part of the MCMC algorithm, and this is usually discarded as the “burn-in” period.

The quick diagnostic to whether convergence has been achieved is that your Markov chains look “grassy” – see Fig. 2.5 below. Another way to check convergence is to update the parameters some more and see if the posterior changes. It is good to confirm convergence using the “R-hat” statistic (\hat{R}) or Brooks-Gelman-Rubin statistic (Gelman et al., 1996)) which should be close to 1 if the Markov chains have converged and sufficient posterior samples have been obtained. In practice, $\hat{R} = 1.2$ is probably good enough for some problems. For some models you can’t actually realize a low \hat{R} . E.g., if the posterior is a discrete mixture of distributions then you can be misled into thinking that your Markov chains have not converged when in fact the chains are just jumping back and forth in the posterior state-space. So, for example, using model selection methods (section XYZ) sometimes suggests non-convergence. Another situation is when one of the parameters is on the boundary of the parameter space which might appear to be very poor mixing, but all within some extreme region of the parameter space.⁴. This

⁴it would be nice if we could compile examples of this later in the book and reference back to this point

kind of stuff is normally ok and you need to think really hard about the context of the model and the problem before you conclude that your MCMC algorithm is ill-behaved.

Some models exhibit “poor mixing” of the Markov chains or what people might also say “have not coveredged” (or “slow convergence”) which is a term we would disagree with because the samples might well be from the posterior (i.e., the Markov chains have converged to the proper stationary distribution) but simply mix around the posterior rather slowly. Anyway, poor mixing can happen for a huge number of reasons – when parameters are highly correlated (even confounded), or barely identified from the data, or the algorithms are very terrible and probably many other reasons. Slow mixing equates to high autocorrelation in the Markov chain – the successive draws are highly correlated, and thus we need to run the MCMC algorithm much longer to get an effective sample size that is sufficient for estimation – or to reduce the MC error to a tolerable level. A strategy often used to reduce autocorrelation is “thinning” – i.e., keep every m^{th} value of the Markov chain output. However, thinning is necessarily inefficient from the stand point of inference – you can always get more precise posterior estimates by using all of the MCMC output regardless of the level of autocorrelation (MacEachern and Berliner, 1994). Practical considerations might necessitate thinning, even though it is statistically inefficient. For example, in models with many parameters or other unknowns being tabulated, the output files might be enormous and unwieldy to work with. In such cases, thinning is perfectly reasonable. In many cases, how well the Markov chains mix is strongly influenced by parameterization, standardization of covariates, and the prior distributions being used. Some things work better than others, and the investigator should experiment with different settings and remain calm when things don’t work out perfectly. MCMC is an art, and a science.

Is the posterior sample large enough? A good rule of thumb is that you should never report MCMC results to more than 2 decimal places – because they will always be different! Look at the MC error which is printed by default in summaries of BUGS output. You want that to be smallish relative to the magnitude of the parameter and this might depend on the purpose of the analysis. For a preliminary analysis you might settle for a few percent whereas for a final analysis then certainly less than 1% is called for, but you can run your MCMC algorithm as long as it takes. Note that MC error in summaries of the posterior is not the same as having an “approximate” solution in a standard likelihood analysis or similar. The approximate SE in likelihood inference is actually wrong in its actual value.... XYZ.

2.6.3 Bayesian confidence intervals

The 95% Bayesian interval based on percentiles of the posterior is not a unique interval – there are many of them – and the so-called “highest posterior density” (HPD) interval is the narrowest interval. We might compute that frequently because

it is easy to do with an integer parameter which N is (See the next chapter). The 95 % HPD is not often exactly 95% but usually slightly more conservative than nominal because it is the narrowest interval that contains at least 95% of the posterior mass.

2.6.4 Estimating functions of parameters

A benefit of analysis by MCMC is that we can seamlessly estimate functions of parameters by simply tabulating the desired function of the simulated posterior draws. For example, if θ is the parameter of interest and let $\theta^{(i)}$ for $i = 1, 2, \dots, M$ be the posterior samples of θ . Let $\eta = \exp(\theta)$, then a posterior sample of η can be obtained simply by computing $\exp(\theta^{(i)})$ for $i = 1, 2, \dots, M$. We give another example in section 2.7.2 below and throughout this book. Almost all SCR models in this book involve at least 1 derived parameter. For example, density D is a derived parameter, being a function of population size N and the area A of the underlying state-space of the point process (see chapter 4).

2.7 BAYESIAN ANALYSIS USING WINBUGS

We won't be too concerned with devising our own MCMC algorithms for every analysis although we will do that a few times for fun. More often, we will rely on the freely available software package **WinBUGS** or **JAGS** for doing this. We will always execute these **BUGS** engines from within **R** using the **R2WinBUGS** (REF XYZ XYZ) or **rjags** packages. **WinBUGS** and **JAGS** are MCMC black boxes that takes a pseudo-code description (i.e., written in the **BUGS** language) of all of the relevant stochastic and deterministic elements of a model and generates an MCMC algorithm for that model. But you never get to see the algorithm. Instead, **WinBUGS/JAGS** will run the algorithm and just return the Markov chain output - the posterior samples of model parameters.

The great thing about using the **BUGS** language is that it forces you to become intimate with your statistical model - you have to write each element of the model down, admit (explicitly) all of the various assumptions, understand what the actual probability assumptions are and how data relate to latent variables and data and latent variables relate to parameters, and how parameters relate to one another.

While we normally use **WinBUGS** or **JAGS** in this book, we note that **OpenBUGS** is the current active development tree of the **BUGS** language. See Kéry (2010, ch.xyz) and Kery and Schaub (2011, appendix xyz) for more on practical analysis in **WinBUGS**. That book should also be consulted for a more comprehensive introduction to using **WinBUGS**. In this example, we're going to accelerate pretty fast.

2.7.1 Linear Regression in WinBUGS

We provide a brief introductory example of a normal regression model using a small simulated data set. The following commands are executed from within your **R**

workspace, the command line being indicated by ‘>’. First, simulate a covariate x and observations y having prescribed intercept, slope and variance:

```

787 > x<-rnorm(10)
788 > mu<- -3.2+ 1.5*x
791 > y<-rnorm(10,mu,sd=4)

```

The **BUGS** model specification for a normal regression model is written within **R** as a character string input to the command `cat()` and then dumped to a text file named `normal.txt`:

```

795 > cat("
796 model {
797   for (i in 1:10){
798     y[i]~dnorm(mu[i],tau)      # the "likelihood"
799     mu[i]<- beta0 + beta1*x[i] # the linear predictor
800   }
801   beta0~dnorm(0,.01)          # prior distributions
802   beta1~dnorm(0,.01)
803   sigma~dunif(0,100)
804   tau<-1/(sigma*sigma)       # tau is a derived parameter
805 }
806 ",file="normal.txt")

```

Alternatively, you can write the model specifications directly within a text file and save it in your current working directory, but we do not usually take that approach in this book.

Remarks: 1. WinBUGS parameterizes the normal in terms of the mean and inverse-variance, called the precision. Thus, `dnorm(0,.01)` implies a variance of 100; **2.** We typically use diffuse normal priors for mean parameters, β_0 and β_1 in this case, but sometimes we might use uniform priors with suitable bounds $-B$ and $+B$. **3.** We typically use a `Unif(0, B)` prior on standard deviation parameters (Gelman XXX 2006 XXXX). But sometimes we might use a gamma prior on the precision parameter τ . **4.** In a **WinBUGS** model file, every variable referenced in the model description has to be either data, which will be input (see below), a random variable which must have a probability distribution associated with it using the “~”, or it has to be a derived parameter connected to variables and data using “<-”.

To fit the model, we need to describe various data objects to **WinBUGS**. In particular, we create an **R** list object called `data` which are the data objects identified in the BUGS model file. In the example, the data consist of two objects which exist as y and x in the **R** workspace and also in the **WinBUGS** model definition. We also have to create an **R** function that produces a list of starting values `inits` that get sent to **WinBUGS**. Finally, we identify the names of the parameters (labeled correspondingly in the **WinBUGS** model specification) that

we want **WinBUGS** to save the MCMC output for. In this example, we will “monitor” the parameters β_0 , β_1 , σ and τ . **WinBUGS** is executed using the **R** command `bugs()`. We set the option `debug=TRUE` if we want the **WinBUGS** GUI to stay open (useful for analyzing MCMC output and looking at the **WinBUGS** error log). Also, we set `working.dir=getwd()` so that **WinBUGS** output files and the log file are saved in the current **R** working directory. All of these activities look like this:

```
library("R2WinBUGS") # "attach" the R2WinBUGS library
data <- list ( "y","x")
inits <- function()
  list ( beta1=rnorm(1),beta0=rnorm(1),sigma=runif(1,0,2) )
parameters <- c("beta0","beta1","sigma","tau")
out<-bugs (data, inits, parameters, "normal.txt", n.thin=2, n.chains=2,
           n.burnin=2000, n.iter=6000, debug=TRUE,working.dir=getwd())
```

Remarks: A common question is “how should my data be formatted?” That depends on how you describe the model in the **BUGS** language, how your data are input into **R** and subsequently formatted. There is no unique way to describe any particular model and so you have some flexibility. We talk about data format further in the context of capture-recapture models and SCR models in chapter 4 and elsewhere. In general, starting values are optional but we recommend to always provide reasonable starting values for structural parameters, but are not always necessary for random effects. Note that the previously created objects defining data, initial values and parameters to monitor are passed to the function `bugs()`. In addition, various other things are declared: The number of Markov chains (`n.chains`), the thinning rate (`n.thin`), the number of burn-in iterations (`n.burnin`) and the total number of iterations (`n.iter`). To develop a detailed understanding of the various parameters and settings used for MCMC, consult a basic reference such as Kéry (2010).

You should execute all of the commands given above and then look at the resulting output. Kill the **WinBUGS** GUI and the data will be read back into **R** (or specify `debug=FALSE`). We don’t want to give instructions on how to navigate and use the GUI - see XYZ REF (XYZ) for that. The object `out` prints important summaries by default (this is slightly edited):

```
> print(out,digits=2)
Inference for Bugs model at "normal.txt", fit using WinBUGS,
2 chains, each with 6000 iterations (first 2000 discarded), n.thin = 2
n.sims = 4000 iterations saved
```

	mean	sd	2.5%	25%	50%	75%	97.5%	Rhat	n.eff
beta0	-2.43	1.84	-6.21	-3.50	-2.42	-1.34	1.27	1	4000
beta1	2.62	1.54	-0.42	1.68	2.62	3.57	5.67	1	4000
sigma	5.29	1.66	3.11	4.14	4.95	6.05	9.39	1	4000
tau	0.05	0.02	0.01	0.03	0.04	0.06	0.10	1	4000
deviance	59.85	3.24	56.18	57.47	59.00	61.37	68.32	1	840

871 For each parameter, `n.eff` is a crude measure of effective sample size,
 872 and `Rhat` is the potential scale reduction factor (at convergence, `Rhat=1`).
 873
 874 DIC info (using the rule, `pD = Dbar-Dhat`)
 875 `pD = 2.6` and `DIC = 62.4`
 876

877 **Remarks:** (1) convergence is assessed using the \hat{R} statistic – which we might
 878 sometimes write “*Rhat*”. A value of *Rhat* near 1 indicates convergence; (2) DIC
 879 is the “deviance information criterion” (Spiegelhalter et al., 2002) (see section 2.8)
 880 which some people use in a manner similar to AIC although it is recognized to
 881 have some problems in hierarchical models (Millar, 2009). We evaluate this in the
 882 context of SCR models in chapter XYZ XYZ.

883 2.7.2 Inference about functions of model parameters

884 Using the MCMC draws for a given model we can easily obtain the posterior distri-
 885 bution of any function of model parameters. We showed this in the above example
 886 by providing the posterior of τ when the model was parameterized in terms of stan-
 887 dard deviation σ . As another example, suppose that the normal regression model
 888 above had a quadratic response function of the form

$$E(y_i) = \beta_0 + \beta_1 x_i + \beta_2 x_i^2$$

889 Then the optimum value of x , i.e., that corresponding to the optimal expected
 890 response, can be found by setting the derivative of this function to 0 and solving
 891 for x . We find that

$$df/dx = \beta_1 + 2 * \beta_2 x = 0$$

892 yields that $x_{opt} = -\beta_1/(2 * \beta_2)$. We can just take our posterior draws for β_1
 893 and β_2 and obtain a posterior sample of x_{opt} by this simple calculation. As an
 894 exercise, take the normal model above and simulate a quadratic response and then
 895 describe the posterior distribution of x_{opt} .

2.8 MODEL CHECKING AND SELECTION

896 In general terms model checking - or assessing the adequacy of the model - and
 897 model selection are quite thorny issues and, despite contrary and, sometimes,
 898 strongly held belief among practitioners, there are not really definitive, general
 899 solutions to either problem. We're against dogma on these issues and think people
 900 need to be open-minded about such things and recognize that models can be useful
 901 whether or not they pass certain statistical tests. Some models are intrinsically
 902 better than others because they make more biological sense or foster understanding
 903 or achieve some objective that some bootstrap or other goodness-of-fit test can't

decide for you. That said, it gives you some confidence if your model seems adequate and we try to provide some fit assessment in most real applications of SCR models. We provide a very brief overview of concepts here, but provide more detailed coverage in chapter 8. See also Kéry (2010, ch. xyz) and Link and Barker (2009, ch. xyz) for specific context related to Bayesian model checking and selection.

2.8.1 Goodness-of-fit

Goodness-of-fit testing is an important element of any analysis because our model represents a general set of hypotheses about the ecological and observation processes that generated our data. Thus, if our model “fits” in some statistical or scientific sense, then we believe it to be consistent with the hypotheses that went into the model. More formally, we would conclude that the data are *not inconsistent* with the hypotheses, or that the model appears adequate. If we have enough data, then of course we will reject any set of statistical hypotheses. Conversely, we can always come up with a model that fits by making the model extremely complex. Despite this paradox, it seems to us that simple models that you can understand should usually be preferred even if they don’t fit, for example if they embody essential mechanisms central to our understanding of things, or if we think that some contributing factors to lack-of-fit are minor or irrelevant to the scientific context and intended use of the model. In other words, models can be useful irrespective of whether they fit according to some formal statistical test of fit. Yet the tension is there to obtain fitting models, and this comes naturally at the expense of models that can be easily interpreted and studied and effectively used. Moreover, conducting goodness-of-fit tests is not always so easy to do. Moreover, it is never really easy (or especially convenient) to decide if your goodness-of-fit test is worth anything. It might have 0 power! Despite this, we recommend attempting to assess model fit in real applications, as a general rule, and we provide some basic guidance here and some more specific to SCR models in chapter 8.

To evaluate goodness-of-fit in Bayesian analyses, we will most often use the Bayesian p-value (Gelman et al., 1996). The basic idea is to define a fit statistic or “discrepancy measure” and compare the posterior distribution of that statistic to the posterior predictive distribution of that statistic for hypothetical perfect data sets for which the model is known to be correct. For example, with count frequency data, a standard measure of fit is the sum of squares of the “Pearson residuals”,

$$D(y_i, \theta) = \frac{(y_i - E(y_i))^2}{\text{Var}(y_i)}$$

The fit statistic based on the squared residuals is

$$FIT = \sum_i D(y_i, \theta)^2$$

which can be computed at each iteration of a MCMC algorithm given the current values of parameters that determine the response distribution. At the same time (i.e., at each MCMC iteration), the equivalent statistic is computed for a “new” data set, simulated using the current parameter values. The Bayesian p-value is simply the posterior probability $\Pr(\text{Fit} > \text{Fit}_{\text{new}})$ ⁵ which should be close to 0.50 for a good model – one that “fits” in the sense that the observed data set is consistent with realizations simulated under the model being fitted to the observed data. In practice we judge “close to 0.50” as being “not too close to 0 or 1” and, as always, closeness is somewhat subjective. We’re happy with anything $> .1$ and $< .9$ but might settle for $> .05$ and < 0.95 . In summary, the Bayesian p-value seems like a bootstrap idea, is easy to compute, and widely used as a result.

Another useful fit statistic is the Freeman-Tukey statistic⁶, in which

$$D(\mathbf{y}, \theta) = \sum_i (\sqrt{y_i} - \sqrt{e_i})^2$$

(Brooks et al., 2000), where y_i is the observed value of observation i and e_i its expected value. In contrast to a chi-square discrepancy, the Freeman-Tukey statistic removes the need to pool cells with small expected values.

2.8.2 Model Selection

For model selection we typically use three different methods: First is, let’s say, common sense. If a parameter has posterior mass concentrated away from 0 then it seems like it should be regarded as important - that is, it is “significant.” This approach seems to have fallen out of favor with all of the interest over the last 10 or 15 years on model selection in ecology. It seems reasonable to us.

For regression problems we sometimes use the factor weighting idea which is to introduce a set of binary variables w_k for variable k , and express the model as, e.g., for a single covariate model:

$$E(y_i) = \alpha + w\beta x_i$$

where w is given a Bernoulli prior distribution with some prescribed probability. E.g., $w \sim \text{Bern}(0.50)$ to provide a prior probability of 0.50 that variable x should be an element of the linear predictor. The posterior probability of the event $w = 1$ is a gauge of the importance of the variable x . i.e., high values of $\Pr(w = 1)$ indicate stronger evidence to support that “ x is in the model” whereas values of $\Pr(w = 1)$ close to 0 suggest that x is less important.

This idea seems to be due to Kuo and Mallick (1998)⁷ and see Royle and Dorazio (2008, ch. XXXX) for an example in the context of logistic regression. This

⁵Check this definition!

⁶Ref for this?

⁷ Is this also what people call Zellner’s G-priors?

approach seems to even work sometimes with fairly complex hierarchical models of a certain form. E.g., Royle (2008) applied it to a random effects model to evaluate the importance of the random effect component of the model. The main problem with this approach is that its effectiveness and results will typically be highly sensitive to the prior distribution on the structural parameters (e.g., see Royle and Dorazio (2008, table xyz)). The reason for this is obvious: If $w = 0$ for the current iteration of the MCMC algorithm, so that β is sampled from the prior distribution, and the prior distribution is very diffuse, then extreme values of β are likely. Consequently, when the current value of β is far away from the mass of the posterior when $w = 1$, then the Markov chain may only jump from $w = 0$ to $w = 1$ infrequently. One seemingly reasonable solution to this problem (Aitken XYZ FIND THIS XXXXX⁸) is to fit the full model to obtain posterior distributions for all parameters, and then use those as prior distributions in a “model selection” run of the MCMC algorithm. This seems preferable to more-or-less arbitrary restriction of the prior support to improve the performance of the MCMC algorithm.

A third method that we advocate is subject-matter context. It seems that there are some situations – some models – where one should not have to do model selection because it is necessitated by the specific context of the problem, thus rendering a formal hypothesis test pointless (Johnson, 1999). SCR models are such an example. In SCR models, we will see that “spatial location” of individuals is an element of the model. The simpler, reduced, model is an ordinary capture-recapture model which is not spatially explicit (i.e., chapter 3), but it seems silly and pointless to think about actually using the reduced model even if we could concoct some statistical test to refute the more complex model. The simpler model is manifestly wrong but, more importantly, not even a plausible data-generating model! Other examples are when effort, area or sample rate is used as a covariate. One might prefer to have such things in models regardless of whether or not they pass some statistical litmus test (although one can always find referees to argue for pedantic procedure over thinking).

Many problems can be approached using one of these methods but there are also broad classes of problems that can’t and, for those, you’re on your own. In later chapters we will address model selection in specific contexts and we hope those will prove useful for a majority of the situations you encounter.

2.9 POISSON GLMS

The Poisson GLM (also known as “Poisson regression”) is probably the most relevant and important class of models in all of ecology. The basic model assumes observations $y_i; i = 1, 2, \dots, n$ follow a Poisson distribution with mean λ which we write

$$y_i \sim \text{Poisson}(\lambda)$$

⁸see Royle 2008 paper for reference

Commonly y_i is a count of animals or plants at some point in space and λ_i might depend on i . For example, i might index point count locations in a forest, BBS route centers, or sample quadrats, or similar. If covariates are available it is typical to model them as linear effects on the log mean. If $x(i)$ is some measured covariate associated with observation i . Then,

$$\log(x(i)) = \alpha + \beta * x(i)$$

While we only specify the mean of the Poisson model directly, the Poisson model (and all GLMs) has a “built-in” variance which is directly related to the mean. In this case, $\text{Var}(y) = E(y) = \lambda$. Thus the model accommodates a linear increase in variance with the mean.

2.9.1 Important properties of the Poisson distribution

There are two properties of the Poisson distribution that make it extremely useful in ecology. First is the property of *compound additivity*. If y_1 and y_2 are Poisson random variables with means λ_1 and λ_2 , then their sum $N = y_1 + y_2$ is Poisson with mean $\lambda_1 + \lambda_2$. Thus, if the observations can be viewed as an aggregate of counts over some finer unit of measurement, then the mean aggregates in a corresponding manner. Secondly, the Poisson distribution has a direct relationship to the multinomial. If y_1 and y_2 are *iid* Poisson then, conditional on their sum $N = y_1 + y_2$, their joint distribution is multinomial with sample size N and cell probabilities $\lambda_1/(\lambda_1 + \lambda_2)$ and $\lambda_2/(\lambda_1 + \lambda_2)$. As a result of this, most multinomial models can be analyzed as a Poisson GLM and *vice versa*.

2.9.2 Example: Breeding Bird Survey Data

As an example we consider a classical situation in ecology where counts of an organism are made at a collection of spatial locations. In this particular example, we have mourning dove counts made along North American Breeding Bird Survey (BBS) routes in Pennsylvania, USA. A route consists of 50 stops separated by 0.5 mile. For the purposes here we are defining y_i = route total count and the sample location will be marked by the center point of the BBS route. The survey is run annually and the data set we have is 1966-1998. BBS data can be obtained online at <http://www.pwrc.usgs.gov/bbs/>. We will make use of the whole data set shortly but for now we’re going to focus on a specific year of counts – 1990 – for the sake of building a simple model. For 1990 there were 77 active routes. We have the data stored in a .csv file⁹ where rows index the unique route, column 1 is the route ID, columns 2-3 are the route coordinates (longitude/latitude), column 4 is a habitat covariate “forest cover” (standardized, see below) and the remaining columns are the yearly counts. Years for which a route was not run are coded as “NA” in the data matrix. We imagine that this will be a typical format for many ecological

⁹check this data format

1043 studies, perhaps with more columns representing covariates. To read in the data
1044 and display the first few elements of this matrix, do this:

```
1045 > a<-read.csv("pa-bbsdovedata-all.csv")
1046 > data[1:2,1:6]
1047      X      lon      lat      habitat X66 X67
1048 1 72002 -80.445 41.501 -0.3871372 NA 24
1049 2 72003 -80.347 41.214 -1.0171629 NA NA
```

1050 It is useful to display the spatial pattern in the observed counts. For that we
1051 use a spatial dot plot - where we plot the coordinates of the observations and mark
1052 the color of the plotting symbol based on the magnitude of the count. We have a
1053 special plotting function for that which is called `spatial.plot()` and it is available
1054 with the supplemental **R** package. Actually, what we want to do here is plot the
1055 log-count (+1 of course) which (Fig. 2.1) displays a notable pattern that could be
1056 related to something. The **R** commands for obtaining this figure are:

```
1057 data<-read.csv("pa-bbsdovedata-all.csv")
1058 y<-data[,29] # pick out 1990
1059 notna<-!is.na(y)
1060 y<-y[notna]
1061 spatial.plot(data[notna,2:3],y)
```

1062 We can ponder the potential effects that might lead to dove counts being
1063 high....corn fields, telephone wires, barn roofs along with misidentification of pi-
1064 geons, these could all correlated reasonably well with the observed count of mourn-
1065 ing doves. Unfortunately we don't have any of that information.

1066 We do have a measure of forest cover in the vicinity of each point which is
1067 contained in the data set (variable "habitat"). This was derived from a larger GIS
1068 coverage of the state (provided in the data file "pahabdata.csv") which can be
1069 plotted using the `spatial.plot` function using the following commands

```
1070 > map('state',regions="penn",lwd=2)
1071 > spatial.plot(pahabdata[,2:3],pahabdata[, "dfor"],cx=2)
1072 > map('state',regions="penn",lwd=2,add=TRUE)
```

1073 where the result appears in Fig. 2.2. We see a prominent pattern that indicates
1074 high forest coverage in the central part of the state and low forest cover in the SE.
1075 Inspecting the previous figure of log-counts suggests a relationship between counts
1076 and forest cover which is perhaps not surprising.

1077 2.9.3 Doing it in WinBUGS

1078 Here we demonstrate how to fit a Poisson GLM in **WinBUGS** using the covariate
1079 x_i = forest cover. It is advisable that x_i be standardized in most cases as this will
1080 improve mixing of the Markov chains. Recall that the data we have stored include
1081 a standardized covariate (forest cover) and so we don't have to worry about that
1082 here. To read the BBS data into **R** and get things set up for **WinBUGS** we issue
1083 the following commands:

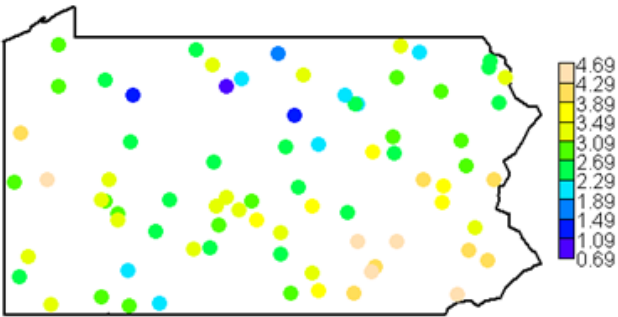


Figure 2.1. Needs a caption

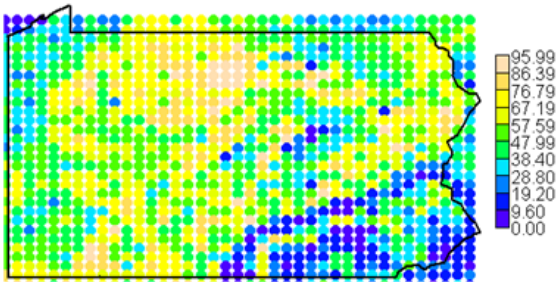


Figure 2.2. Needs a caption

```

1084 data<-read.csv("pa-bbsdovedata-all.csv")
1085 y<-data[,29] # pick out 1990
1086 notna<-!is.na(y)
1087 y<-y[notna] # discard missing
1088 habitat<-data[notna,4] # get habitat data
1089 library("R2WinBUGS") # load R2WinBUGS
1090 data <- list ( "y","M","habitat") # bundle data for WinBUGS

```

Now we write out the Poisson model specification in **WinBUGS** pseudo-code, provide initial values, identify parameters to be monitored and then execute **WinBUGS**:

```

1094 cat("
1095 model {
1096   for (i in 1:M){
1097     y[i]~dpois(lam[i])
1098     log(lam[i])<- beta0+beta1*habitat[i]
1099   }
1100   beta0~dunif(-5,5)
1101   beta1~dunif(-5,5)
1102 }
1103 ",file="PoissonGLM.txt")
1104
1105 inits <- function() list ( beta0=rnorm(1),beta1=rnorm(1))
1106 parameters <- c("beta0","beta1")
1107 out<-bugs (data, inits, parameters, "PoissonGLM.txt", n.thin=2,n.chains=2,
1108           n.burnin=2000,n.iter=6000,debug=TRUE,working.dir=getwd())

```

Remarks: (1) Note the close correspondence in how the model is specified here compared with the normal regression model previously. As an exercise you should discuss the specific differences between the **BUGS** model specifications for the normal and Poisson models.

```

1113 > print(out,digits=3)
1114 Inference for Bugs model at
1115 ‘‘PoissonGLM.txt’’, fit using WinBUGS,
1116 2 chains, each with 4000 iterations (first 1000 discarded), n.thin = 2
1117 n.sims = 3000 iterations saved

```

	mean	sd	2.5%	25%	50%	75%	97.5%	Rhat	n.eff
beta0	3.151	0.025	3.102	3.135	3.151	3.168	3.199	1.001	2300
beta1	-0.498	0.021	-0.539	-0.512	-0.498	-0.484	-0.457	1.001	3000
fit	869.930	19.856	835.500	855.700	868.600	881.900	913.602	1.002	1600
fitnew	76.709	12.519	54.098	68.107	76.215	84.510	102.602	1.001	3000
deviance	1116.605	2.014	1115.000	1115.000	1116.000	1117.000	1122.000	1.001	3000

We might wonder whether this model provides an adequate fit to our data. To evaluate that, we used a Bayesian p-value analysis with fit statistic based on the Freeman-Tukey residual by replacing the model specification above with this:

```

1127 cat("
1128 model {
1129     for (i in 1:M){
1130         y[i]~dpois(lam[i])
1131         log(lam[i])<- beta0+beta1*habitat[i]
1132         d[i]<- pow(pow(y[i],0.5)-pow(lam[i],0.5),2)    #
1133
1134         ynew[i]~dpois(lam[i])
1135         dnew[i]<-pow( pow(ynew[i],0.5)-pow(lam[i],0.5),2)
1136
1137     }
1138     fit<-sum(d[])
1139     fitnew<-sum(dnew[])
1140     beta0~dunif(-5,5)
1141     beta1~dunif(-5,5)
1142 }
1143 ",file="PoissonGLM.txt")

```

1144 The Bayesian p-value is the proportion of times $fitnew > fit$ which, for this
 1145 data set, is 0, which was 1.0 in this case (calculation omitted). This suggests that
 1146 the basic Poisson model does not fit well.

1147 2.9.4 Constructing your own MCMC algorithm

1148 It might be helpful to suffer through a couple examples building custom MCMC
 1149 algorithms. Here, we develop an MCMC algorithm for the Poisson regression model,
 1150 using a Metropolis-within-Gibbs sampling framework.

1151 We will assume that the two parameters have diffuse normal priors, say $[\alpha] =$
 1152 $\text{Norm}(0, 100)$ and $[\beta] = \text{Norm}(0, 100)$ where each has *standard deviation* 100 (recall
 1153 that **WinBUGS** parameterizes the normal in terms of $1/\sigma^2$). We need to assemble
 1154 the relevant elements of the model which are these two prior distributions and the
 1155 likelihood $[\mathbf{y}|\alpha, \beta] = \prod_i [y_i|\alpha, \beta]$ which is, mathematically, the product of the Poisson
 1156 pmf evaluated at each y_i , given particular values of α and β . Next, we need to
 1157 identify the full conditionals $[\alpha|\beta, \mathbf{y}]$ and $[\beta|\alpha, \mathbf{y}]$. We use the all-purpose rule for
 1158 constructing full conditionals (section 2.5.1) to discover that:

$$[\alpha|\beta, \mathbf{y}] \propto \left\{ \prod_i [y_i|\alpha, \beta] \right\} [\alpha]$$

1159 and

$$[\beta|\alpha, \mathbf{y}] \propto \left\{ \prod_i [y_i|\alpha, \beta] \right\} [\beta]$$

1160 Remember, we could replace the “ \propto ” with “ $=$ ” if we put $[y|\beta]$ or $[y|\alpha]$ in the de-
 1161 nominator. But, in general, $[y|\alpha]$ or $[y|\beta]$ will be quite a pain to compute and, more

importantly, it is a constant as far as the operative parameters (α or β , respectively) are concerned. Therefore, the MH acceptance probability will be the ratio of the full-conditional evaluated at a candidate draw to that evaluated at the current draw, and so the denominator required to change \propto to $=$ winds up canceling from the MH acceptance probability. Here we will use the random walk candidate generator so that, for example, $\alpha^* \sim \text{Normal}(\alpha^t, \delta)$ where δ is the standard-deviation of the proposal distribution, which is just a tuning parameter¹⁰. We remark also that calculations are often done on the log-scale to preserve numerical integrity of things when quantities evaluate to small or large numbers, so keep in mind, for example, $a*b = \exp(\log(a) + \log(b))$. The “Metropolis within Gibbs” algorithm for a Poisson regression turns out to be remarkably simple:

```

1173 set.seed(2013)
1174
1175 out<-matrix(NA,nrow=1000,ncol=2)  # matrix to store the output
1176 alpha<- -1                        # starting values
1177 beta <- -.8
1178
1179 # begin the MCMC loop ; do 1000 iterations
1180 for(i in 1:1000){
1181
1182   # update the alpha parameter
1183   lambda<- exp(alpha+beta*habitat)
1184   lik.curr<- sum(log(dpois(y,lambda)))
1185   prior.curr<- log(dnorm(alpha,0,100))
1186   alpha.cand<-rnorm(1,alpha,.25)      # generate candidate
1187   lambda.cand<- exp(alpha.cand + beta*habitat)
1188   lik.cand<- sum(log(dpois(y,lambda.cand)))
1189   prior.cand<- log(dnorm(alpha.cand,0,100))
1190   mhratio<- exp(lik.cand +prior.cand - lik.curr-prior.curr)
1191   if(runif(1)< mhratio)
1192     alpha<-alpha.cand
1193
1194   # update the beta parameter
1195   lik.curr<- sum(log(dpois(y,exp(alpha+beta*habitat))))
1196   prior.curr<- log(dnorm(beta,0,100))
1197   beta.cand<-rnorm(1,beta,.25)
1198   lambda.cand<- exp(alpha+beta.cand*habitat)
1199   lik.cand<- sum(log(dpois(y,lambda.cand)))
1200   prior.cand<- log(dnorm(beta.cand,0,100))
1201   mhratio<- exp(lik.cand + prior.cand - lik.curr - prior.curr)
1202   if(runif(1)< mhratio)
1203     beta<-beta.cand
1204

```

¹⁰ It would help lots of people out to see a non-symmetric proposal distribution, and the extra step needed to account for it.

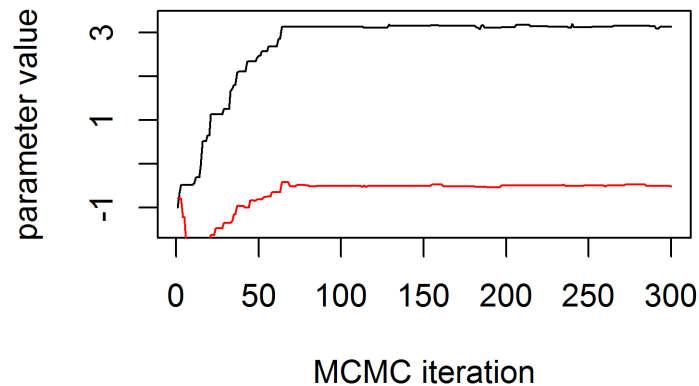


Figure 2.3. MCMC output for Poisson regression parameters (top trace: intercept α ; bottom trace: slope β). This is for $\delta = 0.25$.

```

1205 out[i,]<-c(alpha,beta)          # save the current values
1206 }
1207
1208
1209 plot(out[,1],ylim=c(-1.5,3.3),type="l",lwd=2,ylab="parameter value",
1210      xlab="MCMC iteration")
1211 lines(out[,2],lwd=2,col="red")

```

1212 The first 300 iterations of the MCMC history of each parameter is shown in
 1213 Fig. 2.3. The appearance of this is not very appealing but a couple of things are
 1214 evident: First, the Markov chains clearly stabilize - “burn-in” - after about 60 or
 1215 70 iterations. They also appear to mix very slowly once convergence is achieved,
 1216 although this is not so clear given the scale of the y -axis. We decreased the standard
 1217 deviation of the candidate generating distribution from $\delta = 0.25$ to $\delta = 0.05$ and
 1218 re-ran the MCMC algorithm producing the output shown in Fig. 2.4. We see
 1219 that the burn-in takes longer but it seems to mix better although it takes slightly
 1220 longer to burn-in. Using this value of δ we generated 10,000 posterior samples,
 1221 discarding the first 500 as burn-in, and the result is shown in Fig. 2.5, this time
 1222 separate panels for each parameter. The “grassy” look of the MCMC history is
 1223 diagnostic of Markov chains that are well-mixing and we would generally be very

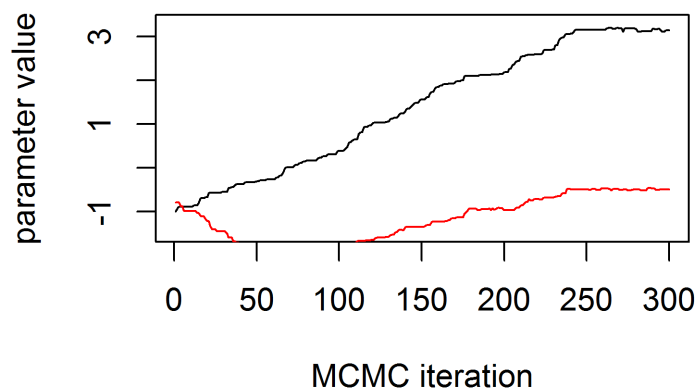


Figure 2.4. Same as previous fig but with $\delta = 0.05$.

1224 satisfied with results that look like this.

1225 **Remarks:** We used a specific set of starting values for these simulations. It
 1226 should be clear that starting values closer to the mass of the posterior distribution
 1227 might cause burn-in to occur faster. As an exercise, evaluate that. (2) Clearly
 1228 the influence of the proposal standard deviation term is important. Small values
 1229 lead to much better mixing but it should be noted that values that are too small
 1230 will slow down burn-in and also lead to high correlation. This suggests there is an
 1231 optimal value of the Metropolis-Hastings tuning parameter¹¹. As an exercise you
 1232 should contemplate finding that optimal value for this problem¹² (3) For the flat
 1233 normal prior distributions here we could leave the prior contribution out of the full
 1234 conditional evaluation since it is locally constant, i.e., constant in the vicinity of the
 1235 posterior mass, and thus has no practical effect. Removing the prior contribution
 1236 from the MH acceptance probability is equivalent to saying that the parameters
 1237 have an improper uniform prior, i.e., $\alpha \sim \text{const}$, which is commonly used for mean
 1238 parameters in practice. Note also that we have used a different prior than in our
 1239 **WinBUGS** model specification given previously. As an exercise, evaluate whether
 1240 this seems to affect the result.

¹¹Defined previously?????

¹²effective sample size definition?

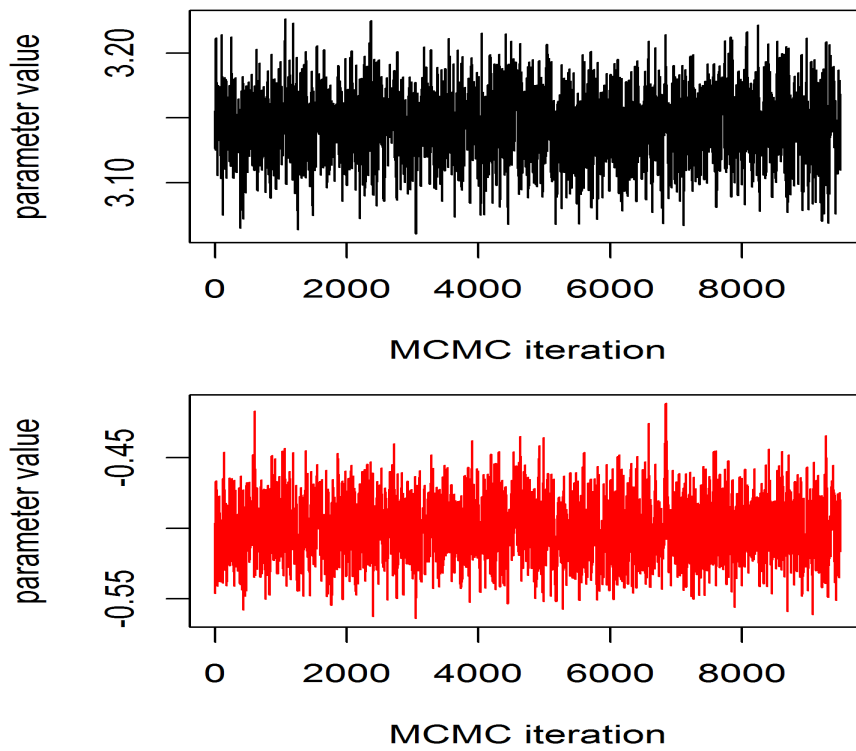


Figure 2.5. nice grassy mcmc output, longer run of previous with $\delta = 0.05$.

2.10 POISSON GLM WITH RANDOM EFFECTS

What we will be doing in most of this book is dealing with random effects in GLM-like models - similar to what are usually referred to as generalized linear mixed models (GLMMs). We provide a brief introduction by way of example, extending our Poisson regression model to include a random effect.

ANDY STOPPED HERE

The Log-Normal mixture: The classical situation involves a GLM with a normally distributed random effect that is additive on the linear predictor. For the Poisson case, we have:

$$\log(\lambda_i) = \alpha + \beta x_i + \eta_i$$

where $\eta_i \sim \text{Normal}(0, \sigma^2)$. A natural alternative is to have multiplicative gamma-distributed noise, $\exp(\eta_i) \sim \text{Gamma}(a, b)$ which would correspond to a negative binomial kind of over-dispersion, implying a different mean/variance relationship to the log-normal mixture (the interested reader should work that out). Choosing between such possibilities is not a topic we will get into here because it doesn't seem possible to provide general guidance on it. For this model we carried-out a goodness-of-fit evaluation using the Bayesian p-value based on a Pearson residual statistic. See also (Kéry, 2010, ch. 18) for an example involving a binomial mixed model¹³. Anyhow, it is really amazingly simple to express this model in **WinBUGS** and have **WinBUGS** draw samples from the posterior distribution using the following code for the BBS dove counts:

```
data<-read.csv("pa-bbsdovedata-all.csv")
locs<-data[,2:3]
habitat<-data[,4]
y<-data[,29]      # grab year 1990
M<-length(y)

set.seed(2013)

cat("
model {
  for (i in 1:M){
    y[i]~dpois(lam[i])
    log(lam[i])<- alpha+ beta*habitat[i] + eta[i]
    frog[i]<-beta*habitat[i] + eta[i]
    eta[i] ~ dnorm(0,tau)
    d[i]<- pow(pow(y[i],0.5)-pow(lam[i],0.5),2)

    ynew[i]~dpois(lam[i])
    dnew[i]<- pow(pow(ynew[i],0.5)-pow(lam[i],0.5),2)
  }
  fit<-sum(d[])
```

¹³Kéry has noticed that such tests probably have 0 power. Should use the marginal frequency of the data

```

1281 fitnew<-sum(dnew[])
1282
1283 alpha~dunif(-5,5)
1284 beta~dunif(-5,5)
1285 sigma~dunif(0,10)
1286 tau<-1/(sigma*sigma)
1287 }
1288
1289 ",file="model.txt")
1290 data <- list ( "y","M","habitat")
1291 inits <- function()
1292   list ( alpha=rnorm(1),beta=rnorm(1),sigma=runif(1,0,4))
1293 parameters <- c("alpha","beta","sigma","tau","fit","fitnew")
1294 library("R2WinBUGS")
1295
1296 out<-bugs (data, inits, parameters, "model.txt", n.thin=2,n.chains=2,
1297   n.burnin=1000,n.iter=5000,debug=TRUE)

```

1298 This produces the following posterior summary statistics:

```

1299 > print(out,digits=2)
1300 Inference for Bugs model at "model.txt", fit using WinBUGS,
1301 2 chains, each with 5000 iterations (first 1000 discarded), n.thin = 2
1302 n.sims = 4000 iterations saved
1303      mean    sd  2.5%   25%   50%   75%  97.5% Rhat n.eff
1304 alpha    2.98 0.08  2.82  2.93  2.98  3.03  3.12 1.00  1400
1305 beta   -0.53 0.07 -0.68 -0.58 -0.53 -0.49 -0.38 1.01   350
1306 sigma    0.60 0.06  0.49  0.56  0.59  0.64  0.73 1.00  2000
1307 tau     2.88 0.57  1.88  2.47  2.86  3.24  4.12 1.00  2000
1308 fit     26.58 3.72 19.87 23.96 26.37 29.01 34.46 1.00  4000
1309 fitnew   26.83 3.90 19.60 24.12 26.68 29.36 35.04 1.00  4000
1310 deviance 445.94 12.18 424.00 437.40 445.20 453.90 471.50 1.00  4000
1311
1312 [... some output deleted ...]

```

1313 The Bayesian p-value for this model is

```

1314 > mean(out$sims.list$fit>out$sims.list$fitnew)
1315 [1] 0.4815

```

1316 indicating a pretty good fit. Given the site-level random effect, it would be surprising for this model to not fit! One thing we notice is that the posterior standard deviations of the regression parameters are much higher, a result of the excess variation. We would also notice much less precise predictions of hypothetical new observations.

1320 ANDY STOPPED HERE.

2.11 BINOMIAL GLMS

Another extremely important class of models in ecology are binomial models. We use binomial models for count data whenever the observations are counts or frequencies and it is natural to condition on a “sample size”, say K , the maximum frequency possible in a sample. The random variable, $y \leq K$, is then the frequency of occurrences out of K “trials”. The parameter of the binomial models is p , often called “success probability” which is related to the expected value of y by $E(y) = pK$. Usually we are interested in modeling covariates that affect the parameter p , and such models are called binomial GLMs, binomial regression models or logistic regression, although logistic regression really only applies when the logistic link is used to model the relationship between p and covariates (see below).

One of the most typical binomial GLMs occurs when the sample size equals 1 and the outcome, y , is “presence” ($y = 1$) or “absence” ($y = 0$) of a species. This is a classical “species distribution” modeling situation. A special situation occurs when presence/absence is observed with error (MacKenzie et al., 2002; Tyre et al., 2003). In that case, $K > 1$ samples are usually needed for effective estimation of model parameters.

In standard binomial regression problems the sample size is fixed by design but interesting models also arise when the sample size is itself a random variable. These are the N -mixture models (Royle, 2004; Kéry et al., 2005; Royle and Dorazio, 2008; Kéry, 2010) and related models (in this case, N being the sample size, which we labeled K above)¹⁴. Another situation in which the binomial sample size is “fixed” is closed population capture-recapture models in which a population of individuals is sampled K times. The number of times each individual is encountered is a binomial outcome with parameter - encounter probability - p , based on a sample of size K . In addition, the total number of unique individuals observed, n , is also a binomial random variable based on population size N . We consider such models in the chapter 3.

2.11.1 Binomial regression

In binomial models, covariates are modeled on a suitable transformation (the link function) of the binomial success probability, p . Let x_i denote some measured covariate for sample unit i and let p_i be the success probability for unit i . The standard choice is the “logit” link function which is:

$$\log(p_i/(1 - p_i)) = \alpha + \beta * x_i.$$

¹⁴Some of the jargon is actually a little bit confusing here because the binomial index is customarily referred to as “sample size” but in the context of N -mixture models N is actually the “population size”

1354 The inverse-logit (or “expit”) is

$$p_i = \text{expit}(\alpha + \beta * x_i) = \frac{\exp(\alpha + \beta * x_i)}{1 + \exp(\alpha + \beta * x_i)}$$

1355 There are many other possible link functions. However, ecologists seem to adopt
 1356 the logit link function without question in most applications¹⁵. We sometimes use
 1357 the “complementary log-log” (= “cloglog”) link function in ecological applications
 1358 because it arises naturally in many situations (Royle and Dorazio, 2008, p. 150).
 1359 For example, consider the “probability of observing a count greater than 0” under
 1360 a Poisson model: $\Pr(y > 0) = 1 - \exp(-\lambda)$. In that case,

$$\text{cloglog}(p) = \log(-\log(1 - p)) = \log(\lambda)$$

1361 So that if you have covariates in your linear predictor for $E(y)$ under a Poisson
 1362 model then they are linear on the complementary log-log link of p . In models of
 1363 species occurrence it seems natural to view occupancy as being derived from local
 1364 abundance N (Royle and Nichols, 2003; Royle and Dorazio, 2006; Dorazio, 2007).
 1365 Therefore, models of local abundance in which $N \sim \text{Poisson}(A\lambda)$ for a habitat patch
 1366 of area A implies a model for occupancy ψ of the form

$$\text{cloglog}(\psi) = \log(A) + \log(\lambda).$$

1367 We will use the cloglog link in some analyses of SCR models in chapter 4 and
 1368 elsewhere.

1369 2.11.2 Example: Waterfowl Banding Data

1370 It would be easy to consider a standard “distribution modeling” application where
 1371 $K = 1$ and the outcome is occurrence ($y = 1$) or not ($y = 0$) of some species.
 1372 Such examples abound in books (e.g., Royle and Dorazio (2008, ch. 3); Kéry (2010,
 1373 ch. 21); Kéry and Schaub (2011, ch. 13)) and in the literature. Instead, we will
 1374 consider an example involving band returns of waterfowl which were analyzed by
 1375 Royle and Dubovsky (2001)¹⁶.

1376 For these data, y_i is the number of waterfowl bands recovered out of B_i birds
 1377 banded at some location \mathbf{s}_i . In this case B_i is fixed. Thinking about recovery rate
 1378 as being proportional to harvest rate, we use these data to explore geographic gra-
 1379 dients in recovery rate resulting from variability in harvest pressure experienced by
 1380 populations depending on their migration ecology. As such, we fit a basic binomial
 1381 GLM with a linear response to geographic coordinates (including an interaction
 1382 term). The data are provided with the **R** package **scrbook**. Here we provide the
 1383 part of the script for creating the model and fitting the model in **WinBUGS** using

¹⁵a notable exception is distance sampling, which is all about choosing among link functions

¹⁶I hate this example. Anyone got a better one thats not distribution modeling?

the `bugs` function. There are few structural differences between this model and the Poisson GLM fitted previously. The main things are due to the data structure (we have a matrix here instead of a vector) and otherwise we change the main distributional assumption to binomial (specified with `dbin`) and then use the `logit` function to relate the parameter p_{it} to the covariates. Here is the script:

```

1389 load("mallarddata") # not sure how this will look
1390
1391 sink("model.txt")
1392 cat("
1393 model {
1394   for(t in 1:5){
1395     for (i in 1:nobs){
1396       y[i,t] ~ dbin(p[i,t], B[i,t])
1397       logit(p[i,t]) <- alpha0[t] + alpha1*X[i,1] + alpha2*X[i,2] + alpha3*X[i,1]*X[i,2]
1398     }
1399   }
1400   alpha1~dnorm(0,.001)
1401   alpha2~dnorm(0,.001)
1402   alpha3~dnorm(0,.001)
1403   for(t in 1:5){
1404     alpha0[t] ~ dnorm(0,.001)
1405   }
1406 }
1407 ",fill=TRUE)
1408 sink()
1409
1410 data <- list(B=mallard.bandings, y=mallard.recoveries,
1411             nobs=nrow(banding.locs),X=banding.locs)
1412 inits <- function(){
1413   list(alpha0=rnorm(5),alpha1=0,alpha2=0,alpha3=0) }
1414 parms <- list('alpha0','alpha1','alpha2','alpha3')
1415 out <- bugs(data,inits, parms,"model.txt",n.chains=3,
1416            n.iter=2000,n.burnin=1000, n.thin=2,debug=TRUE)

```

Posterior summaries of model parameters are as follows:

```

1418 > print(out,digits=3)
1419 Inference for Bugs model at "model.txt", fit using WinBUGS,
1420 3 chains, each with 2000 iterations (first 1000 discarded), n.thin = 2
1421 n.sims = 1500 iterations saved

```

	mean	sd	2.5%	25%	50%	75%	97.5%	Rhat	n.eff
alpha0[1]	-2.346	0.036	-2.417	-2.370	-2.346	-2.323	-2.277	1.001	1500
alpha0[2]	-2.356	0.032	-2.420	-2.379	-2.356	-2.335	-2.292	1.001	1500
alpha0[3]	-2.220	0.035	-2.291	-2.244	-2.219	-2.197	-2.153	1.001	1500
alpha0[4]	-2.144	0.039	-2.225	-2.169	-2.143	-2.116	-2.068	1.000	1500
alpha0[5]	-1.925	0.034	-1.990	-1.949	-1.924	-1.901	-1.856	1.004	570

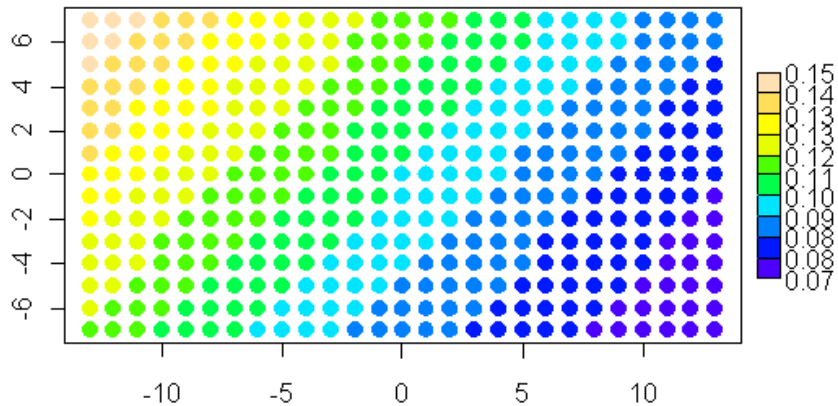


Figure 2.6. Predicted recovery rate of bands.

```
1428 alpha1      -0.023 0.003   -0.028  -0.025  -0.023  -0.022  -0.018  1.001  1500
1429 alpha2       0.020 0.006    0.009   0.016   0.020   0.024   0.031  1.001  1500
1430 alpha3       0.000 0.001   -0.002  -0.001   0.000   0.000   0.002  1.001  1500
1431 deviance  1716.001 4.091 1710.000 1713.000 1715.000 1718.000 1726.000 1.001  1500
1432
1433 [... some output deleted ...]
```

1434 The basic result suggests a negative east-west gradient and a positive south to
1435 north gradient but no interaction. A map of the response surface is shown in Fig.
1436 2.6. We did an additional MCMC run where we saved the binomial parameter
1437 p and computed the Bayesian p-value (double use of “p” here is confusing, but I
1438 guess that happens sometimes!) using a fit statistic based on the Freeman-Tukey
1439 statistic (see Section XXX above). The result indicates that the linear response
1440 surface model does not provide an adequate fit of the data. The reader should
1441 contemplate whether this invalidates the basic interpretation of the result.

2.12 SUMMARY AND OUTLOOK

1442 GLMs and GLMMs are the most useful statistical methods in all of ecology. The
1443 principles and procedures underlying these methods are relevant to nearly all mod-
1444 eling and analysis problems in every branch of ecology. Moreover, understanding
1445 how to analyze these models is crucial in a huge number of diverse problems. If you

1446 understand and can conduct classical likelihood and Bayesian analysis of Poisson
1447 and binomial GLM(M)s, then you will be successful analyzing and understanding
1448 more complex classes of models that arise. We will see shortly that spatial capture-
1449 recapture models are a type of GLMM and thus having a basic understanding of
1450 the conceptual origins and formulation of GLM(M)s and their analysis is extremely
1451 useful.

1452 We note that GLM(M)s are routinely analyzed by likelihood methods but we
1453 have focused on Bayesian analysis here in order to develop the tools that are less
1454 familiar to most ecologists. In particular, Bayesian analysis of models with ran-
1455 dom effects is relatively straightforward because the models are easy to analyze
1456 conditional on the random effect, using methods of MCMC. Thus, we will often
1457 analyze SCR models in later chapters by MCMC, explicitly adopting a Bayesian
1458 inference framework. In that regard, the various **BUGS** engines (**WinBUGS**,
1459 **OpenBUGS**, **JAGS**) are enormously useful because they provide an accessible
1460 platform for carrying out analyses by MCMC by just describing the model, and not
1461 having to worry about how to actually build MCMC algorithms. That said, the
1462 **BUGS** language is more important than just to the extent that it enables one to
1463 do MCMC - it is useful as a modeling tool because it fosters understanding, in the
1464 sense that it forces you to become intimate with your model. You have to write
1465 down all of the probability assumptions, the relationships between observations and
1466 latent variables and parameters. This is really a great learning paradigm that you
1467 can grow with.

1468 While we have emphasized Bayesian analysis in this chapter, and make primary
1469 use of it through the book, we will provide an introduction to likelihood analysis
1470 in chapter 6 and use those methods also from time to time. Before getting to that,
1471 however, it will be useful to talk about more basic, conventional closed population
1472 capture-recapture models and these are the topic of the next chapter.

3

ESTIMATING THE SIZE OF A CLOSED POPULATION

In this chapter we will consider ordinary capture-recapture (CR) models for estimating population size in closed populations. We will see that such models are closely related to binomial (or logistic) regression type models. In fact, when N is known, they are precisely such models. We consider some important extensions of ordinary closed population models that accommodate various types of “individual effects” — either in the form of explicit covariates (sex, age, body mass) or unstructured “heterogeneity” in the form of an individual random effect. In general, these models are variations of generalized linear or generalized linear mixed models (GLMMs). Because of the paramount importance of this concept, we focus mainly on fairly simple models in which the observations are individual encounter frequencies, y_i = the number of encounters of individual i out of K replicate samples of the population which, for the models we consider here, is the outcome of a binomial random variable. Along the way, we consider the spatial context of capture-recapture data and models and demonstrate that density cannot be formally estimated when spatial information is ignored. We also review some of the informal methods of estimating density using CR methods, and consider some of their limitations. We will be exposed to our first primitive spatial capture-recapture models which arise as relatively minor variations of so-called “individual covariate models” (of the Huggins (1989) and Alho (1990) variety). In a sense, the point of this chapter is to establish that linkage in a direct and concise manner beginning with the basic “Model M0” and extensions of that model to include individual heterogeneity and also individual covariates. A special type of individual covariate models is distance sampling, which could be thought of as the most primitive spatial capture-recapture model. In later chapters we further develop and extend

ideas introduced in this chapter.

We emphasize Bayesian analysis of capture-recapture models and we accomplish this using a method related to classical “data augmentation” from the statistics literature Tanner and Wong (e.g., 1987)). This is a general concept in statistics but, in the context of capture-recapture models where N is unknown, it has a consistent implementation across classes of capture-recapture models and one that is really convenient from the standpoint of doing MCMC (Royle et al., 2007). We use data augmentation throughout this book and thus emphasize its conceptual and technical origins and demonstrate applications to closed population models. We refer the reader to Kery and Schaub (2011, ch. 6) for an accessible and complimentary development of ordinary closed population models.

3.1 THE SIMPLEST CLOSED POPULATION MODEL: MODEL M0

We suppose that there exists a population of N individuals which we subject to repeated sampling, say over K nights, where individuals are captured, marked, and subsequently recaptured. We suppose that individual encounter histories are obtained, and these are of the form of a sequence of 0’s and 1’s indicating capture ($y = 1$) or not ($y = 0$) during any sampling occasion (“sample”). As an example, suppose $K = 5$ sampling occasions, then an individual captured during sample 2 and 3 but not otherwise would have an encounter history of the form $\mathbf{y} = (0, 1, 1, 0, 0)$. Thus, the observation \mathbf{y}_i for each individual (i) is a vector having elements denoted by y_{ik} for $k = 1, 2, \dots, K$. Usually this is organized as a row of a matrix with elements y_{ik} , see Table 3.1. Except where noted explicitly, we suppose that observations are independent within individuals and among individuals. Formally, this allows us to say that y_{ik} are Bernoulli random variables and we may write $y_{ik} \sim \text{Bern}(p)$. Consequently, for this very simple model in which p is in fact constant, then we can declare that the individual encounter frequencies (total captures), $y_i = \sum_k y_{ik}$, have a binomial distribution based on a sample of size K . That is

$$y_i = \sum_k y_{ik} \sim \text{Bin}(p, K)$$

for every individual in the population. This is a remarkably simple model that forms the cornerstone of almost all of classical capture-recapture models, including most spatial capture-recapture models discussed throughout this book. Evidently, the basic capture-recapture model structure is precisely a simplistic version of a logistic-regression model with only an intercept term ($\text{logit}(p) = \text{constant}$). To say that all capture-recapture models are just logistic regressions is only slightly inaccurate. In fact, we are proceeding here “conditional on N ”, i.e., as if we knew N . In practice we don’t, of course, and that is kind of the point of capture-recapture models as estimating N is the central objective. But, by proceeding conditional on N , we can specify a simple model and then deal with the fact that N is unknown

Table 3.1. a capture-recapture data set with $n = 6$ observed individuals and $K = 5$ samples.

indiv i	Sample occasion					y_i
	1	2	3	4	5	
1	1	0	0	1	0	2
2	0	1	0	0	1	2
3	1	0	0	1	0	2
4	1	0	1	0	1	3
5	0	1	0	0	0	1
$n = 6$	1	0	0	0	0	1

using standard methods that you are already familiar with (i.e., GLMs - see chapter 2).

Assuming individuals of the population are observed independently, the joint probability distribution of the observations is the product of N binomials

$$\begin{aligned} \Pr(y_1, \dots, y_N | p) &= \prod_{i=1}^N \text{Bin}(y_i | K, p) \\ &= \prod_{k=0}^K \pi(k)^{n_k} \end{aligned}$$

where $\pi(k) = \text{Bin}(k | K, p)$ and where $n_k = \sum_{i=1}^N I(y_i = k)$ denotes the number of individuals captured k times in K surveys. We emphasize that this is conditional on N , in which case we get to observe the $y = 0$ observations and the resulting data are just *iid* binomial counts. Because this is a binomial regression model of the variety described in chapter 2, fitting this model using a BUGS engine poses no difficulty.

The essential problem in capture-recapture, however, is that N is not known because the number of uncaptured/missing individuals (i.e., those in the zero cell that occur with probability $\pi(0)$) is unknown. Consequently, the observed capture frequencies n_k are no longer independent. Instead, their joint distribution is multinomial (e.g., see Illian (2008a, p. xyz)):

$$n_1, n_2, \dots, n_K \sim \text{Multin}(N, \pi(1), \pi(2), \dots, \pi(K)) \quad (3.1.1)$$

Note that in our notation the number of uncaptured/missing individuals is denoted by $n_0 = N - n$, where $n = \sum_{k=1}^K n_k$ denotes the total number of distinct individuals seen in the K samples.

To fit the model in which N is *unknown*, we can regard N as a parameter and maximize the multinomial likelihood directly. While direct likelihood analysis of the multinomial model is straightforward, that does not prove to be too useful in practice because we seldom are concerned with models for the aggregated encounter history frequencies. In many instances, including for spatial capture-recapture (SCR)

models, we require a formulation of the model that can accommodate individual level covariates which we address subsequently in this chapter.

3.1.1 The Spatial Context of Capture-Recapture

A common assumption made is that of population “closure” which is really just a colloquial way of saying (in part) the Bernoulli assumptions stated explicitly above. In the biological context, closure means, strictly, no additions or subtractions from the population during study. This is manifest by the statement that the encounters are independent and identically distributed (iid) Bernoulli trials. In practice, closure is usually interpreted by the manner in which potential violations of that assumption arise. In particular, two important elements of the closure assumption are “demographic” and “geographic” closure. If an individual dies then subsequent values of y_{ik} are clearly no longer Bernoulli trials with the same parameter p . If there is no mortality or recruitment in the population, then we say that demographic closure is satisfied. Similarly, animals may emigrate or immigrate. If they do not, then geographic closure is satisfied. Sometimes a distinction is made between temporary and permanent emigration or immigration. That is a relevant distinction in spatial capture-recapture models, because SCR models explicitly accommodate “temporary emigration” of a certain type, due to individuals moving about their home range. The demographic closure assumption can also be relaxed using SCR models, but we will save that discussion for chapter 4.

3.1.2 Conditional likelihood

We saw that a basic closed population model is a simple logistic regression model if N is known and, when N is unknown, the model is multinomial with index or sample size parameter N . This multinomial model, being conditional on N , is sometimes referred to as the “joint likelihood” the “full likelihood” or the “unconditional likelihood” (or model in place of likelihood). This formulation differs from the so-called “conditional likelihood” approach in which the likelihood of the observed encounter histories is devised conditional on the event that an individual is captured at least once. To construct this likelihood, we have to recognize that individuals appear or not in the sample based on the value of the random variable y_i , that is, we capture them if and only if $y_i > 0$. The observation model is therefore based on $\Pr(y|y > 0)$. For the simple case of Model M0, the resulting conditional distribution is a “zero truncated” binomial distribution which accounts for the fact that we cannot observe the value $y = 0$ in the data set (see Royle and Dorazio, 2008, section XYZ). Both the conditional or unconditional models are legitimate modes of analysis in all capture-recapture types of studies, and they provide equally valid descriptions of the data and for many practical purposes provide equivalent inferences, at least in large sample sizes (Sanathanan, 1972).

Mode of analysis	parameters in model	statistical model
Joint likelihood	p, N	multinomial with index N
Conditional likelihood	p	zero-truncated binomial
Data augmentation	p, ψ	zero-inflated binomial

Table 3.2. Modes of analysis of capture-recapture models.

In this book we emphasize Bayesian analysis of capture-recapture models using data augmentation (discussed subsequently), which produces yet a third distinct formulation of capture-recapture-models based on the zero-*inflated* binomial distribution that we describe in the next section. Thus, there are 3 distinct formulations of the model – or models of analysis – for analyzing all capture-recapture models based on the (1) binomial model for the joint or unconditional specification; (2) zero-truncated binomial that arises “conditional on n ”; and (3) the zero-inflated binomial that arises under data augmentation. Each formulation has a distinct complement of model parameters (shown in Table 3.2 for Model M0).

3.2 DATA AUGMENTATION

We consider a method of analyzing closed population models using data augmentation (DA) which is useful for Bayesian analysis and, in particular, analysis of models using the various BUGS engines and other software. Data augmentation is a general statistical concept that is widely used in statistics in many different settings. The classical reference is Tanner and Wong (1987) but see also Liu and Wu (1999). Data augmentation can be adapted to provide a very generic framework for Bayesian analysis of capture-recapture models with unknown N . This idea was introduced for closed populations by Royle et al. (2007), and has subsequently been applied to a number of different contexts including individual covariate models (Royle, 2009), open population models (Royle and Dorazio, 2008, 2010; Gardner et al., 2010), spatial capture-recapture models (Royle and Young, 2008; Royle, 2010; Gardner, 2009), and many others.

Conceptually, data augmentation takes the data you wish you had - that is, the data set with N rows - the known- N data set - and embeds that data set into a larger data set having $M > N$ rows.¹ It is always possible, in practice, to choose M pretty easily for a given problem and context. Then, under data augmentation, analysis is focused on the “augmented data set.” That is, we analyze the bigger data set - the one having M rows - with an appropriate model that accounts for the augmentation. Inference is focused directly on estimating the proportion $\psi = E[N]/M$, instead of directly on N , where ψ is the “data augmentation parameter.”

¹ RC: Might be just me, but I find that formulation a little confusing... I think it's the 'data you wish you had because that's effectively data you don't have. I think it might be easier to grasp if this were explained with the data you do have - based on n .

3.2.1 DA links occupancy models and closed population models

We provide a heuristic description of data augmentation based on the close correspondence between so-called “occupancy” models and closed population models following Royle and Dorazio (2008, sec. xyz).

In occupancy models (MacKenzie et al., 2002; Tyre et al., 2003) the sampling situation is that M sites, or patches, are sampled multiple times to assess whether a species occurs at each site. This yields encounter data such as that illustrated in the left panel of Table 3.3. The important problem is that a species may occur at a site, but go undetected, yielding the “all-zero” encounter histories which are observed. However, some of the all-zeros may well correspond to sites where the species in fact *does not* occur. Thus, while the zeros are observed, there are too many of them and, in a sense, the inference problem is to allocate the zeros into “structural” (fixed) and “sampling” (or stochastic) zeros. More formally, inference is focused on the parameter ψ , the probability that a site is occupied. In contrast, in classical closed population studies, we observe a data set as in the middle panel of Table 3.3 where *no* zeros are observed. The inference problem is, essentially, to estimate how many sampling zeros there are - or should be - in a “complete” data set. The inference objective (how many sampling zeros?) is precisely the same for both types of problems if an upper limit M is specified for the closed population model. The only distinction being that, in occupancy models, M is set by design (i.e., the number of sites to visit) whereas a natural choice of M for capture-recapture models may not be obvious. However, by assuming a uniform prior for N on the integers $[0, M]$, this upper bound is induced (Royle et al., 2007). Then, one can analyze capture-recapture models by adding $M - n$ all-zero encounter histories to the data set and regarding the augmented data set, essentially, as a site-occupancy data set.

Thus, the heuristic motivation of data augmentation is to fix the size of the data set by adding *too many* all-zero encounter histories to create the data set shown in the right panel of Table 3.3 - and then analyze the augmented data set using an occupancy type model which includes both “unoccupied sites” as well as “occupied sites” at which detections did not occur. We call these $M - n$ all-zero histories “potential individuals” because they exist to be recruited (in a non-biological sense) into the population, for example during an analysis by MCMC.

To analyze the augmented data set, we recognize that it is a zero-inflated version of the known- N data set. That is, some of the augmented all-zeros are sampling zeros (corresponding to actual individuals that were missed) and some are “structural” zeros, which do not correspond to individuals in the population. For a basic closed-population model, the resulting likelihood under data augmentation - that is, for the data set of size M - is a simple zero-inflated binomial likelihood. The zero-inflated binomial model can be described “hierarchically”, by introducing a set of binary latent variables, z_1, z_2, \dots, z_M , to indicate whether each individual i is ($z_i = 1$) or is not ($z_i = 0$) a member of the population of N individuals exposed

1670 to sampling. We assume that $z_i \sim \text{Bern}(\psi)$ where ψ is the probability that an
 1671 individual in the data set of size M is a member of the sampled population - in the
 1672 sense that $1 - \psi$ is the probability of realizing a “structural zero” in the augmented
 1673 data set. The zero-inflated binomial model which arises under data augmentation
 1674 can be formally expressed by the following set of assumptions:

$$\begin{aligned} y_i | z_i = 1 &\sim \text{Bin}(K, p) \\ y_i | z_i = 0 &\sim \delta(0) \\ z_i &\stackrel{iid}{\sim} \text{Bern}(\psi) \\ \psi &\sim \text{Unif}(0, 1) \\ p &\sim \text{Unif}(0, 1) \end{aligned}$$

1675 for $i = 1, \dots, M$, where $\delta(0)$ is a point mass at $y = 0$.

1676 We note that N is no longer an explicit parameter of this model. Instead, we
 1677 estimate ψ and functions of the latent variables. In particular, under the assump-
 1678 tions of the zero-inflated model, $z_i \stackrel{iid}{\sim} \text{Bern}(\psi)$; therefore, N is a function of these
 1679 latent variables:

$$N = \sum_{i=1}^M z_i.$$

1680 Further, we note that the latent z_i parameters can be removed from the model by
 1681 integration, in which case the joint probability of the data is

$$\Pr(y_1, \dots, y_M | p, \psi) = \prod_{i=1}^M \psi \text{Bin}(y_i | K, p) + I(y_i = 0)(1 - \psi) \quad (3.2.1)$$

1682 Which can be maximized directly to obtain the MLEs of the structural parameters
 1683 ψ and p or those of other more complex models (e.g., see Royle, 2006). We could
 1684 estimate these parameters and then use them to obtain an estimator of N using
 1685 the so-called “Best unbiased predictor” (see Royle and Dorazio, 2011).

1686 3.2.2 Model M_0 in BUGS

1687 For model M_0 in which we can aggregate the encounter data to individual-specific
 1688 encounter frequencies, the augmented data are given by the vector of frequencies
 1689 $(y_1, \dots, y_n, 0, 0, \dots, 0)$. The zero-inflated model of the augmented data combines
 1690 the model of the latent variables, $z_i \sim \text{Bern}(\psi)$ with the conditional-on- z binomial
 1691 model:

$$\begin{aligned} y_i | z_i = 0 &\sim \delta(0) \\ y_i | z_i = 1 &\sim \text{Bin}(K, p) \end{aligned}$$

Table 3.3. Hypothetical occupancy data set (left), capture-recapture data in standard form (center), and capture-recapture data augmented with all-zero capture histories (right).

Occupancy data				Capture-recapture				Augmented C-R			
site	k=1	k=2	k=3	ind	k=1	k=2	k=3	ind	k=1	k=2	k=3
1	0	1	0	1	0	1	0	1	0	1	0
2	1	0	1	2	1	0	1	2	1	0	1
3	0	1	0	.	0	1	0	3	1	0	1
4	1	0	1	.	1	0	1	4	1	0	1
5	0	1	1	.	0	1	1	5	1	0	1
.	0	1	1	.	0	1	1	.	0	1	1
.	1	1	1	.	1	1	1	.	0	1	1
.	1	1	1	.	1	1	1	.	1	1	1
.	1	1	1	.	1	1	1	.	1	1	1
n	1	1	1	n	1	1	1	n	1	1	1
.	0	0	0					.	0	0	0
.	0	0	0					.	0	0	0
	0	0	0						0	0	0
	0	0	0						0	0	0
	0	0	0						0	0	0
	0	0	0					N	0	0	0
.	0	0	0					.	0	0	0
.	0	0	0						0	0	0
M	0	0	0					.	0	0	0
							
							
							
								M	0	0	0

1692 It is convenient to express the conditional-on- z observation model concisely as:

$$y_i|z_i \sim \text{Bin}(K, pz_i)$$

1693 Thus, if $z_i = 0$ then the success probability of the binomial distribution is identically
 1694 0 whereas, if $z_i = 1$, then the success probability is p . This is useful in describing
 1695 the model in the **BUGS** language, as shown below. Note the last line of the
 1696 model specification here provides the expression for computing N from the data
 1697 augmentation variables z_i .

```
1698 p ~ dunif(0,1)
1699 psi~dunif(0,1)
1700
1701 # nind = number of individuals captured at least once
1702 # nz = number of uncaptured individuals added for PX-DA
1703 for(i in 1:(nind+nz)) {
1704     z[i]~dbern(psi)
1705     mu[i]<-z[i]*p
1706     y[i]~dbin(mu[i],K)
1707 }
1708
1709 N<-sum(z[1:(nind+nz)])
```

1710 Specification of a more general model in terms of the individual encounter obser-
 1711 vations y_{ik} is not much more difficult than for the individual encounter frequencies.
 1712 We define the observation model by a double loop and change the indexing of things
 1713 accordingly, i.e.,

```
1714 for(i in 1:(nind+nz)) {
1715     z[i]~dbern(psi)
1716     for(k in 1:K){
1717         mu[i,k]<-z[i]*p
1718         y[i,k]~dbin(mu[i,k],1)
1719     }
1720 }
```

1721 In this manner, it is straightforward to incorporate covariates on p (see discussion
 1722 of this below and also chapt. 8 (REF XYZ) and consider other extensions.

1723 3.2.3 Formal development of data augmentation

1724 Use of DA for solving inference problems with unknown N can be justified as
 1725 originating from the choice of uniform prior on N . The $\text{Unif}(0, M)$ prior for N is
 1726 innocuous in the sense that the posterior associated with this prior is equal to the
 1727 likelihood for sufficiently large M . One way of inducing the $\text{Unif}(0, M)$ prior on N

1728 is by assuming the following hierarchical prior:

$$\begin{aligned} N &\sim \text{Bin}(M, \psi) \\ \psi &\sim \text{Unif}(0, 1) \end{aligned} \tag{3.2.2}$$

1729 which includes a new model parameter ψ . This parameter denotes the probability
 1730 that an individual in the super-population of size M is a member of the population
 1731 of N individuals exposed to sampling. The model assumptions, specifically the
 1732 multinomial model (eq. XYZ) and eq. 3.2.2, may be combined to yield a reparam-
 1733 eterization of the conventional model that is appropriate for the augmented data
 1734 set of known size M :

$$(n_1, n_2, \dots, n_K) \sim \text{Multin}(M, \psi\pi(1), \psi\pi(2), \dots, \psi\pi(K)) \tag{3.2.3}$$

1735 This arises by removing N from Eq. multinomial XYZ by integrating over the
 1736 binomial prior distribution for N . Thus, the models we analyze under data aug-
 1737 mentation arise formally by removing the parameter N from the ordinary model -
 1738 the model conditional on N - by integrating over a binomial prior distribution for
 1739 N .

1740 Note that the $M - n$ unobserved individuals in the augmented data set have
 1741 probability $\psi\pi(0) + (1 - \psi)$, indicating that these unobserved individuals are a
 1742 mixture of individuals that are sampling zeros ($\psi\pi_0$, and belong to the population
 1743 of size N) and others that are “structural zeros” (occurring in the augmented
 1744 data set with probability $1 - \psi$). In Eq. 3.2.3 N has been eliminated as a formal
 1745 parameter of the model by marginalization (integration) and replaced with the new
 1746 parameter ψ , which we will call the “data augmentation parameter.” However, the
 1747 full likelihood containing both N and ψ can be analyzed (see Royle et al., 2007).

1748 3.2.4 Remarks on Data Augmentation

1749 Data augmentation may seem like a strange and mysterious black-box, and likely
 1750 it is unfamiliar to most people even those with extensive experience with capture-
 1751 recapture models. However, it really is a formal reparameterization of capture-
 1752 recapture models in which N is removed from the ordinary (conditional-on- N)
 1753 model by integration. In the case of Model M0, data augmentation produces the
 1754 zero-inflated binomial which is distinct from the original observation model, but
 1755 only in the sense that it embodies, explicitly, the $\text{Unif}(0, M)$ prior for N . Choice of
 1756 M might be cause for some concern related to potential sensitivity to choice of M .
 1757 The guiding principle is that it should be chosen large enough so that the posterior
 1758 for N is not truncated, but no larger because large values entail more computational
 1759 burden. It seems likely that the properties of the Markov chains should be affected
 1760 by M and so some optimality might exist (Gopalaswamy, 2012), as in occupancy
 1761 models (Mackenzie and Royle, 2005). Formal analysis of this is required.

We emphasize the motivation for data augmentation being that it produces a data set of fixed size, so that the parameter dimension in any capture-recapture model is also fixed. As a result, MCMC is a relatively simple proposition using standard Gibbs Sampling. Consider the simplest context - analyzing Model M0 using the occupancy model. In this case, DA converts Model M0 to a basic occupancy model and the parameters p and ψ have known full-conditional distributions (in fact, beta distributions) that can be sampled from directly. Furthermore, the data augmentation variables - the latent data augmentation variables z , can be sampled from Bernoulli full conditionals. MCMC is not too much more difficult for complicated models - sometimes the hyperparameters need to be sampled using a Metropolis-Hastings step, but nothing more sophisticated than that is required.

There are other approaches to analyzing models with unknown N , using reversible jump MCMC (RJMCMC) or other so-called “trans-dimensional” (TD) algorithms² (Durbin and Elston, 2012; King, missing; Schofield and Barker, missing). What distinguishes DA from RJMCMC and related TD methods is that DA is used to create a distinctly new model that is unconditional on N and we (usually) analyze the unconditional model. The various TD/RJMCMC approaches seek to analyze the conditional-on- N model in which the dimensional of the parameter space is a variable function of N . TD/RJMCMC approaches might appear to have the advantage that one can model N explicitly or consider alternative priors for N . However, despite that N is removed as an explicit parameter in DA, it is possible to develop hierarchical models that involve structure on N (Converse and Royle, 2010; Royle et al., 2011a) which we consider in chapt. XYZ.

3.2.5 Example: Black Bear Study on Fort Drum

To illustrate the analysis of Model M0 using data augmentation, we use a data set collected at Fort Drum Military Installation in upstate New York by the Department of Defense, Cornell University and colleagues. These data have been analyzed in various forms by Gardner (2009); Gardner et al. (2010), and Wegan (2008). The specific data used here are encounter histories on 47 individuals obtained from an array of 38 baited “hair snares” (Fig. 3.1) during June and July 2006. Barbed wire traps were baited and checked for hair samples each week for eight weeks, thus we have $K = 8$ sample intervals. The data are provided on the Web Supplement and the analysis can be set up and run as follows. Here, the data were augmented with $M - n = 128$ ($M = 175$) all-zero encounter histories.

```
# Consider adding comments to your code.
## Good idea. This will be done in final draft
trapmat<-read.csv("FDtrapmat.csv")
bearArray<-source("FDbeararray.R")$value
nind<-dim(bearArray)[1]
```

²Look these citations up in Royle-Dorazio EURING paper

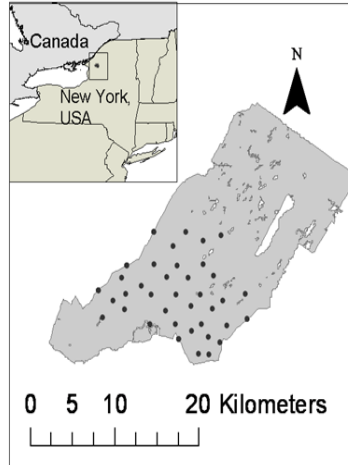


Figure 3.1. Fort Drum study area and hair snare locations.

```

1801 K<-dim(bearArray)[3]
1802 ntraps<-dim(bearArray)[2]
1803
1804 M=175
1805 nz<-M-nind
1806
1807 Xaug <- array(0, dim=c(M,ntraps,K))
1808 Xaug[1:nind,,]<-bearArray
1809 y<- apply(Xaug,c(1,3),sum)
1810 y[y>1]<-1
1811 ytot<-apply(y,1,sum) # total encounters out of K

```

Note that the raw data, \mathbf{y} , is an $M \times K$ array of individual encounter events (i.e., $y_{ik} = 1$ if individual i was encountered in any trap and 0 otherwise). For $i = 48, \dots, 175$, $y_{ik}=0$ as these are augmented observations. For Model M0 it is sufficient to reduce the data to individual encounter frequencies which we have labeled \mathbf{y}_{tot} above. The BUGS model file along with commands to fit the model are as follows:

```

1818 set.seed(2013) # to obtain the same results each time
1819 data0<-list(y=y,M=M,K=K)
1820 params0<-list('psi','p','N')
1821 zst=c(rep(1,nind),rbinom(M-nind, 1, .5))
1822 inits = function() {list(z=zst, psi=runif(1), p=runif(1)) }
1823
1824 cat("

```

```

1825 model {
1826
1827   psi~dunif(0, 1)
1828   p~dunif(0,1)
1829
1830   for (i in 1:M){
1831     z[i]~dbern(psi)
1832     for(k in 1:K){
1833       tmp[i,k]<-p*z[i]
1834       y[i,k]~dbin(tmp[i,k],1)
1835     }
1836   }
1837   N<-sum(z[1:M])
1838 }
1839 ",file="modelM0.txt")
1840
1841 fit0 = bugs(data0, inits, params0, model.file="modelM0.txt",
1842            n.chains=3, n.iter=2000, n.burnin=1000, n.thin=1,
1843            debug=TRUE,working.directory=getwd())

```

1844 The posterior summary statistics from this analysis are as follows:

```

1845 > print(fit0,digits=2)
1846 Inference for Bugs model at "modelM0.txt", fit using WinBUGS,
1847 3 chains, each with 2000 iterations (first 1000 discarded)
1848 n.sims = 3000 iterations saved
1849      mean    sd  2.5%   25%   50%   75%  97.5% Rhat n.eff
1850 psi      0.29 0.04  0.22  0.26  0.29  0.31  0.36   1 3000
1851 p        0.30 0.03  0.25  0.28  0.30  0.32  0.35   1 3000
1852 N        49.94 1.99 47.00 48.00 50.00 51.00 54.00   1 3000
1853 deviance 489.05 11.28 471.00 480.45 488.80 495.40 513.70   1 3000
1854
1855 [.. some output deleted ...]

```

1856 **WinBUGS** did well in choosing an MCMC algorithm for this model – we have
1857 $\hat{R} = 1$ for each parameter, and an effective sample size of 3000, equal to the total
1858 number of posterior samples. We see that the posterior mean of N under this model
1859 is 49.94 and a 95% posterior interval is (48, 54). We revisit these data later in the
1860 context of more complex models.

1861 In order to obtain an estimate of density, D , we need an area to associate with
1862 the estimate of N , and commonly used procedures to conjure up such an area
1863 include buffering the trap array by the home range radius, often estimated by the
1864 mean maximum distance moved (MMDM)³, 1/2 MMDM (Dice, 1938) or directly
1865 from telemetry data (REF XXX NEED REF HERE XXXXX). Typically, the trap
1866 array is defined by the convex hull around the trap locations, and this is what we

³really MMDM? How can this be an estimate of the home range radius? Reference for this?

1867 applied a buffer to. We computed the buffer by using an estimate of the mean female
 1868 home range radius (2.19 km) estimated from telemetry studies (Bales et al., 2005)
 1869 instead of using an estimate based on our relatively more sparse recapture data⁴.
 1870 For the Fort Drum study, the convex hull has area 157.135 km^2 , and the buffered
 1871 convex hull has area 277.011 km^2 . To create this we used functions contained in the
 1872 **R** package **rgeos** and created a utility function **bcharea** which is in our **R** package
 1873 **scrbook**. The commands are as follows:

```
1874 library("rgeos")
1875
1876 bcharea<-function(buff,traplocs){
1877   p1<-Polygon(rbind(traplocs,traplocs[1,]))
1878   p2<-Polygons(list(p1=p1),ID=1)
1879   p3<-SpatialPolygons(list(p2=p2))
1880   p1ch<-gConvexHull(p3)
1881   bp1<-gBuffer(p1ch, width=buff)
1882   plot(bp1, col='gray')
1883   plot(p1ch, border='black', lwd=2, add=TRUE)
1884   gArea(bp1)
1885 }
1886
1887 bcharea(2.19,traplocs=trapmat)
```

1888 The resulting buffered convex hull is shown in Fig. 3.2.

1889 To conjure up a density estimate under model M_0 , we compute the appropriate
 1890 posterior summary of N and the prescribed area (277.011 km^2):

```
1891 > summary(fit0$sims.list$N/277.011)
1892   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
1893 0.1697 0.1733 0.1805 0.1803 0.1841 0.2130
1894
1895 > quantile(fit0$sims.list$N/277.011,c(0.025,0.975))
1896      2.5%      97.5%
1897 0.1696684 0.1949381
```

1898 which yields a density estimate of about 0.18 ind/ km^2 , and a 95% Bayesian confi-
 1899 dence interval of (0.170, 0.195).

1900 The obvious limitation of this estimate and, indeed, of the whole process, is that
 1901 our choice of “area” is completely subjective - which area should we use? MMDM?
 1902 One-half MMDM? Estimated from telemetry data? And, furthermore, how certain
 1903 are we of this area? Can we quantify our uncertainty about this quantity? More
 1904 important, what exactly is the meaning of this area and, in this context, how do
 1905 we gauge bias and/or variance of “estimators” of it? (i.e., what is it estimating?).

⁴BETH: Why?

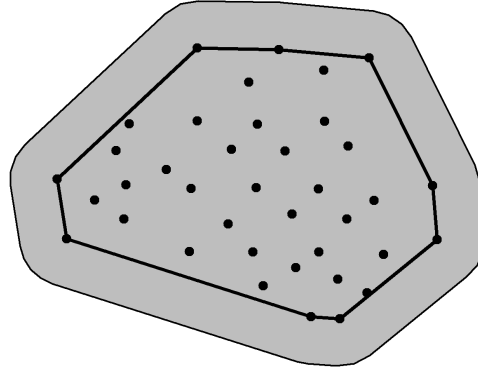


Figure 3.2. buffered convex hull of the bear hair snare array

3.3 TEMPORALLY VARYING AND BEHAVIORAL EFFECTS

The purpose of this chapter is mainly to emphasize the central importance of the binomial model in capture-recapture and so we have considered models for individual encounter frequencies - the number of times individuals are captured out of K samples. Sometimes it is not acceptable to aggregate the encounter data for each individual - such as when encounter probability varies over time among samples. A type of time-varying response that seems relevant in most capture-recapture studies is “effort” such as amount of search time, number of observers, or trap effort or when p depends on date (Kéry et al., 2010; Gardner et al., 2010). A common situation is that in which there exists a “behavioral response” to trapping (even if the animal is not physically trapped).

Behavioral response is an important concept in carnivore studies because individuals might learn to come to baited traps or avoid traps due to trauma related to being encountered. There are a number of ways to parameterize a behavioral response to encounter. The distinction between persistent and ephemeral was made by Yang and Chao (2005) who considered a general behavioral response model of the form:

$$\text{logit}(p_{ik}) = \alpha_0 + \alpha_1 * y_{i,k-1} + \alpha_2 x_{ik}$$

where x_{ik} is a covariate indicator variable of previous capture (i.e., $x_{ik} = 1$ if

captured in any previous period). Therefore, encounter probability changes depending on whether an individual was captured in the immediate previous period (ephemeral behavioral response) or in any previous period (persistent behavioral response). The former probably models a behavioral response due to individuals moving around their territory relatively slowly over time and the latter probably accommodates trap happiness due to baiting or shyness due to trauma. In spatial capture-recapture models it makes sense to consider a local behavioral response that is trap-specific (Royle et al., 2011c) - that is, the encounter probability is modified for individual traps depending on previous capture in specific traps.

Models with temporal effects are easy to describe in the **BUGS** language and analyze and we provide a number of examples in chapt. 8. XXXXX ?? XXXXX

3.4 MODELS WITH INDIVIDUAL HETEROGENEITY

Here we consider models with individual-specific encounter probability parameters, say p_i , which we model according to some probability distribution, $g(\theta)$. We denote this basic model assumption as $p_i \sim g(\theta)$. This type of model is similar in concept to extending a GLM to a GLMM but in the capture-recapture context N is unknown. The basic class of models is often referred to as “Model M_h ” but really this is a broad class of models, each being distinguished by the specific distribution assumed for p_i . There are many different varieties of Model M_h including parametric and various putatively non-parametric approaches (Burnham and Overton, 1978; Norris III and Pollock, 1996; Pledger, 2000). One important practical matter is that estimates of N can be extremely sensitive to the choice of heterogeneity model (Fienberg et al., 1999; Dorazio and Royle, 2003; Link, 2003). Indeed, Link (2003) showed that in some cases it’s possible to find models that yield precisely the same expected data, yet produce wildly different estimates of N . In that sense, N for most practical purposes is not identifiable across classes of mixture models, and this should be understood before fitting any such model. One solution to this problem is to seek to model explicit factors that contribute to heterogeneity, e.g., using individual covariate models (See 3.5 below). Indeed, spatial capture-recapture models seek to do just that, by modeling heterogeneity due to the spatial organization of individuals in relation to traps or other encounter mechanism. For additional background and applications of Model M_h see Royle and Dorazio (2008, chapt. 6) and Kery and Schaub (2011, chapt. 6).

Model M_h has important historical relevance to spatial capture-recapture situations (Karanth, 1995) because investigators recognized that the juxtaposition of individuals with the array of trap locations should yield heterogeneity in encounter probability, and thus it became common to use some version of Model M_h in spatial trapping arrays to estimate N . While this doesn’t resolve the problem of not knowing the area relevant to N , it does yield an estimator that accommodates the heterogeneity in p induced by the spatial aspect of capture-recapture studies.

To see how this juxtaposition induces heterogeneity, we have to understand

the relevance of movement in capture-recapture models. Imagine a quadrat that can be uniformly searched by a crew of biologists for some species of reptile (see Royle and Young (2008)). Figure 3.3 shows a sample quadrat searched repeatedly over a period of time. Further, suppose that species exhibits some sense of spatial fidelity in the form of a home range or territory, and individuals move about their home range (home range centroids are given by the blue dots) in some kind of random fashion. It is natural to think about it in terms of a movement process and sometimes that movement process can be modeled explicitly using hierarchical models (Royle and Young, 2008; Royle et al., 2011b). Heuristically, we imagine that each individual in the vicinity of the study area is liable to experience variable exposure to encounter due to the overlap of its home range with the sampled area - essentially the long-run proportion of times the individual is within the sample plot boundaries, say ϕ . We might model the exposure of an individual to capture by supposing that $z_i = 1$ if individual i is available to be captured (i.e., within the survey plot) during any sample, and 0 otherwise. Then, $\Pr(z_i = 1) = \phi$. In the context of spatial studies, it is natural that ϕ should depend on *where* an individual lives, i.e., it should be individual-specific ϕ_i (Chandler et al., 2011). This system describes, precisely, that of “random temporary emigration” (Kendall, 1997) where ϕ_i is the individual-specific probability of being “available” for capture.

Conceptually, SCR models aim to deal with this problem of variable exposure to sampling due to movement in the proximity of the trapping array explicitly and formally with auxiliary spatial information. If individuals are detected with probability p_0 , *conditional* on $z_i = 1$, then the marginal probability of detection of individual i is

$$p_i = p_0 \phi_i$$

so we see clearly that individual heterogeneity in encounter probability is induced as a result of the juxtaposition of individuals (i.e., their home ranges) with the sample apparatus and the movement of individuals about their home range.

We will work with a specific type of Model M_h here, that in which we extend the basic binomial observation model of Model M_0 so that

$$\text{logit}(p_i) = \mu + \eta_i$$

where

$$\eta_i \sim \text{Normal}(0, \sigma_p^2)$$

We could as well write

$$\text{logit}(p_i) \sim \text{Normal}(\mu, \sigma_p^2)$$

This “logit-normal mixture” was analyzed by Coull and Agresti (1999) and elsewhere. It is a natural extension of the basic model with constant p , as a mixed GLMM, and similar models occur throughout statistics. It is also natural to consider a beta prior distribution for p_i (Dorazio and Royle, 2003) and so-called “finite-mixture” models are also popular (Norris III and Pollock, 1996; Pledger, 2000).

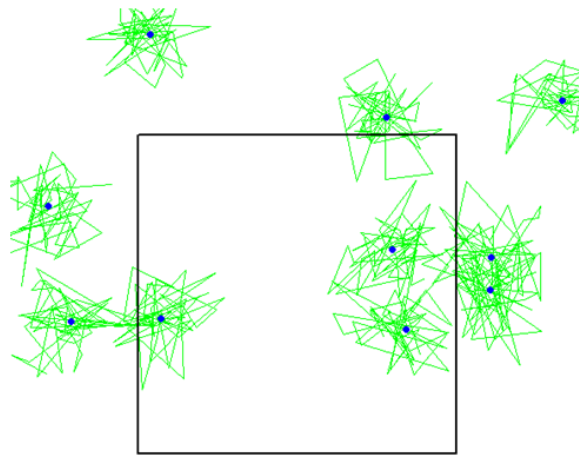


Figure 3.3. A quadrat searched for lizards and the locations of each lizard over some period of time.

3.4.1 Analysis of Model Mh

If N is known, it is worth taking note of the essential simplicity of Model M_h as a binomial GLMM. This is a type of model that is widely applied in just about every scientific discipline and using standard methods of inference based either on integrated likelihood (Laird and Ware, 1982; Berger et al., 1999) which we discuss in chapt. 6 or standard Bayesian methods. However, because N is not known, inference is somewhat more challenging. We address that here using Bayesian analysis based on data augmentation (DA). Although we use data augmentation in the context of Bayesian methods here, we note that heterogeneity models formulated under DA are easily analyzed by conventional likelihood methods as zero-inflated binomial mixtures (Royle, 2006) and more traditional analysis of model M_h based on integrated likelihood, without using data augmentation, has been considered by Coull and Agresti (1999), Dorazio and Royle (2003), and others.

As with model M_0 , we have the Bernoulli model for the zero-inflation variables: $z_i \sim \text{Bern}(\psi)$ and the model of the observations expressed conditional on the latent variables z_i . For $z_i = 1$, we have a binomial model with individual-specific p_i :

$$y_i | z_i = 1 \sim \text{Bin}(K, p_i)$$

and otherwise $y_i | z_i = 0 \sim \delta(0)$. Further, we prescribe a distribution for p_i . Here we assume

$$\text{logit}(p_i) \sim \text{Normal}(\mu, \sigma^2)$$

The basic **BUGS** description for this model, assuming a $\text{Unif}(0, 1)$ prior for $p_0 = \text{logit}^{-1}(\mu)$, is given as follows:

```

model{
  p0 ~ dunif(0,1)          # prior distributions
  mup<- log(p0/(1-p0))
  taup~dgamma(.1,.1)
  psi~dunif(0,1)

  for(i in 1:(nind+nz)){
    z[i]~dbern(psi)        # zero inflation variables
    lp[i] ~ dnorm(mup,taup) # individual effect
    logit(p[i])<-lp[i]
    mu[i]<-z[i]*p[i]
    y[i]~dbin(mu[i],J)     # observation model
  }

  N<-sum(z[1:(nind+nz)])  # N is a derived parameter
}
```

3.4.2 Analysis of the Fort Drum data

The logit-normal heterogeneity model was fitted to the bear data from the Fort Drum study, and we used data augmentation to produce a data set of $M = 500$ individuals. We ran the model using **JAGS** with the instructions given as follows⁵.

```

2040 [... get data as before ....]
2041
2042 set.seed(2013)
2043
2044 cat("
2045 model{
2046   p0 ~ dunif(0,1)          # prior distributions
2047   mup<- log(p0/(1-p0))
2048   sigmap ~ dunif(0,10)
2049   taup<- 1/(sigmap*sigmap)
2050   psi~dunif(0,1)
2051
2052   for(i in 1:(nind+nz)){
2053     z[i]~dbern(psi)        # zero inflation variables
2054     lp[i] ~ dnorm(mup,taup) # individual effect
2055     logit(p[i])<-lp[i]
2056     mu[i]<-z[i]*p[i]
2057     y[i]~dbin(mu[i],K)    # observation model
2058   }
2059
2060   N<-sum(z[1:(nind+nz)])
2061 }
2062 ",file="modelMh.txt")
2063
2064 data1<-list(y=ytot, nz=nz, nind=nind,K=K)
2065 params1= c('p0','sigmap','psi','N')
2066 inits = function() {list(z=as.numeric(ytot>=1), psi=.6, p0=runif(1),
2067   sigmap=runif(1,.7,1.2),lp=rnorm(M,-2)) }
2068
2069 library("rjags")
2070 jm<- jags.model("modelMh.txt", data=data1, inits=inits, n.chains=4,
2071   n.adapt=1000)
2072 jout<- coda.samples(jm, params1, n.iter=200000, thin=1)
2073
2074   This produces the posterior distribution for  $N$  shown in Fig. 3.4. Posterior
2075   summaries of parameters are given as follows:
2076
2077 > summary(jout)
2078
2079 Iterations = 2001:202000

```

⁵For WinBUGS, should provide starts for lp and sigma or sometimes WinBUGS breaks

```

2078 Thinning interval = 1
2079 Number of chains = 4
2080 Sample size per chain = 2e+05
2081
2082 1. Empirical mean and standard deviation for each variable,
2083    plus standard error of the mean:
2084
2085           Mean          SD Naive SE Time-series SE
2086 N      117.7740 56.31633 6.296e-02      1.960115
2087 p0      0.0728 0.05522 6.174e-05      0.001655
2088 psi     0.2366 0.11362 1.270e-04      0.003909
2089 sigmap  2.0795 0.53096 5.936e-04      0.016789
2090
2091 2. Quantiles for each variable:
2092
2093           2.5%      25%      50%      75%      97.5%
2094 N      62.000000 82.00000 102.00000 134.0000 277.0000
2095 p0      0.003143 0.02842 0.06077 0.1066 0.2036
2096 psi     0.117269 0.16377 0.20522 0.2712 0.5560
2097 sigmap  1.211900 1.69434 2.02113 2.4028 3.2694

```

We used $M = 500$ for this analysis and we note that while the posterior mass of N is concentrated away from this upper bound (Fig. 3.4), the posterior has an extremely long right tail, with some posterior values at the upper bound $N = 500$. Maybe or maybe not sufficient data augmentation.⁶ The model runs effectively in **WinBUGS** but sometimes with apparently inefficient mixing for reasons that may be related to bad starting values. In some cases this was resolved if we supplied starting values for the $\text{logit}(p_i)$ parameters and τ .

Because of the skewed posterior we see that the posterior mean ($N = 117$) is considerably higher than the posterior mode ($N = 102$). Moreover, posterior summaries are estimated with a relatively high error (“Time-series SE” of around 2.0)⁷. Further, it may be surprising that the posterior mode does not compare well with the MLE. To compute the posterior mode we could easily find the posterior value of N with the highest mass because N is discrete. But we want to smooth out some of the Monte Carlo error a bit so we used a smoothing spline to the posterior frequencies of N as follows:

```

2113 > tt<-table(jout[[1]][, "N"])[1:80]
2114 > xg<-as.numeric(names(tt))
2115 > plot(xg,tt)
2116 > sp<- smooth.spline(xg,tt,df=9)
2117 > sp$x[sp$y==max(sp$y)]
2118 [1] 80

```

⁶ to do: insert final results. longer run. more data augmentation. compare with winbugs.

⁷ need to define this somewhere

The `df` argument controls the degree of smoothing and we find in this case that the modal value (i.e., 80) is not too sensitive to the smoothing parameter but this should be checked in any specific instance⁸.

To compare with the MLE, we used the **R** code contained in Panel 6.1 of Royle and Dorazio (2008). The MLE of $\log(n_0)$, the logarithm of the number of uncaptured individuals, is $\log(n_0) = 3.86$ and therefore the MLE is $\hat{N} = \exp(3.86) + 47 = 94.47$ which is not at all consistent with the apparent mode in Fig. 3.4.⁹

Comments: First of all the posterior for this model and data set is very sensitive to prior distributions. While MLEs are invariant to transformation of the parameters, the posterior distribution definitely is *not* invariant. In the present case, the use of a $\text{Unif}(0, 1)$ prior for $p_0 = \text{expit}(\mu)$ is somewhat informative – in particular, it is not at all “flat” on the scale of μ – and this affects the posterior. We generally always recommend use of a $\text{Unif}(0, 1)$ prior for $\text{expit}(\mu)$ in such models. That said, we were surprised at this result, and we experimented with other prior configurations including putting a flat prior on μ directly. This specific prior suggests the possibility that the posterior distribution may be improper for that prior specification. This kind of small sample instability has been widely noted in Model M_h (Fienberg et al., 1999; Dorazio and Royle, 2003) and is not unrelated to sensitivity to model which has also been identified as an important issue in model M_h (Dorazio and Royle, 2003; Link, 2003). Conclusion: The mode is well-defined but the data set is sparse and hence inferences are poor and sensitive to model choices. Get over it.

3.4.3 Building your own MCMC algorithm

For fun, we construct our own MCMC algorithm using a Metropolized Gibbs sampler for Model M_h . In chapter 7 we devise MCMC algorithms for spatial capture-recapture models and the basic conceptual and technical considerations are entirely analogous to Model M_h .

To begin, we first collect all of our model components which are as follows: $[y_i|p_i, z_i]$, $[p_i|\mu_p, \sigma_p]$, and $[z_i|\psi]$ for each $i = 1, 2, \dots, M$ and then prior distributions $[\mu_p]$, $[\sigma_p]$ and $[\psi]$. The joint posterior distribution of all unknown quantities in the model is proportional to the joint distribution of all elements y_i, p_i, z_i and also the prior distributions of the prior parameters:

$$\left\{ \prod_{i=1}^M [y_i|p_i, z_i][p_i|\mu_p, \sigma_p][z_i|\psi] \right\} [\mu_p, \sigma_p, \psi]$$

For prior distributions, we assume that μ_p, σ_p, ψ are mutually independent and for μ_p and σ_p we use improper uniform priors, and $\psi \sim \text{Unif}(0, 1)$. Note that the

⁸we need to give examples of using `density()` to obtain modes

⁹We note that the result is inconsistent with Gardner et al. (2009) who reported an MLE of 104.1 ($\text{density} = 0.437 \text{ inds/km}^2$) although we do not know the reason for this at the present time.

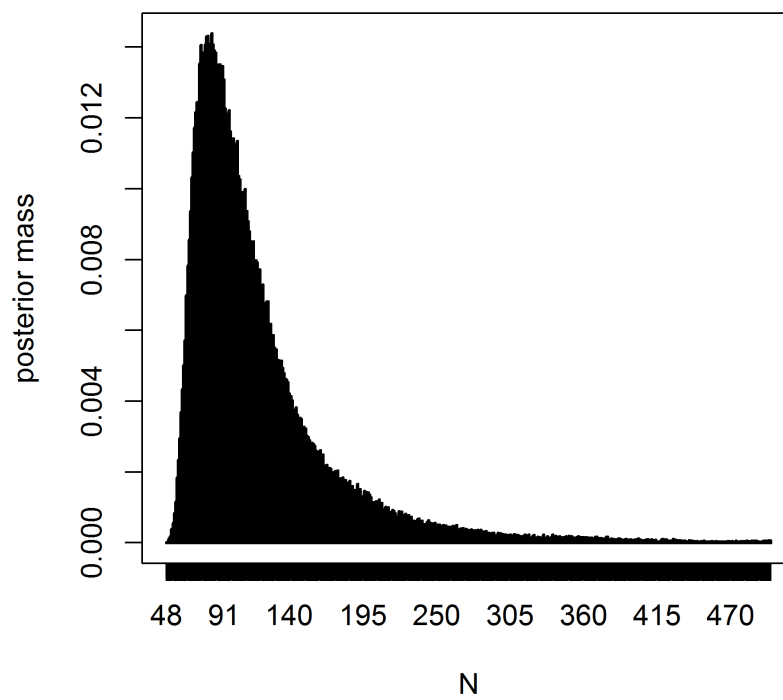


Figure 3.4. Posterior of N for Fort Drum bear study data under the logit-normal version of model M_h . From WinBUGS output. 200k samples.

likelihood contribution for each individual, when conditioned on p_i and z_i , does not depend on ψ , μ_p , or σ_p . As such, the full-conditionals for the structural parameters ψ only depends on the collection of data augmentation variables z_i , and that for μ_p and σ_p will only depends on the collection of latent variables p_i ; $i = 1, 2, \dots, M$. The full conditionals for all the unknowns are as follows:

(1) For p_i :

$$[p_i | y_i, \mu_p, \sigma_p, z_i = 1] \propto [y_i | p_i][p_i | \mu_p, \sigma_p^2] \text{ if } z_i = 1$$

$$[p_i | \mu_p, \sigma_p] \text{ if } z_i = 0$$

(2) for z_i :

$$z_i | \cdot \propto [y_i | z_i * p_i] \text{Bern}(z_i | \psi)$$

(3) For μ_p :

$$[\mu_p | \cdot] \sim \prod_i [p_i | \cdot] * \text{const}$$

(4) For σ_p :

$$[\sigma_p | \cdot] \sim \prod_i [p_i | \cdot] * \text{const}$$

(5) For ψ :

$$\psi | \cdot \sim \text{Beta}(1 + \sum z_i, 1 + M - \sum z_i)$$

We've identified each of the full conditional distributions in sufficient detail to apply the Metropolis-Hastings algorithm. With the exception of ψ which has a convenient analytic solution – it is a beta distribution which we can easily sample directly. In truth, we could also sample μ_p and σ_p^2 directly with certain choices of prior distributions. For example, if $\mu_p \sim \text{Normal}(0, 1000)$ then the full conditional for μ_p is also normal, etc.. We implement an MCMC algorithm for this model in the following block of **R** code. The basic structure is: initialize the parameters and create any required output or intermediate data holders, and then begin the main MCMC loop which, in this case, generates 100000 samples.¹⁰

```
## obtain the bear data by executing the previous data grabbing
## function
temp<-getdata()
M<-temp$M
K<-temp$K
ytot<-temp$ytot

###
### MCMC algorithm for Model Mh
```

¹⁰This data grabbing function is not implemented yet

```

2182
2183 out<-matrix(NA,nrow=100000,ncol=4)
2184 dimnames(out)<-list(NULL,c("mu","sigma","psi","N"))
2185 lp<- rnorm(M,-1,1)
2186 p<-expit(lp)
2187 mu<- -1
2188 p0<-exp(mu)/(1+exp(mu))
2189 sigma<- 1
2190 psi<- .5
2191 z<-rbinom(M,1,psi)
2192 z[ytot>0]<-1
2193
2194 for(i in 1:100000){
2195
2196   ### update the logit(p) parameters
2197   lpc<- rnorm(M,lp,1) # 0.5 is a tuning parameter
2198   pc<-expit(lpc)
2199   lik.curr<-log(dbinom(ytot,K,z*p)*dnorm(lp,mu,sigma))
2200   lik.cand<-log(dbinom(ytot,K,z*pc)*dnorm(lpc,mu,sigma))
2201   kp<- runif(M) < exp(lik.cand-lik.curr)
2202   p[kp]<-pc[kp]
2203   lp[kp]<-lpc[kp]
2204
2205   p0c<- rnorm(1,p0,.05)
2206   if(p0c>0 & p0c<1){
2207     muc<-log(p0c/(1-p0c))
2208     lik.curr<-sum(dnorm(lp,mu,sigma,log=TRUE))
2209     lik.cand<-sum(dnorm(lp,muc,sigma,log=TRUE))
2210     if(runif(1)<exp(lik.cand-lik.curr)) {
2211       mu<-muc
2212       p0<-p0c
2213     }
2214   }
2215
2216   sigmac<-rnorm(1,sigma,.5)
2217   if(sigmac>0){
2218     lik.curr<-sum(dnorm(lp,mu,sigma,log=TRUE))
2219     lik.cand<-sum(dnorm(lp,mu,sigmac,log=TRUE))
2220     if(runif(1)<exp(lik.cand-lik.curr))
2221       sigma<-sigmac
2222   }
2223
2224   ### update the z[i] variables
2225   zc<- ifelse(z==1,0,1) # candidate is 0 if current = 1, etc..
2226   lik.curr<- dbinom(ytot,K,z*p)*dbinom(z,1,psi)
2227   lik.cand<- dbinom(ytot,K,zc*p)*dbinom(zc,1,psi)

```

```

2228 kp<- runif(M) < (lik.cand/lik.curr)
2229 z[kp]<- zc[kp]
2230
2231 psi<-rbeta(1, sum(z) + 1, M-sum(z) + 1)
2232
2233 out[i,]<- c(mu,sigma,psi,sum(z))
2234 }

```

Remarks: (1) for parameters with bounded support, i.e., σ_p and p_0 , we are using a random walk candidate generator but rejecting draws outside of the parameter space. (2) We mostly use Metropolis-Hastings except for the data augmentation parameter ψ which we sample directly from its full-conditional distribution which is a beta distribution. (3) Even the latent data augmentation variables z_i are updated using Metropolis-Hastings although they too can be updated directly from their full-conditional.

2242 3.4.4 Exercises related to model Mh

- 2243 (1) Enclose the MCMC algorithm in an R function and provide arguments for some
- 2244 of the parameters of the function that a user might wish to modify.
- 2245 (2) Execute the function and compare the results to those generated from Win-
- 2246 BUGS in the previous section
- 2247 (3) Note that the prior distribution for the “mean” parameter is given on $p_0 =$
- 2248 $\exp(\mu)/(1 + \exp(\mu))$. Reformulate the algorithm with a flat prior on μ and see
- 2249 what happens. Contemplate this.
- 2250 (4) Using Bayes rule, figure out the full conditional for z_i so that you don’t have
- 2251 to use MH for that one. It might be more efficient. Is it?
- 2252 (5) Modify the MCMC algorithm so that the prior for μ_p is an improper flat prior.
- 2253 i.e., $[\mu_p] \propto 1$. Describe the posterior distribution of N .

2254 3.5 INDIVIDUAL COVARIATE MODELS: TOWARD SPATIAL CAPTURE-RECAPTURE

2254 A standard situation in capture-recapture models is when an individual covariate
2255 is measured, and this covariate is thought to influence encounter probability. As
2256 with other closed population models, we begin with the basic binomial observation
2257 model:

$$y_i \sim \text{Bin}(K, p_i)$$

2258 and we assume also a model for encounter probability according to:

$$\text{logit}(p_i) = \alpha + \beta x_i \quad (3.5.1)$$

2259 Classical examples of covariates influencing detection probability are type of animal
2260 (juvenile/adult or male/female), a continuous covariate such as body mass (Royle

and Dorazio, 2008, ch. 6), or a discrete covariate such as group or cluster size. For example, in models of aerial survey data, it is natural to model detection probabilities as a function of the observation-level individual covariate, “group size” (Royle, 2008, 2009; ?).

Such “individual covariate models” are similar in structure to Model M_h , except that the individual effects are *observed* for the n individuals that appear in the sample. These models are important here because spatial capture-recapture models are precisely a form of individual covariate model, an idea that we will develop here and elsewhere. Specifically, they are such models, but where the individual covariate is a partially observed latent variable similar.. That is, unlike Model M_h , we do have some direct information about the latent variable, which comes from the spatial locations/distribution of individual recaptures.

Traditionally, estimation of N in individual covariate models is achieved using methods based on ideas of unequal probability sampling (i.e., Horwitz-Thompson estimation; see Huggins (1989) and Alho (1990)). An estimator of N is

$$\hat{N} = \sum_i^n \frac{1}{\tilde{p}_i}$$

where \tilde{p}_i is the probability that individual i appeared in the sample. That is, $\tilde{p}_i = \Pr(y_i > 0)$ where, in closed population capture-recapture models,

$$\Pr(y_i > 0) = (1 - (1 - p_i)^K)$$

where p_i is a function of parameters α and β according to Eq. 3.5.1. In practice, parameters are estimated from the conditional-likelihood of the observed encounter histories which is, for observation y_i ,

$$\mathcal{L}(\alpha, \beta | y_i) = \frac{\text{Bin}(y_i | \alpha, \beta)}{\tilde{p}_i}.$$

Here we take a formal model-based approach to Bayesian analysis of such models using data augmentation (Royle, 2009). Classical likelihood analysis of the so-called “full likelihood” is covered by Borchers et al. (2002). For Bayesian analysis of individual covariate models, because the individual covariate is unobserved for the $N - n$ uncaptured individuals, we require a model to describe variation among individuals, essentially allowing the sample to be extrapolated to the population¹¹. For our present purposes, we consider a continuous covariate and we assume that it has a normal distribution:

$$x_i \sim \text{Normal}(\mu, \sigma^2)$$

¹¹weak argument

2289 Data augmentation can be applied directly to this class of models. In particular,
 2290 reformulation of the model under DA yields a basic zero-inflated binomial model of
 2291 the form:

$$\begin{aligned} z_i &\sim \text{Bern}(\psi) \\ y_i|z_i=1 &\sim \text{Bin}(K, p_i(x_i)) \\ y_i|z_i=0 &\sim \delta(0) \\ x_i &\sim \text{Normal}(\mu, \sigma^2) \end{aligned}$$

2292 Fully spatial capture-recapture models use this formulation with a latent covariate
 2293 that is directly related to the individual detection probability (see next Section).
 2294 As with the previous models, implementation is trivial in the **BUGS** language.
 2295 The **BUGS** specification is very similar to that for model M_h , but we require the
 2296 distribution of the covariate to be specified, along with priors for the parameters of
 2297 that distribution.

2298 3.5.1 Example: Location of capture as a covariate.

2299 If we had a regular grid of traps over some closed geographic system then we imagine
 2300 that the average location of capture would be a decent estimate (heuristically) of
 2301 an individual's home range center. Intuitively some measure of typical distance
 2302 from home range center to traps for an individual should be a decent covariate to
 2303 explain heterogeneity in encounter probability, i.e., individuals with more exposure
 2304 to traps should have higher encounter probabilities and vice versa. A version of
 2305 this idea was put forth by Boulanger and McLellan (2001) (see also Ivan (2012)),
 2306 but using the Huggins-Alho estimator and with covariate "distance to edge" of the
 2307 trapping array. A limitation of this approach is that it does not provide a solution
 2308 to the problem that the trap area is fundamentally ill-defined, nor does it readily
 2309 accommodate the inherent and heterogeneous variation in this measured covariate.

2310 Here, we provide an example of this type of heuristically motivated approach
 2311 using the fully model-based individual covariate model described above analyzed
 2312 by data augmentation. We take a slightly different approach than that adopted
 2313 by Boulanger and McLellan (2001). By analyzing the full likelihood and placing a
 2314 prior distribution on the individual covariate, we resolve the problem of having an
 2315 ill-defined area over which the population size is distributed. After you read later
 2316 chapters of this book, it will be apparent that SCR models represent a formalization
 2317 of this heuristic procedure.

2318 For our purposes here, we define $x_i = ||\mathbf{s}_i - \mathbf{x}_0||$ where \mathbf{s}_i is the average encounter
 2319 location of individual i and \mathbf{x}_0 is the centroid of the trap array. Conceptually,
 2320 individuals in the middle of the array should have higher probability of encounter
 2321 and, as x_i increases, p_i should therefore decrease. We note that we have defined
 2322 s_i in terms of a sample quantity - the observed mean - which is ad hoc but maybe
 2323 satisfactory under the circumstances. That said, for an expansive, dense trapping

grid then we might expect the sample mean encounter location to be a good estimate of home range center but, clearly this is biased for individuals that live around the edge (or off) the trapping array. Regardless, it should be good enough for our present purposes of demonstrating this heuristically appealing application of an individual covariate model. A key point is that s_i is missing for each individual that is not encountered and thus so is x_i . Thus, it is a latent variable, or random effect, and we need therefore to specify a probability distribution for it. As a measurement of distance we know it must be positive-valued. Thinking about this like a distance sampling problem lets first try to make x_i uniform from 0 to some large number, say D_{max} , beyond which it would be difficult to imagine an individual being captured. For example, D_{max} should be at a home range diameter past the furthest trap from the center. As such, we use this distribution for the individual covariate “distance from center of the trap array”

$$x_i \sim \text{uniform}(0, D_{max})$$

where D_{max} is a specified constant. In practice, people have used distance from edge of the trap array but that is less easy to define and compute.

ANDY STOPPED RIGHT HERE

Note: I’m a little cautious about stumbling through an analysis that is blatantly wrong – but I’m hoping it will evolve into a learning experience for the reader.

Fort Drum Bear Study

We have to do a little bit of data processing to fit this individual covariate model to the Fort Drum data. We compute the individual covariate \mathbf{x}_i using the **R** script provided in *scrbook*. We picked $D_{max} = 11.5 \text{ km}^2$ which is about the distance from the array center to the furthest trap. Once we specific D_{max} then the implication is that the population size parameter applies to the area within 11.5 units of the trap centroid¹². The **BUGS** model specification and **R** commands to package the data and fit the model are as follows:

```
cat("
model{
  p0 ~ dunif(0,1)      # prior distributions
  mup<- log(p0/(1-p0))
  psi~dunif(0,1)
  beta~dnorm(0,.01)

  for(i in 1:(nind+nz)){
    xcent[i]~dunif(0,maxD)
    z[i]~dbern(psi)      # DA variables
    lp[i] <- mup + beta*xcent[i] # individual effect
```

¹²To be convincing this might need a little bit of hand-holding

```

2362   logit(p[i])<-lp[i]
2363   mu[i]<-z[i]*p[i]
2364   y[i]~dbin(mu[i],K) # observation model
2365 }
2366 N<-sum(z[1:(nind+nz)])
2367 }
2368 ",file="modelMcov.txt")
2369 data2<-list(y=ytot,nz=nz,nind=nind,K=K,xcent=xcent,maxD=11.5)
2370 params2<-list('p0','psi','N','beta')
2371 inits = function() {list(z=z, psi=psi, p0=runif(1),beta=rnorm(1) ) }
2372 fit2 = bugs(data2, inits, params2, model.file="modelMcov.txt",working.directory=getwd(),
2373           debug=T, n.chains=3, n.iter=4000, n.burnin=1000, n.thin=4)

```

Posterior summaries are given in Table ?? XYZ, and the posterior distribution of N is given in Fig. 3.5. It might be perplexing that the estimated N is much lower than obtained by model M_h but there is a good explanation for this, discussed subsequently. That issue notwithstanding, it is worth pondering how this model could be an improvement (conceptually or technically) over some other model/estimator including M_0 and M_h considered previously. Well, for one, we have accounted formally for heterogeneity due to spatial location of individuals relative to exposure to the trap array, characterized by the centroid of the array. Moreover, we have done so using a model that is based on an explicit mechanism, as opposed to a phenomenological one such as Model M_h . Moreover, importantly, using our new model, *the estimated N applies to an explicit area which is defined by our prescribed value of $\max D$* . That is, this area is a fixed component of the model and the parameter N therefore has explicit spatial context, as the number of individuals with home range centers less than $\max D$ from the centroid of the trap array. As such, the implied “effective trap area”¹³ for any $\max D$ is that of a circle with radius $\max D$.

```

2390 %% Not sure whether this should be a table or verbatim print-out
2391 \begin{table}
2392 \tabular{cccccccc}
2393 Node statistics
2394 node mean sd MC error 2.5% median 97.5% start sample
2395 N 58.89 5.483 0.2199 50.0 58.0 71.0 251 2250
2396 beta -0.246 0.06087 0.003892 -0.3592 -0.2457 -0.126 251 2250
2397 deviance 459.4 13.29 0.4496 435.7 458.4 487.8 251 2250
2398 p0 0.5409 0.06817 0.004052 0.4072 0.544 0.6678 251 2250
2399 psi 0.1706 0.02572 7.759E-4 0.1247 0.1692 0.2242 251 2250
2400 \end{tabular}
2401 \caption{..... xyz .....}
2402 \end{table}

```

¹³This is a bad use of this term. We have never defined ETA or ESA. What is it, exactly?

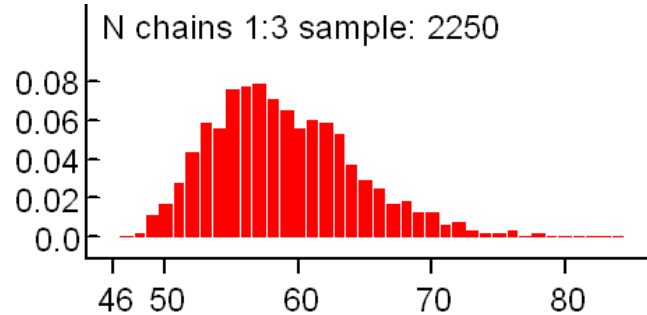


Figure 3.5. Needs a caption

2403 `\label{tab.maxD}`

2404 We'll remake this figure in R. For now, insert it as is.

2405 3.5.2 Extension of the Model

2406 One important issue in understanding the meaning of estimates produced under
 2407 the individual covariate model is that the uniform distribution on maxD implies
 2408 that density is *not constant* over space. In particular, this model implies that it
 2409 *decreases* as we move away from the centroid of the trap array. This is one reason
 2410 we have a lower estimate of density than that obtained previously and also why,
 2411 if we were to increase maxD , we would see density continue to decrease: $x[i] \sim$
 2412 $\text{Uniform}(0, \text{maxD})$ implies constant N in each distance band from the centroid but
 2413 obviously the *area* of each distance band is increasing. The reader can verify this
 2414 as a homework exercise.

2415 Obviously, the use of an individual covariate model is *not* restricted to use of
 2416 this specific distribution for the individual covariate. Clearly, it is a bad choice and,
 2417 therefore, we should think about whether we can choose a better distribution for
 2418 maxD - one that doesn't imply a decreasing density as distance from the centroid
 2419 increases. Conceptually, what we want to do is impose a prior on distance from
 2420 the centroid, x , such that density is proportional to the amount of area in each
 2421 successive distance band as you move farther away from the centroid. In fact, there
 2422 is theory that exists which tells us what the correct distribution of x is $2x/\text{maxD}^2$.
 2423 This can be derived by noting that $F(x) = \text{Pr}(X < x) = \pi x^2 / \pi \text{maxD}^2$.
 2424 Then, $f(x) = dF/dx = 2x / (\text{maxD}^2)$. This might be called a
 2425 triangular distribution, I think, which makes sense because the incremental area
 2426 in each additional distance band increases linearly with radius (i.e., distance from
 2427 centroid). It is sometimes comforting to verify things empirically:

```

2428 > u<-runif(10000,-1,1)
2429 > v<-runif(10000,-1,1)
2430 > d<- sqrt(u*u+v*v)
2431 > hist(d[d<1])
2432 > hist(d[d<1],100)
2433 > hist(d[d<1],100,probability=TRUE)
2434 > abline(0,2)

```

It would be useful if we could describe this distribution in *BUGS but there is not a built-in way to do this. One possibility is to use a discrete version of the pdf. We might also be able to use what is referred to in WinBUGS jargon as the “zeros trick” (see Advanced BUGS tricks) although we haven’t pursued this approach. Instead, we consider using a discrete version and break D_{\max} into L distance classes of width δ , with probabilities proportional to $2 * x$. In particular, if the cut-points are $xg[1] = 0, xg[2], \dots, xg[L + 1] = D_{\max}$ and the interval midpoints are $xm[i] = xg[i + 1] - \delta$. Then, the interval probabilities are $p[i] = 2 * xm[i] * \delta / (D_{\max} * D_{\max})$, which we can compute once and then send them to WinBUGS as data.

The R script is as follows. In the model description the variable x (observed home range center) has been rounded so that the discrete version of the $f(x)$ can be used as described previously. The new variable labeled `xround` is actually then the integer category label in units of δ from 0. Thus, to convert back to distance in the expression for $lp[i]$, `xround[i]` has to be multiplied by δ .

```

2449 delta<-.2
2450 xround<-xcent/%delta + 1
2451 Dgrid<- seq(delta,maxD,delta)
2452 xprobs<- delta*(2*Dgrid/(maxD*maxD))
2453 xprobs<-xprobs/sum(xprobs)
2454
2455 cat("
2456 model{
2457   p0 ~ dunif(0,1)          # prior distributions
2458   mup<- log(p0/(1-p0))
2459   psi~dunif(0,1)
2460   beta~dnorm(0,.01)
2461
2462   for(i in 1:(nind+nz)){
2463     xround[i]~dcat(xprobs[])
2464     z[i]~dbern(psi)          # zero inflation variables
2465     lp[i] <- mup + beta*xround[i]*delta # individual effect
2466     logit(p[i])<-lp[i]
2467     mu[i]<-z[i]*p[i]
2468     y[i]~dbin(mu[i],K)      # observation model
2469   }

```

```

2470
2471 N<-sum(z[1:(nind+nz)])
2472 }
2473 ",file="modelMcov.txt")

```

2474 To fit the model we do this - keeping in mind that the data objects required
 2475 below have been defined in previous analyses of this chapter:

```

2476 data2<-list(y=ytot,nz=nz,nind=nind,K=T,xround=xround,xprobs=xprobs,delta=delta)
2477 params2<-list('p0','psi','N','beta')
2478 inits = function() {list(z=z, psi=psi, p0=runif(1),beta=rnorm(1) ) }
2479 fit = bugs(data2, inits, params2, model.file="modelMcov.txt",working.directory=getwd(),
2480           debug=FALSE, n.chains=3, n.iter=11000, n.burnin=1000, n.thin=2)

```

2481 This is a useful model because it induces a clear definition of area in which
 2482 the population of N individuals reside. Under this model, that area is defined by
 2483 specification of D_{max} . We can apply the model for different values of D_{max} and
 2484 observe that the estimated N varies with D_{max} . Fortunately, we see empirically,
 2485 that while N seems highly sensitive to the prescribed value of D_{max} , density seems
 2486 to be invariant to D_{max} as long as it is chosen to be sufficiently large. We fit the
 2487 model for $D_{max} = 12$ (points in close proximity to the trap arra) to 20 for and the
 2488 results are given in Table ??.

```

2489 \begin{table}
2490 \caption{Table: Analysis of Fort Drum bear hair snare data using the individual covariate }
2491 \begin{tabular}{cccc}
2492     maxD  mn   SD
2493 [1,]   12 0.230 0.038
2494 [2,]   15 0.244 0.041
2495 [3,]   17 0.249 0.044
2496 [4,]   18 0.249 0.043
2497 [5,]   19 0.250 0.043
2498 [6,]   20 0.250 0.044
2499 \end{tabular}
2500 \end{table}

```

2501 We see that the posterior mean and SD of density (individuals per square km)
 2502 appear insensitive to choice of $maxD$ once we get a slight ways away from the
 2503 maximum observed value of about 11.5. The estimated density of 0.250 per km² is
 2504 actually quite a bit lower than we reported using model Mh (0.37, see section XYZ
 2505 above) for which sample area is not an explicit feature of the model. On the other
 2506 hand it is higher than that reported from Model M0 using the buffered area (0.195).
 2507 There is no basis really for comparing or contrasting these various estimates and
 2508 it would be a useful philosophical exercise for the reader to discuss this matter.

In particular, application of model M0 and Mh are distinctly *not* spatially explicit models – the area within which the population¹⁴ resides is not defined under either model. There is therefore no reason at all to think that the estimates produced under either model, using a buffered area, are justifiable based on any theory. In fact, we would get exactly the same estimate of N no matter what we declare the area to be. On the other hand, the individual covariate model explicitly describes a distribution for “distance from centroid” that is a reasonable and standard null model - it posits, in the absence of direct information, that individual home range centers are randomly distributed in space and that probability of detection depends on the distance between home range center and the centroid of the trap array. Under this definition of the system, we see that density is invariant to the choice of sample area which seems like a desirable feature. The individual covariate model is not ideal, however, because it does not make full use of the spatial information in the data set, i.e., the trap locations and the locations of each individual encounter.

3.5.3 Invariance of density to maxD

Under the model above, and also under models that we consider in later chapters, a general property of the estimators is that while N increases with the prescribed trap area (equivalent to maxD in this case), we expect that density estimators should be invariant to this area. In the model used above, we note that $Area(maxD) = \pi * maxD * maxD$ and $E[N(maxD)] = \lambda * A(maxD)$ and thus $E[Density(maxD)] = \lambda$ which is constant. This should be interpreted as the *prior* density. Absent data, then realizations under the model will have density λ regardless of what $maxD$ is prescribed to be. As we verified empirically above, the posterior density is also invariant if $maxD$ as long as the implied area (implied by maxD) is large enough so that the data no longer provide information about density (i.e., “far away”), then our estimator of density should become insensitive.

3.5.4 Toward Fully Spatial Capture-recapture Models

We developed this model for the average observed location and equated it to home range center s_i . Intuitively, taking the average encounter location as an estimate of home range center makes sense but more so when the trapping grid is dense and expansive relative to typical home range sizes. However, our approach also ignored the variable precision with which each $s[i]$ is estimated and also, as noted previously, estimates of $s[i]$ around the “edge” (however we define that) are biased because the observations are truncated (we can only observe locations within the trap array). In the next Chapter we provide a further extension of this individual covariate model that definitively resolves the ad hoc nature of the individual covariate approach we took here. In that model we build a model in which $s[i]$ are regarded as latent

¹⁴We need to look back at Chapter 1 and make sure we quit calling this “sample area” - it really isn’t that at all, but rather the area within which N resides.

variables and the observation locations (i.e., trap specific encounters) are linked to those latent variables with an explicit model. We note that the model fitted previously could be adapted easily to deal with s_i as a latent variable, simply by adding a prior distribution for s_i . The reader should contemplate how to do this in WinBUGS.

3.6 DISTANCE SAMPLING: A PRIMITIVE SPATIAL CAPTURE-RECAPTURE MODEL

Distance sampling is one of the most popular methods for estimating animal abundance. One of the great benefits of distance sampling is that it provides explicit estimates of *density*. The distance sampling model is a special case of a closed population model with a covariate. The covariate in this case, x_i , is the distance between an individual's location " u " and the observation location or transect. In fact, the model underlying distance sampling is precisely the same model as that which applies to the individual-covariate models, except that observations are made at only $K = 1$ sampling occasion. In a sense, distance sampling is a spatial capture-recapture model, but without the "recapture." This first and most basic spatial capture-recapture model has been used routinely for decades and, formally, it is a spatially-explicit model in the sense that it describes, explicitly, the spatial organization of individual locations (although this is not always stated explicitly) and, as a result, somewhat general models of how individuals are distributed in space can be specified (Royle et al., 2004; Johnson, 2010; Sillett, 2011). As before, the distance sampling model, under data augmentation, includes a set of M zero-inflation variables z_i and the binomial model expressed conditional on z (binomial for $z = 1$, and fixed zeros for $z = 0$). In distance sampling we pay for having only a single sample (i.e., $K = 1$) by requiring constraints on the model of detection probability. A standard model is

$$\log(p_i) = b * x_i^2$$

for $b < 0$, where x_i denotes the distance at which the i th individual is detected relative to some reference location where perfect detectability ($p = 1$) is assumed. This function corresponds to the "half-normal" detection function (i.e., with $b = 1/\sigma^2$). If $K > 1$ then the intercept alpha is identifiable and such models are usually called "capture-recapture distance sampling" (Borchers, missing) and others XYZ????).

As with previous examples, we require a distribution for the individual covariate x_i . The customary choice is

$$x_i \sim \text{Uniform}(0, B)$$

wherein $B > 0$ is a known constant, being the upper limit of data recording by the observer (i.e., the point count radius, or transect half-width). In practice, this is sometimes asserted to be infinity, but in such cases the distance data are usually

truncated. Specification of this distance sampling model in the BUGS language is shown in Panel 3.1. Royle and Dorazio (2008), p. xyz) provide a distance sampling example analyzed by DA using the famous Impala data.

```

b~dunif(0,10)
psi~dunif(0,1)

for(i in 1:(nind+nz)){
  z[i]~dbern(psi)      # DA Variables
  x[i]~dunif(0,B)      # B=strip width
  p[i]<-exp(logp[i])    # DETECTION MODEL
  logp[i]<- -((x[i]*x[i])*b)
  mu[i]<-z[i]*p[i]
  y[i]~dbern(mu[i])    # OBSERVATION MODEL
}
N<-sum(z[1:(nind+nz)])
D<- N/striparea # area of transects

```

Panel 3.1: Distance sampling model in WinBUGS, using a “half-normal” detection function.

As with the individual covariate model in the previous section, the distance sampling model can be equivalently specified by putting a prior distribution on individual *location* instead of distance between individual and observation point (or transect). Thus we can write the general distance sampling model as

$$\text{logit}(p[i]) = \alpha + \beta * ||u[i] - x_0||$$

Along with

$$\mathbf{u}_i \sim \text{Uniform}(\mathcal{S})$$

where x_0 is a fixed point (or line) and $u[i]$ is the individual’s location which is observable for n individuals. In practice it is easier to record distance instead of location. Basic math can be used to argue that if individuals have a uniform distribution in space, then the distribution of Euclidean distance is also uniform. In particular, if a transect of length L is used and x is distance to the transect then $F(x) = \text{Pr}(X \leq x) = L * x/L * B = x/B$ and $f(x) = dF/dx = (1/B)$. For measurements of radial distance, see the previous section.

In the context of our general characterization of SCR models (chapter 1.XYZ), we suggested that every SCR model can be described, conceptually, by a hierarchical model of the form:

$$[y|u][u|s][s].$$

Distance sampling ignores s , and treats u as observed data¹⁵. Thus, we are left

¹⁵Formally we could also say that $[u] = \int [y|s][s]ds$

2600 with

$$[y|u][u].$$

2601 In contrast, as we will see in the next chapters, basic SCR models (chapter 4) ignore
2602 u and condition on s , which is not observed:

$$[y|s][s]$$

2603 Since $[u]$ and $[s]$ are both assumed to be uniformly distributed, these are structurally
2604 equivalent models! The main differences have to do with interpretation of model
2605 components and whether or not the latent variables are observable (in distance
2606 sampling they are).

2607 So why bother with SCR models when distance sampling yields density esti-
2608 mates and accounts for spatial heterogeneity in detection? For one, imagine try to
2609 collect distance sampling data on tigers! Clearly, distance sampling requires that
2610 one can collect large quantities of distance data, which is not always possible. For
2611 tigers, it is much easier, efficient, and safer to employ camera traps or tracking
2612 plates and then apply SCR models. Furthermore, as we will see in Ch XYZ, SCR
2613 models can use distance data to estimate all the parameters of our enchilada, al-
2614 lowing us to study distribution, movement, and density. Thus, SCR models are
2615 much more flexible than distance sampling models, and can accommodate data
2616 from virtually all animal survey designs.

2617 3.6.1 Example: Muntjac deer survey from Nagarahole, India

2618 Here we fit distance sampling models to distance sampling data on the muntjac deer
2619 (*Muntiacus muntjak*) collected in the year 2004 from Nagarahole National Park in
2620 southern India (Kumar, missing)(Kumar et al. unpublished data). The muntjac
2621 is a solitary species and distance measurements were made on 57 groups that were
2622 largely singletons with XYZ pairs of individuals. Commands for reading in and
2623 organizing the data for WinBUGS, followed by writing the model to a text file.
2624 Note that the total sampled area of the transects is fed in as “striparea” which is
2625 708 (km of transect) multiplied by the strip width ($B=150 = 0.15$ km) multiplied
2626 by 2.

```
2627 library("R2WinBUGS")
2628 data<- read.csv("Muntjac.csv")
2629 nind<-nrow(data)
2630 y<-rep(1,nind)
2631 nz<-400
2632 y<-c(y,rep(0,nz))
2633 x<-data[,3]
2634 x<-c(x,rep(NA,nz))
2635 z<-y
```

```

2636 data<-list(y=y,x=x,nz=nz,nind=nind,B=150,striparea=708*.15*2)
2637
2638 cat("
2639 model{
2640   b~dunif(0,10)
2641   psi~dunif(0,1)
2642
2643   for(i in 1:(nind+nz)){
2644     z[i]~dbern(psi)    # DA Variables
2645     x[i]~dunif(0,B)    # B=strip width
2646     p[i]<-exp(logp[i])  # DETECTION MODEL
2647     logp[i]<- -((x[i]*x[i])*b)
2648     #logp[i]<- -b*log(x[i]+1)
2649     mu[i]<-z[i]*p[i]
2650     y[i]~dbern(mu[i])  # OBSERVATION MODEL
2651   }
2652   N<-sum(z[1:(nind+nz)])
2653   D<- N/striparea    # area of transects
2654 }
2655 ",file="dsamp.txt")

```

Next, we provide inits, indicate which parameters to monitor, and then pass those things to WinBUGS:

```

2658 params<-list('b','N','D','psi')
2659 inits = function() {list(z=z, psi=runif(1), b=runif(1,0,.02) )}
2660 fit = bugs(data, inits, params, model.file="dsamp.txt",
2661 working.directory=getwd(),debug=T, n.chains=3, n.iter=4000, n.burnin=1000, n.thin=2)

```

Posterior summaries are provided in the following table. Estimated density is pretty low, 1.1 individuals per sq. km.¹⁶

```

2664 node mean sd MC error 2.5% median 97.5% start sample
2665 D 1.096 0.1694 0.009122 0.8098 1.078 1.474 501 4500
2666 N 232.8 35.99 1.938 172.0 229.0 313.0 501 4500
2667 b 5.678E-4 1.05E-4 4.129E-6 3.867E-4 5.616E-4 7.949E-4 501 4500
2668 deviance 681.2 16.72 0.7536 650.8 680.6 716.6 501 4500
2669 psi 0.5099 0.08238 0.004442 0.3681 0.5033 0.6918 501 4500

```

3.7 SUMMARY AND OUTLOOK

Traditional closed population capture-recapture models are closely related to binomial generalized linear models. Indeed, the only real distinction is that in capture-

¹⁶ much lower than Samba's : Observers walked about 708 km from 39 transects in Nagarahole and the muntjac density is about 3 per sq km.. I need to get to the bottom of this.

recapture models, the population size parameter N (corresponding also to the size of a hypothetical “complete” data set) is unknown. This requires special consideration in the analysis of capture-recapture models. The classical approach to inference recognizes that the observations don’t have a standard binomial distribution but, rather, a truncated binomial (from which the so-called “conditional likelihood” derives) since we only have encounter frequency data on observed individuals. If instead we analyze the models using data augmentation, the observations can be modeled using a zero-inflated binomial distribution. In short, when we deal with the unknown- N problem using data augmentation then we are left with zero-inflated GLM and GLMMs instead of ordinary GLM or GLMMs. The analysis of such zero-inflated models is practically convenient, especially using the various Bayesian analysis packages that use the BUGS language.

Spatial capture-recapture models that we will consider in the rest of the chapters of this book are closely related to what have been called individual covariate models. Heuristically, spatial capture-recapture models arise by defining individual covariates based on observed locations of individuals – we can think of using some function of mean encounter location as an individual covariate. We did this in a novel way, by using distance to the centroid of the trapping array as a covariate. We analyzed the “full likelihood” using data augmentation, and placed a prior distribution on the individual covariate which was derived from an assumption that individual locations are, a priori, uniformly distributed in space. This assumption provides for invariance of the density estimator to the choice of population size area (induced by maximum distance from the centroid of the). The model addressed some important problems in the use of closed population models: it allows for heterogeneity in encounter probability due to the spatial context of the problem and it also provides a direct estimate of density because area is a feature of the model (via the prior on the individual covariate). The model is still not completely general because the model does not make use of the fully spatial encounter histories, which provide direct information about the locations and density of individuals. A specific individual covariate model that is in widespread use is classical “distance sampling.” The model underlying distance sampling is precisely a special kind of SCR model - but one without replicate samples. Understanding distance sampling and individual covariate models more broadly provides a solid basis for understanding and analyzing spatial capture-recapture models.

4

FULLY SPATIAL CAPTURE-RECAPTURE MODELS

In previous sections we discussed some classes of models that could be viewed as primitive spatial capture-recapture models. We looked at a basic distance sampling model and we also considered a classical individual covariate modeling approach in which we defined a covariate to be the distance from (estimated) home range center to the center of the trap array. These were spatial in the sense that they included some characterization of where individuals live but, on the other hand, only a primitive or no characterization of trap location. That said, very little distinguishes these two models from spatial capture-recapture models that we consider in this chapter which fully recognize the spatial attribution of both individual animals *and* the locations of encounter devices.

Fully spatial capture-recapture models must accommodate the spatial organization of individuals and the encounter devices because the encounter process occurs at the level of individual traps. Failure to consider the trap-specific collection of data is the key deficiency with classical ad-hoc approaches which aggregate encounter information to the resolution of the entire trap array. We have seen previously some problems that this induces - imbalance in trap-level effort over time is problematic, and not being able to deal with trap-specific behavioral responses. Here, we resolve that by developing what is basically an individual covariate model but operating at the level of traps. That is, we develop our first fully spatial capture-recapture model which turns out to be precisely the model considered in section 3.XXX but instead of defining the individual covariate to be distance to centroid of the array we define J individual covariates - the distance to *each* trap. And, instead of using estimates of individual locations \mathbf{s} , we consider a fully hierarchical model in which we regard \mathbf{s} as a latent variable and impose a prior distribution on

it. We can think of having J independent capture-recapture studies generating one data set for each trap, and applying the individual covariate model with random activity centers, and that is all the basic SCR model is.

In the following sections of this chapter we investigate the basic spatial capture-recapture model and address some important considerations related to its analysis in **WinBUGS**. We also demonstrate how to summarize posterior output for the purposes of producing density maps or spatial predictions of density.

4.1 SAMPLING DESIGN AND DATA STRUCTURE

In our development here, we will assume a standard sampling design in which an array of J traps is operated for K time periods (say, nights) producing encounters of n individuals. Because sampling occurs by traps and also over time, the most general data structure yields encounter histories for *each individual* that are temporally *and* spatially indexed. Thus a typical data set will include an encounter history *matrix* for each individual. For the most basic model, there are no time-varying covariates that influence encounter, there are no explicit individual-specific covariates, and there are no covariates that influence density we will develop models in this chapter for encounter data that are aggregated over the temporal replicates. For example, suppose we observe 6 individuals in sampling at 4 traps over 3 nights of sampling then a plausible data set is the 6×4 matrix of encounters, out of 3, of the form:

	trap1	trap2	trap3	trap4
[1,]	1	0	0	0
[2,]	0	2	0	0
[3,]	0	0	0	1
[4,]	0	1	0	0
[5,]	0	0	1	1
[6,]	1	0	1	0

We develop models in this chapter for devices such as “hair snares” or other DNA sampling methods (Kéry et al., 2010; Gardner et al., 2010) and related types of sampling problems so that we can suppose that “traps” may capture any number of individuals and an individual may be captured in any number of traps during each occasion but individuals can be encountered at most 1 time in a trap during any occasion. Thus, this is a “multi-catch” type of sampling (?, p. xyz). The statistical assumptions are that individual encounters within and among traps are independent. These basic (but admittedly at this point somewhat imprecise) assumptions define the basic spatial capture-recapture model, which we will refer to as “SCR0” henceforth¹ so that we may use that model as a point of reference

¹RC: It would be nice to have a running series of figures to display the various types of models. Each figure could have the same set of traps, use the same symbols, etc... It’s probably worth showing example data (and latent variables) in a table too

without having to provide a long-winded enumeration of assumptions and sampling design each time we do. We will make things more precise as we develop a formal statistical definition of the model shortly.

While the model is mostly directly relevant for hair snares and other DNA sampling methods for which multiple detections of an individual are not distinguishable, we will also make use of the model for data that arise from camera-trapping studies. In practice, with camera trapping, individuals might be photographed several times in a night but we will typically distill such data into a single binary encounter event for reasons discussed later in chapter 6.

4.2 THE BINOMIAL OBSERVATION MODEL

We assume that the individual and trap-specific encounters, y_{ij} , are mutually independent outcomes of a binomial random variable:

$$y_{ij} \sim \text{Bin}(K, p_{ij}) \quad (4.2.1)$$

This is the basic model underlying “logistic regression” (chapter 2) as well as standard closed population models (chapter 3). The key element of the model is that the encounter probability p_{ij} is indexed by (i.e., depends on) both individual and trap. In a sense, then, we can think of each *trap* as producing individual level encounter history data of the classical variety - an $n_{\text{ind}} \times n_{\text{rep}}$ matrix of 0's and 1's (this is the “encountered at most 1 time” assumption).

As we did in section XXX.YYY, we will make explicit the notion that p_{ij} is defined conditional on “where” individual i lives. Naturally, we think about defining an individual home range and then relating p_{ij} explicitly to the centroid of the individual's home range, or its center of activity (Efford, 2004; Borchers and Efford, 2008; Royle and Young, 2008). Therefore, define \mathbf{s}_i , a two-dimensional spatial coordinate, to be the activity center for individual i . Then, the basic SCR model postulates that encounter probability, p_{ij} , is related by a decreasing function to distance between trap j , having location \mathbf{x}_j , and \mathbf{s}_i . Naturally, if we think of modeling binomial counts using logistic regression, we might specify the model according to:

$$\text{logit}(p_{ij}) = \alpha_0 + \theta * ||\mathbf{s}_i - \mathbf{x}_j|| \quad (4.2.2)$$

where, here, $||\mathbf{s}_i - \mathbf{x}_j||$ is the distance between \mathbf{s}_i and \mathbf{x}_j . We sometimes write $||\mathbf{s}_i - \mathbf{x}_j|| = \text{dist}(\mathbf{s}_i, \mathbf{x}_j) = d_{ij}$. Alternatively, if we think about distance sampling then we might use the “half-normal” model of the form:

$$p_{ij} = p_0 * \exp(-\theta * ||\mathbf{s}_i - \mathbf{x}_j||^2)$$

Or any of a large number of standard detection models that are commonly used (we consider more in chapter XYZ). The half-normal model implies

$$\log(p_{ij}) = \log(p_0) - \theta * ||\mathbf{s}_i - \mathbf{x}_j||^2 \quad (4.2.3)$$

2802 Whatever model encounter probability we choose, we should always keep in mind
 2803 that the model is described conditional on \mathbf{s}_i , which is an unobserved random
 2804 variable. Thus, to be precise about this, we should write the observation model as

$$y_{ij}|\mathbf{s}_i \sim \text{Bin}(K, p(\mathbf{s}_{ij}; \theta))$$

2805 Note that we probably expect that the parameter θ in Eq. 4.2.2 or 4.2.3 should
 2806 be negative, so that the probability of encounter decreases with distance between the
 2807 trap and individual home range center. The joint likelihood for the data, conditional
 2808 on the collection of individual activity centers, can therefore be expressed as

$$\mathcal{L}(\theta|\{\mathbf{y}_i, \mathbf{s}_i\}_{i=1}^N) = \prod_i \prod_j \text{Bin}(y_{ij}|p_{ij}(\theta))$$

2809 Which, if we switch the indices on the product operators, this shows the SCR
 2810 likelihood (conditional on \mathbf{s}) to be the product of J *independent* capture-recapture
 2811 likelihoods - one for each trap. However, the data have a “repeated measures” type
 2812 of structure, with each of the j likelihood contributions for each individual being
 2813 grouped by individual. Thus, we cannot analyze the model meaningfully by J trap-
 2814 specific models. In classical repeated measures types of models, we accommodate
 2815 the group structure of the data using random effects (random individual or group
 2816 level variables). For SCR models we take the same basic approach, which we develop
 2817 subsequently.

2818 4.2.1 Distance as a latent variable

2819 If we knew precisely every \mathbf{s}_i in the population (and how many, N), then the model
 2820 specified by eqs. 4.2.1 and 4.2.2 or 4.2.3 is just an ordinary logistic regression type
 2821 of a model which we learned how to fit using **WinBUGS** previously (chapt. 2),
 2822 with a covariate d_{ij} . However, the activity centers are unobservable even in the best
 2823 possible circumstances. In that case, d_{ij} is an unobserved variable, analogous to
 2824 classical “random effects” models. We need to therefore extend the model to accom-
 2825 modate these random variables with an additional model component. A standard,
 2826 and perhaps not unreasonable, assumption is the so-called “uniformity assumption”
 2827 which is to say that the \mathbf{s}_i are uniformly distributed over space (the obvious next
 2828 question “which space?” is addressed below). This uniformity assumption amounts
 2829 to a uniform prior distribution on \mathbf{s}_i , i.e., the pdf of \mathbf{s}_i is constant, which we may
 2830 express

$$\Pr(\mathbf{s}_i) \text{propto} \text{const} \quad (4.2.4)$$

2831 To summarize the preceeding model developing, a basic SCR model is defined
 2832 by 3 essential components:

- 2833 (1) Observation model: $y_{ij}|\mathbf{s}_i \sim \text{Bin}(K, p_{ij})$
- 2834 (2) Encounter probability: $\text{logit}(p_{ij}) = \alpha_0 + \theta * \|\mathbf{s}_i - \mathbf{x}_j\|$

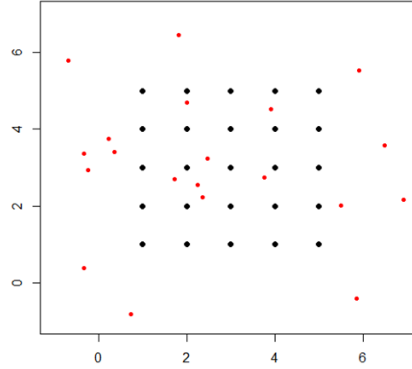


Figure 4.1. Realization of a binomial point process

2835 (3) Point process model: $\Pr[\mathbf{s}_i] \propto \text{const}$

2836 Therefore, the SCR model is little more than an ordinary capture-recapture model
 2837 for closed populations. It is such a model, but augmented with a set of “individual
 2838 effects”, \mathbf{s}_i , which relate some sense of individual location to encounter probability.
 2839 As it turns out, assumption (3) is usually not precise enough to fit a model in
 2840 practice for reasons we discuss in the following section. We will give another way to
 2841 represent this prior distribution that is more concrete, but it depends on specifying
 2842 the “state-space” of the random variable \mathbf{s}_i . The term “state-space” is a technical
 2843 way of saying “possible outcomes”.

4.3 THE BINOMIAL POINT-PROCESS MODEL

2844 The collection of individual activity centers $\mathbf{s}_1, \dots, \mathbf{s}_N$ represent a realization of a
 2845 *binomial point process* (Illian, 2008a, p. xyz). The binomial point process (BPP)
 2846 is analogous to a Poisson point process in the sense that it represents a “random
 2847 scatter” of points in space - except that the total number of points is *fixed*, whereas,
 2848 in a Poisson point process it is random (having a Poisson distribution). As an
 2849 example, we show in Fig. 4.1 locations of 20 individual activity centers (black
 2850 dots) in relation to a grid of 25 traps. For a Poisson point process the number of
 2851 such points in the prescribed state-space would be random whereas often we will
 2852 simulate fixed numbers of points, e.g., for evaluating the performance of procedures
 2853 such as how well does our estimator perform of $N = 50$?

It is natural to consider a binomial point process in the context of capture-recapture models because it preserves N in the model and thus preserves the linkage directly with closed population models. In fact, under the binomial point process model then Model M0 and other closed models are simple limiting cases of SCR models. In addition, use of the BPP model allows us to use data augmentation for Bayesian analysis of the models as in chapter 3, thus yielding a methodologically coherent approach to analyzing the different classes of models. Despite this, making explicit assumptions about N , such as Poisson, is convenient in some cases (see chapt. XYZ).

One consequence of having fixed N , in the BPP model, is that the model is not strictly a model of “complete spatial randomness”. This is because if one forms counts $n(A_1), \dots, n(A_k)$ in any set of disjoint regions say A_1, \dots, A_k , then these counts are *not* independent. In fact, they have a multinomial distribution (see Illian, 2008a, p. XYZ). Thus, the BPP model introduces a slight bit of dependence in the distribution of points. However, in most situations this will have no practical effect on any inference or analysis and, as a practical matter, we will usually regard the BPP model as one of spatial independence among individual activity centers because each activity center is distributed independently of each other activity center. Despite this implicit independence we see in Fig. 4.1 that realizations of randomly distributed points will typically exhibit distinct non-uniformity. Thus, independent, uniformly distributed points will almost never appear regularly, uniformly or systematically distributed. For this reason, the basic binomial (or Poisson) point process models are enormously useful in practical settings. More relevant for SCR models is that we actually have a little bit of data for some individuals and thus the resulting posterior point pattern can deviate strongly from uniformity (we should note this elsewhere too). The uniformity hypothesis is only a *prior* distribution which is directly affected by the quantity and quality of observations.

4.3.1 Definition of home range center

Some will be offended by our use of the concept of “home range center” and thus will have difficulty in believing that the resulting model is really useful for anything. Indeed, the idea of a home range or activity center is a vague concept anyway, a purely phenomenological construct. Despite this, it doesn’t really matter whether or not a home range makes sense for a particular species - individuals of any species inhabit *some* region of space and we can define the “home range center” to be the center of the space that individual was occupying (or using) during the period in which traps were active. Thinking about it in that way, it could even be observable (almost) as the centroid of a very large number of radio fixes over the course of a survey period or a season. Thus, this practical version of a home range center is a well-defined construct regardless of whether one thinks the home range concept is meaningful, even if individuals are not particularly territorial. This is why we usually use the term “activity center” or maybe even “centroid of space usage”

and we recognize that this construct is a transient thing which applies only to a well-defined period of study.

4.3.2 The state-space of the point process

Shortly we will focus on Bayesian analysis of this model with N known so that we can directly apply what we learned in chapter 2 to this situation. To do this, we note that the individual effects $\mathbf{s}_1, \dots, \mathbf{s}_N$ are unknown quantities and we will need to be able to simulate each \mathbf{s}_i in the population from the posterior distribution. It should be self-evident that we cannot simulate the \mathbf{s}_i unless we describe precisely the region over which those \mathbf{s}_i 's are uniformly distributed. This is the quantity referred to above as the state-space, denoted henceforth by \mathcal{S} , which is a region or a set of points comprising the potential values of \mathbf{s}_i . Thus, an equivalent explicit statement of the “uniformity assumption” is

$$\mathbf{s}_i \sim \text{Unif}(\mathcal{S})$$

Prescribing the state-space

Evidently, we need to define the state-space, \mathcal{S} . How can we possibly do this objectively? Prescribing any particular \mathcal{S} seems like the equivalent of specifying a “buffer” which we criticized previously as being ad hoc. How is it that choosing a state-space is *not* ad hoc? As a practical matter, it turns out that estimates of density are insensitive to choice of the state-space. As we observed in chapter 3, it is true that N increases with \mathcal{S} , but only at the same rate as \mathcal{S} under the prior assumption of constant density. As a result, we say that density is invariant to \mathcal{S} as long as \mathcal{S} is sufficiently large. Thus, while choice of \mathcal{S} is (or can be) essentially arbitrary, once \mathcal{S} is chosen, it defines the population being exposed to sampling, which scales appropriately with the size of the state-space.

For our simulated system developed previously in this chapter, we defined the state space to be a square within which our traps were centered perfectly. For many practical situations this might be an acceptable approach to defining the state-space. We provide an example of this in section 4.7 below in which the trap array is irregular and also situated within a realistic landscape that is distinctly irregular. In general, it is most practical to define the state-space as a regular polygon (e.g., rectangle) containing the trap array without differentiating unsuitable habitat. Although defining the state-space to be a regular polygon has computational advantages (e.g., we can implement this more efficiently in **WinBUGS** and cannot for irregular polygons), a regular polygon induces an apparent problem of admitting into the state-space regions that are distinctly non-habitat (e.g., oceans, large lakes, ice fields, etc.). It is difficult to describe complex sets in mathematical terms that can be admitted to this spatial model. As an alternative, we can provide a representation of the state-space as a discrete set of points (section 4.9) that will allow specific points to be deleted or not depending on whether they represent habitat, or we can define the state-space as an intersection of polygons, and analysis of

models with state-space defined in that way can be analyzed easily using MCMC (see section XYZ in chapt. 6). In what follows below we provide an analysis of the camera data defining the state-space to be a regular continuous polygon (a rectangle).

4.3.3 Invariance and the State-space as a model assumption

We will assert for all models we consider in this book that density is invariant to the size and extent of \mathcal{S} , if \mathcal{S} is sufficiently large. In fact, this only holds as long as our model relating p_{ij} to \mathbf{s}_i is a decreasing function of distance. We can prove this thinking about a 1-d case where $E[y]$ for the “last cell” (i.e., for $d > B$ for B large enough) is 0. So it always contributes nothing to the likelihood, i.e., $E[n(\text{lastcell})] = 0$. [sketch out a proof of this], in regular situations in which the detection function decays monotonically with distance and prior density is constant. Sometimes our estimate of density can be influenced if we make \mathcal{S} too small but this might be sensible if \mathcal{S} is naturally well-defined. As we discussed in chapter 1, **choice of \mathcal{S} is part of the model and thus it makes sense that estimates of density might be sensitive to its definition in problems where it is natural to restrict \mathcal{S} .** One could imagine however that in specific cases where you’re studying a small population with well-defined habitat preferences that a problem could arise because changing the state-space around based on differing opinions and GIS layers really changes the estimate of total population size. But this is a real biological problem and a natural consequence of the spatial formalization of capture-recapture models - a feature, not a bug or some statistical artifact - and it should be resolved with better information and research, and not some arbitrary statistical artifact. For situations where there is not a natural choice of \mathcal{S} , we should default to choosing \mathcal{S} to be very large in order to achieve invariance or otherwise evaluate sensitivity of density estimates by trying a couple of different values of \mathcal{S} . This is a standard “sensitivity to prior” argument that Bayesians always have to be conscious of. We demonstrate this in our analysis of section 4.7 below. Note that $area(\mathcal{S})$ affects data augmentation. If you increase $area(\mathcal{S})$ then there are more individuals to account for and therefore the size of the augmented data set M must increase.

We have been told that one can carry-out non-Bayesian analyses of SCR models without having to specify the state-space of the point process or perhaps while only specifying it imprecisely. This assertion is incorrect. We assume people are thinking this because *they* don’t have to specify it explicitly because someone else has done it for them in a package that does integrated likelihood. Even to do integrated likelihood (see chapter 6) we have to integrate the conditional-on- \mathbf{s} likelihood over some 2-dimensional space. It might work that the integration can be done from $-\infty$ to $+\infty$ but that is a mathematical artifact of specific detection functions, and an implicit definition of a state-space that doesn’t make biological sense, even though it may in fact be innocuous;

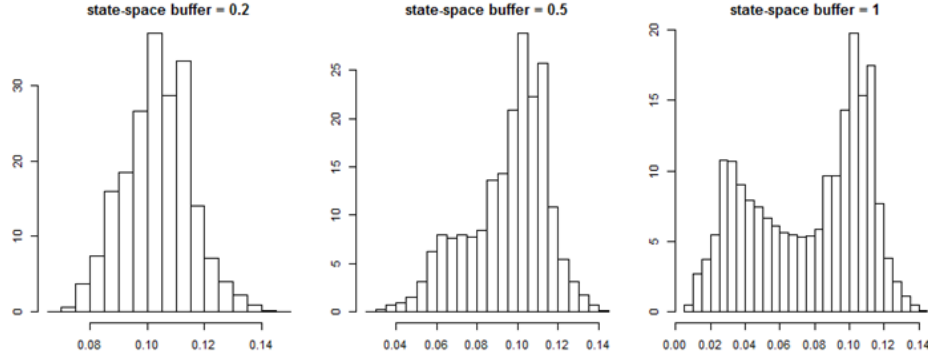


Figure 4.2. Needs a caption

4.3.4 Connection to Model Mh

SCR models are closely related to heterogeneity models. In SCR models, heterogeneity in encounter probability is induced by both the effect of distance in the model for detection probability and also from specification of the state-space. Clearly then the state-space is explicitly part of the model. To understand this, we have a random effect with some prior distribution:

$$\mathbf{s} \sim \text{uniform}(\mathcal{S})$$

And $p(\mathbf{s}) = p(y = 1|\mathbf{s})$ is some function of \mathbf{s} . Therefore, for any specific $g(p)$ and \mathcal{S} we can work out what the implied heterogeneity model is for example, the mean, variance or other moments of the population distribution of p can be evaluated by integrating $p(\mathbf{s})$ over the state-space of \mathbf{s} . Obviously the choice of $p(\mathbf{s})$ and the choice of \mathcal{S} interact to determine the effective heterogeneity in p . We show an illustration in Fig. 4.2 below which shows a histogram of p for a hypothetical population of 100000 individuals on a state-space enclosing our 5×5 trap array above, under the logistic model for distance. **R** code is provided in the **R** package **scrbook** to produce this analysis for the logistic and half-normal models. The histogram shows the encounter probability under buffers of 0.2, 0.5 and 1.0. We see the mass shifts to the left as the buffer increases, implying more individuals in the population but with lower encounter probability as their home range centers increase in distance from the trap array.

Another way to understand this is by representing \mathcal{S} as a set of discrete points on a grid. In the coarsest possible case where \mathcal{S} is a single arbitrary point, then every individual has exactly the same p . As we increase the number of points in \mathcal{S} then more distinct values of p are possible. As such, when \mathcal{S} is characterized by discrete points then SCR models are precisely a type of finite-mixture model (Norris III and

Pollock, 1996; Pledger, 2000), except where we have some information about which group an individual belong (i.e., where their activity center is), as a result of their captures in traps.

This context suggests the problem raised by Link (2003). He showed that in most practical situations N may not be identifiable across classes of mixture distributions which in the context of SCR models is the pair (g, \mathcal{S}) . The difference, however, is that we do obtain some direct information about \mathbf{s} in SCR models and therefore N is identifiable across models characterized by (g, \mathcal{S}) .

4.3.5 Connection to Distance Sampling

It is worth emphasizing that the basic SCR model is a binomial encounter model in which distance is a covariate. As such, it is striking similarity to a classical distance sampling model. Both have distance as a covariate but in classical distance sampling problems the focus is on the distance between the observer and the animal at an instant in time, not the distance between a trap and an animal's home range center. Thus in distance sampling, "distance" is *observed* for those individuals that appear in the sample. Conversely, in SCR problems, it is only imperfectly observed (we have partial information in the form of trap observations). Clearly, it is preferable to observe distance if possible, but as we will discuss in chapter XYZ, distance sampling requires field methods that are often not practical in many situations, e.g. when surveying tigers. Furthermore, SCR models allow us to relax many of the assumption made in classical distance sampling, and SCR models allow for estimates of quantities other than density, such as home range size.

4.4 SIMULATING SCR DATA

It is always useful to simulate data because it allows you to understand the system that you're modeling and also calibrate your understanding with the parameter values of the model. That is, you can simulate data using different parameter values until you obtain data that "looks right" based on your knowledge of the specific situation that you're interested in. Here we provide a simple script to illustrate how to simulate spatial encounter history data. In this exercise we simulate data for 100 individuals and a 25 trap array laid out in a 5×5 grid of unit spacing. The specific encounter model is the half-normal model given above and we used this code to simulate data used in subsequent analyses. The 100 activity centers were simulated on a state-space defined by a 8×8 square within which the trap array was centered (thus the trap array is buffered by 2 units). Therefore, the density of individuals in this system is fixed at $100/64$.

```
set.seed(2013)
# create 5 x 5 grid of trap locations with unit spacing
traplocs<- cbind(sort(rep(1:5,5)),rep(1:5,5))
```

```

3036 Dmat<-e2dist(traplocs,traplocs) # in cases where speed doesn't matter, it might be
3037                                # clearer to just show the slow for-loop.
3038                                # Plus, people will want to copy/paste this stuff
3039 ntraps<-nrow(traplocs)
3040
3041 # define state-space of point process. (i.e., where animals live).
3042 # "delta" just adds a fixed buffer to the outer extent of the traps.
3043 delta<-2
3044 Xl<-min(traplocs[,1] - delta)
3045 Xu<-max(traplocs[,1] + delta)
3046 Yl<-min(traplocs[,2] - delta)
3047 Yu<-max(traplocs[,2] + delta)
3048
3049 N<-100    # population size
3050 K<- 20    # number nights of effort
3051
3052 sx<-runif(N,Xl,Xu)    # simulate activity centers
3053 sy<-runif(N,Yl,Yu)
3054 S<-cbind(sx,sy)
3055 D<- e2dist(S,traplocs) # distance of each individual from each trap
3056
3057 alpha0<- -2.5    # define parameters of encounter probability
3058 sigma<- 0.5      #
3059 theta<- 1/(2*sigma*sigma)
3060 probcap<- expit(-2.5)*exp( - theta*D*D)    # probability of encounter
3061 # now generate the encounters of every individual in every trap
3062 Y<-matrix(NA,nrow=N,ncol=ntraps)
3063 for(i in 1:nrow(Y)){
3064     Y[i,<-rbinom(ntraps,K,probcap[i,])
3065 }

```

Subsequently we will generate data using this code packaged in an R function called `simSCR0.fn` which takes a number of arguments including `discard0` which, if `TRUE`, will return only the encounter histories for captured individuals. A second argument is `array3d` which, if `TRUE`, returns the 3-d encounter history array instead of the aggregated `nind × ntraps` encounter frequencies (see below). Finally we provide a random number seed, `sd` which we always set to 2013 in our analyses. Thus we obtain a data set as above using the following command

```

3073 data<-simSCR0.fn(discard0=TRUE,array3d=FALSE,sd=2013)

```

The **R** object `data` is a list, so let's take a look at what's in the list and then harvest some of its elements for further analysis below.

```

3076 > names(data)
3077 [1] "Y"          "traplocs" "xlim"      "ylim"      "N"          "alpha0"    "beta"
3078 [8] "sigma"      "K"

```

```

3079 > Y<-data$Y
3080 > traplocs<-data$traplocs

```

3081 4.4.1 Formatting and manipulating real data sets

3082 Conventional capture-recapture data are easily stored and manipulated as a 2-
 3083 dimensional array, an $nind \times nperiod$ matrix, which is maximally informative for
 3084 any conventional capture-recapture model, but not for spatial capture-recapture
 3085 models. For SCR models we must preserve the spatial information in the encounter
 3086 history information. We will routinely analyze data from 3 standard formats:

- 3087 (1) The basic 2-dimensional data format, which is an $nind \times ntraps$ encounter
 3088 frequency matrix such as that simulated previously;
- 3089 (2) The maximally informative 3-dimensional array which we establish here the
 3090 convention that it has dimensions $nind \times nperiods \times ntraps$ and
- 3091 (3) We use a compact format - the “SCR flat format” - which we describe below
 3092 in section 4.7.

3093 To simulate data in the most informative format - the “3-d array” - we can use the
 3094 **R** commands given previously but replace the last 4 lines with the following:

```

3095 Y<-array(NA,dim=c(N,K,ntraps))
3096 for(i in 1:nrow(Y)){
3097   for(j in 1:ntraps){
3098     Y[i,1:K,j]<-rbinom(K,1,probcap[i,j])
3099   }
3100 }

```

3101 We see that a collection of K binary encounter events are generated for *each*
 3102 individual and for *each* trap. The probabilities have those Bernoulli trials are
 3103 computed based on the distance from each individuals home range center and the
 3104 trap (see calculation above), and those are housed in the matrix probcap. Our
 3105 data simulator function `simSRC0.fn` will return the full 3-d array if `array3d=TRUE`
 3106 is specified in the function call. To recover the 2-d matrix from the 3-d array, and
 3107 subset the 3-d array to individuals that were captured, we do this:

```

3108 Y2d<- apply(Y,c(1,3),sum) # sum over the ‘‘replicates’’ dimension (2nd margin of the array)
3109 ncaps<-apply(Y2d,1,sum)   # compute how many times each individual was captured
3110 Y<-Y[ncaps>0,,]          # keep those individuals that were captured

```

4.5 FITTING AN SCR MODEL IN BUGS

3111 Clearly if we somehow knew the value of N then we could fit this model directly
 3112 because, in that case, it is a special kind of logistic regression model - one with a
 3113 random effect, but that enters into the model in a peculiar fashion - and also with
 3114 a distribution (uniform) which we don’t usually think of as standard for random

3115 effects models. So our aim here is to analyze the known- N problem, using our
 3116 simulated data, as an incremental step in our progress toward fitting more generally
 3117 useful models.

3118 To begin, we use our simulator to grab a data set and then harvest the elements
 3119 of the resulting object for further analysis.

```
3120 data<-simSCR0.fn(discard0=FALSE,sd=2013)
3121 y<-data$Y
3122 traplocs<-data$traplocs
3123 nind<-nrow(y)
3124 X<-data$traplocs
3125 J<-nrow(X)
3126 y<-rbind(y,matrix(0,nrow=(100-nrow(y)),ncol=J ) )
3127 Xl<-data$xlim[1]
3128 Yl<-data$ylim[1]
3129 Xu<-data$xlim[2]
3130 Yu<-data$ylim[2]
```

3131 Note that we specify `discard0 = FALSE` so that we have a "complete" data set,
 3132 i.e., one with the all-zero encounter histories corresponding to uncaptured individ-
 3133 uals. Now, within an **R** session, we can create the **BUGS** model file and fit the
 3134 model using the following commands. This model describes the half-normal detec-
 3135 tion model but it would be trivial to modify that to various others including the
 3136 logistic described above. One consequence of using the half-normal is that we have
 3137 to constrain the encounter probability to be in $[0, 1]$ which we do here by defining
 3138 `alpha0` to be the logit of the intercept parameter `p0`. Note that the distance co-
 3139 variate is computed within the **BUGS** model specification given the matrix of trap
 3140 locations, `X`, which is provided to **WinBUGS** as data.

```
3141 cat("
3142 model {
3143   alpha0~dnorm(0,.1)
3144   logit(p0)<- alpha0
3145   theta~dnorm(0,.1)
3146   for(i in 1:N){
3147     s[i,1]~dunif(Xl,Xu)
3148     s[i,2]~dunif(Yl,Yu)
3149     for(j in 1:J){
3150       d[i,j]<- pow(pow(s[i,1]-X[j,1],2) + pow(s[i,2]-X[j,2],2),0.5)
3151       y[i,j] ~ dbin(p[i,j],K)
3152       p[i,j]<- p0*exp(- theta*d[i,j]*d[i,j])
3153     }
3154   }
3155 }
3156 ",file = "SCR0a.txt")
3157
```

Next we do a number of organizational activities including bundling the data for **WinBUGS**, defining some initial values, the parameters to monitor and some basic MCMC settings. We choose initial values for the activity centers **s** by generating uniform random numbers in the state-space but, for the observed individuals, we replace those values by each individual's mean trap coordinate for all encounters

```

3158 sst<-cbind(runif(nind,Xl,Xu),runif(nind,Yl,Yu)) # starting values for s
3159 for(i in 1:nind){
3160   if(sum(y[i,])==0) next
3161   sst[i,1]<- mean( X[y[i,]>0,1] )
3162   sst[i,2]<- mean( X[y[i,]>0,2] )
3163 }
3164
3165 data <- list (y=y,X=X,K=K,N=nind,J=J,Xl=Xl,Yl=Yl,Xu=Xu,Yu=Yu)
3166 inits <- function(){
3167   list (alpha0=rnorm(1,-4,.4),theta=runif(1,1,2),s=sst)
3168 }
3169
3170 library("R2WinBUGS")
3171 parameters <- c("alpha0","theta")
3172 nthin<-1
3173 nc<-3
3174 nb<-1000
3175 ni<-2000
3176 out <- bugs (data, inits, parameters, "SCROa.txt", n.thin=nthin,
3177 n.chains=nc, n.burnin=nb,n.iter=ni,debug=TRUE,working.dir=getwd())

```

There is little to say about the preceding basic operations other than to suggest that the interested reader explore the output and additional analyses by running the script provided in the **R** package **scrbook**. We ran 1000 burn-in and 1000 after burn-in, 3 chains, to obtain 3000 posterior samples. Because we know N for this particular data set we only have 2 parameters of the detection model to summarize (**alpha0** and **theta**). When the object **out** is produced we print a summary of the results as follows:

```

3190 > print(out,digits=3)
3191 Inference for Bugs model at "SCROa.txt", fit using WinBUGS,
3192 3 chains, each with 2000 iterations (first 1000 discarded)
3193 n.sims = 3000 iterations saved
3194
3195      mean      sd    2.5%    25%    50%    75%   97.5%  Rhat n.eff
3196 alpha0   -2.496  0.224  -2.954  -2.648  -2.48  -2.340  -2.091  1.013   190
3197 theta     2.442  0.419   1.638   2.145   2.44   2.721   3.303  1.005   530
3198 deviance 292.803 21.155 255.597 277.500 291.90 306.000 339.302 1.006   380
3199
3200 For each parameter, n.eff is a crude measure of effective sample size,
3201 and Rhat is the potential scale reduction factor (at convergence, Rhat=1).
3202
3203 DIC info (using the rule, pD = Dbar-Dhat)

```


3203 `pD = -138.8 and DIC = 154.0`
 3204 `DIC is an estimate of expected predictive error (lower deviance is better).`

3205 We know the data were generated with `alpha0 = -2.5` and `theta = -2`. The
 3206 estimates look reasonably close to those data-generating values and we probably feel
 3207 pretty good about the performance of the Bayesian analysis and MCMC algorithm
 3208 that WinBUGS cooked-up based on our sample size of 1 data set. It is worth noting
 3209 that the Rhat statistics indicate reasonable convergence but, as a practical matter,
 3210 we might choose to run the MCMC algorithm for additional time to bring these
 3211 closer to 1.0 and to increase the effective posterior sample size (`n.eff`). Other
 3212 summary output includes “deviance” and related things including the deviance
 3213 information criterion (DIC). We discuss these things in chapter XXXX.

4.6 UNKNOWN N

3214 In all real applications N is unknown and that fact is kind of an important feature
 3215 of the capture-recapture problem! We handled this important issue in chapter 3
 3216 using the method of data augmentation which we apply here to achieve a realistic
 3217 analysis of Model SCR0. As with the basic closed population models considered
 3218 previously, we formulate the problem here by augmenting our observed data set
 3219 with a number of “all zero” encounter histories - what we referred to in Chapter
 3220 3 as potential individuals. If n is the number of observed individuals, then let
 3221 $M - n$ be the number of potential individuals in the data set. For the basic y_{ij}
 3222 data structure (individuals x traps encounter frequencies) we simply add additional
 3223 rows of “all 0” observations to that data set. This is because such “individuals” are
 3224 unobserved, and therefore necessarily have $y_{ij} = 0$ for all j . A data set, say with 4
 3225 traps and 6 individuals, augmented with 4 pseudo-individuals therefore might look
 3226 like this:

	trap1	trap2	trap3	trap4
[1,]	1	0	0	0
[2,]	0	2	0	0
[3,]	0	0	0	1
[4,]	0	1	0	0
[5,]	0	0	1	1
[6,]	1	0	1	0
[7,]	0	0	0	0
[8,]	0	0	0	0
[9,]	0	0	0	0
[10,]	0	0	0	0

3238 We typically have more than 4 traps and, if we’re fortunate, many more indi-
 3239 viduals in our data set.

3240 For the augmented data, we introduce a set of binary latent variables (the data
 3241 augmentation variables), z_i , and the model is extended to describe $\Pr(z_i = 1)$
 3242 which is, in the context of this problem, the probability that an individual in the

augmented data set is a member of the population that was sampled. In other words, if $z_i = 1$ for one of the “all zero” encounter histories, this is implied to be a sampling zero whereas observations for which $z_i = 0$ are “structural zeros” under the model.

How big does the augmented data set have to be? We discussed this issue in chapt. 3 where we noted that the size of the data set is equivalent to the upper limit of a uniform prior distribution on N . Practically speaking, it should be sufficiently large so that the posterior distribution for N is not truncated. On the other hand, if it is too large then unnecessary calculations are being done. An approach to choosing M by trial-and-error is indicated. You can take a ballpark estimate of the probability that an individual is captured (at all during the study), obtain N as $n/pcap$, and then set $M = 2 * N$, as a first guess. Do a short MCMC run and then consider whether you need to do something different. See chapt. 7 for an example of this. Kery and Schaub (2011, ch. 6) provide an assessment of choosing M in closed population models.

Analysis by data augmentation removes N as an explicit parameter of the model. Instead, N is a derived parameter, computed by $N = \sum_{i=1}^M z_i$. Similarly, *density*, D , is also a derived parameter computed as $D = N/area(\mathcal{S})$. For our simulator, we’re using an 8×8 state-space and thus we will compute D as $D = N/64$.

4.6.1 Analysis using data augmentation in WinBUGS

As before we begin by obtaining a data set using our `simSCR0.fn` routine and then harvesting the required data objects from the resulting data list. Note that we use the `discard0=TRUE` option this time so that we get a “real” data set with no all-zero encounter histories. After harvesting the data we produce the **WinBUGS** model specification which now includes M encounter histories including the augmented potential individuals, the data augmentation parameters z_i , and the data augmentation parameter ψ .

```

data<-simSCR0.fn(discard0=TRUE,sd=2013)
y<-data$Y
traplocs<-data$traplocs
nind<-nrow(y)
X<-data$traplocs
J<-nrow(X)
Xl<-data$xlim[1]
Yl<-data$ylim[1]
Xu<-data$xlim[2]
Yu<-data$ylim[2]

cat("
model {
  alpha0~dnorm(0,.1)
  logit(p0)<- alpha0

```

```

3285 theta~dnorm(0,.1)
3286 psi~dunif(0,1)
3287
3288 for(i in 1:M){
3289   z[i] ~ dbern(psi)
3290   s[i,1]~dunif(Xl,Xu)
3291   s[i,2]~dunif(Yl,Yu)
3292   for(j in 1:J){
3293     d[i,j]<- pow(pow(s[i,1]-X[j,1],2) + pow(s[i,2]-X[j,2],2),0.5)
3294     y[i,j] ~ dbin(p[i,j],K)
3295     p[i,j]<- z[i]*p0*exp(- theta*d[i,j]*d[i,j])
3296   }
3297 }
3298 N<-sum(z[])
3299 D<-N/64
3300 }
3301 ",file = "SCR0a.txt")

```

3302 To prepare our data we have to augment the data matrix y with $M - n$ all-
 3303 zero encounter histories, we have to create starting values for the variables z_i and
 3304 also the activity centers s_i of which, for each, we require M values. Otherwise the
 3305 remainder of the code for bundling the data, creating initial values and executing
 3306 **WinBUGS** looks much the same as before except with more or differently named
 3307 arguments.

```

3308 ## Data augmentation stuff
3309 M<-200
3310 y<-rbind(y,matrix(0,nrow=M-nind,ncol=ncol(y)))
3311 z<-c(rep(1,nind),rep(0,M-nind))
3312
3313 sst<-cbind(runif(M,Xl,Xu),runif(M,Yl,Yu)) # starting values for s
3314 for(i in 1:nind){
3315   if(sum(y[i,])==0) next
3316   sst[i,1]<- mean( X[y[i,]>0,1] )
3317   sst[i,2]<- mean( X[y[i,]>0,2] )
3318 }
3319 data <- list (y=y,X=X,K=K,M=M,J=J,Xl=Xl,Yl=Yl,Xu=Xu,Yu=Yu)
3320 inits <- function(){
3321   list (alpha0=rnorm(1,-4,.4),theta=runif(1,1,2),s=sst,z=z)
3322 }
3323
3324 library("R2WinBUGS")
3325 parameters <- c("alpha0","theta","N")
3326 nthin<-1
3327 nc<-3
3328 nb<-1000
3329 ni<-2000

```

```

3330 out <- bugs (data, inits, parameters, "SCR0a.txt", n.thin=nthin,n.chains=nc,
3331   n.burnin=nb,n.iter=ni,debug=TRUE,working.dir=getwd())

```

3332 **Remarks:** (1) Note the differences in this new **WinBUGS** model with that
 3333 appearing in the known- N version. (2) Also the input data has changed - the
 3334 augmented data set has more rows of all-zeros. Previously we knew that $N = 100$
 3335 but in this analysis we pretend not to know N , but think that $N = 200$ is a good
 3336 upper-bound; (3) Population size $N(S)$ is a derived parameter, being computed by
 3337 summing up all of the data augmentation variables z_i (as we've done previously);
 3338 (4) Density, $D \equiv D(S)$, is also a derived parameter. Summarizing the output from
 3339 **WinBUGS** produces:

```

3340 > print(out1,digits=2)
3341 Inference for Bugs model at "SCR0a.txt", fit using WinBUGS,
3342   3 chains, each with 2000 iterations (first 1000 discarded)
3343   n.sims = 3000 iterations saved
3344
3345      mean    sd   2.5%   25%   50%   75%  97.5% Rhat n.eff
3346 alpha0  -2.57  0.23  -3.04  -2.72  -2.56  -2.41  -2.15  1.01   320
3347 theta    2.46  0.42   1.63   2.16   2.46   2.73   3.33  1.02   120
3348 N       113.62 15.73  86.00 102.00 113.00 124.00 147.00 1.01   260
3349 D         1.78  0.25   1.34   1.59   1.77   1.94   2.30  1.01   260
3350 deviance 302.60 23.67 261.19 285.47 301.50 317.90 354.91 1.00  1400
3351
3352 For each parameter, n.eff is a crude measure of effective sample size,
3353 and Rhat is the potential scale reduction factor (at convergence, Rhat=1).
3354
3355 DIC info (using the rule, pD = var(deviance)/2)
3356 pD = 279.9 and DIC = 582.5
3357 DIC is an estimate of expected predictive error (lower deviance is better).

```

3357 The column labeled “MC error” is the Monte Carlo error - the error inherent in
 3358 the attempt to compute these posterior summaries by MCMC. It is desirable to run
 3359 the Markov chain algorithm long enough so as to reduce the MC error to a tolerable
 3360 level. What constitutes tolerable is up to the investigator. Certainly less than 1% is
 3361 called for. As a general rule, Rhat gets closer to 1 and MC error decreases toward 0
 3362 as the number of iterations increases. We see that the estimated parameters (α_0 and
 3363 θ) are comparable to the previous results obtained for the known- N case, and also
 3364 not too different from the data-generating values. The posterior of N overlaps the
 3365 data-generating value substantially with a mean of 113.62. To obtain these results
 3366 we fitted the true data-generating model, that based on the half-normal detection
 3367 model, to a single simulated data set. For fun and excitement we fit the *wrong*
 3368 model - that with the logistic-linear detection model - to the same data set. This is
 3369 easily achieved by modifying the **WinBUGS** model specification above, although
 3370 we provide the **R** script in the **R** package **scrbook**. Those results are given below.
 3371 We see that the estimate of N , the main parameter of interest, is very similar to
 3372 that obtained under the correct model, convergence is worse (as measured by Rhat)
 3373 which probably doesn't have anything to do with the model being wrong, and the

posterior deviance and DIC favor the correct model. We consider the use of DIC for carrying-out model selection in chapter 8.

```

> print(out2,digits=2)
Inference for Bugs model at "SCR0a.txt", fit using WinBUGS,
  3 chains, each with 2000 iterations (first 1000 discarded)
  n.sims = 3000 iterations saved

```

	mean	sd	2.5%	25%	50%	75%	97.5%	Rhat	n.eff
alpha0	-1.59	0.27	-2.16	-1.77	-1.58	-1.42	-1.07	1.05	60
beta	3.77	0.43	2.92	3.48	3.79	4.05	4.66	1.04	70
N	122.57	18.67	90.00	109.00	122.00	135.00	163.00	1.00	3000
D	1.92	0.29	1.41	1.70	1.91	2.11	2.55	1.00	3000
deviance	312.67	22.43	271.00	297.20	311.50	327.00	359.60	1.02	130

```

For each parameter, n.eff is a crude measure of effective sample size,
and Rhat is the potential scale reduction factor (at convergence, Rhat=1).

DIC info (using the rule, pD = var(deviance)/2)
pD = 247.5 and DIC = 560.1
DIC is an estimate of expected predictive error (lower deviance is better).

```

4.6.2 Use of other BUGS engines: JAGS

There are two other popular **BUGS** engines in widespread use: **OpenBUGS** (Thomas et al., 2006) and **JAGS** (Plummer, 2003). Both of these are easily called from **R**. **OpenBUGS** can be used instead of **WinBUGS** by changing the package option in the bugs call to `package=OpenBUGS`. **JAGS** can be called using the function `jags()` in package **R2JAGS** which has nearly the same arguments as `bugs()`. We prefer to use the **R** library `rjags` (Plummer, 2009) which has a slightly different implementation that we demonstrate here as we reanalyze the simulated data set in the previous section (note: the same **R** commands are used to generate the data and package the data, inits and parameters to monitor). The function `jags.model` is used to initialize the model and run the MCMC algorithm for a period in which adaptive rejection (XXXX not sure XXXXX???) sampling is used. Then the Markov chains are updated using `coda.samples()` to obtain posterior samples for analysis, as follows:

```

jm<- jags.model("SCR0a.txt", data=data, inits=inits, n.chains=nc,
               n.adapt=nb))
jm<- coda.samples(jm, parameters, n.iter=ni-nb, thin=nthin)

```

We find that JAGS seems to be 20-30% faster for the basic SCR model which the reader can evaluate using the script `jags.winbugs.R` in the **R** package `scrbook`.

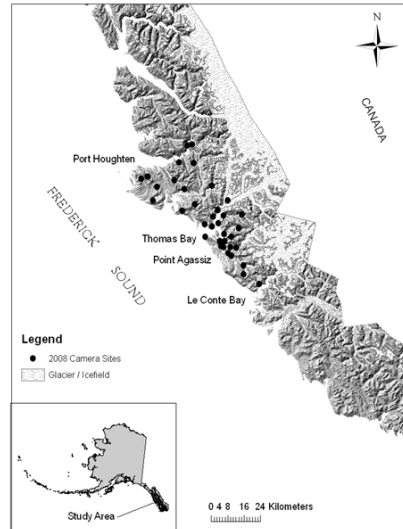


Figure 4.3. Wolverine camera trap locations from Magoun et al. (2011).

4.7 CASE STUDY: WOLVERINE CAMERA TRAPPING STUDY

We provide an analysis here of A. Magoun’s wolverine data (Magoun et al., 2011; Royle et al., 2011c). The study took place in SE Alaska (Fig. 4.3) where 37 cameras were operational for variable periods of time (min = 5 days, max = 108 days, median = 45 days). A consequence of this is that the binomial sample size K (see Eq. 4.2.1) is variable for each camera. Thus, we must provide a matrix of sample sizes as data to BUGS and modify the model specification in sec. 4.6 accordingly. Our treatment of the data here is based on the analysis of Royle et al. (2011c).

To carry-out an analysis of these data, we require the matrix of trap coordinates and the encounter history data. We store data in an the “scr flat format” (see sec. 4.4.1 above), an efficient file format which is easily manipulated and also used as the input file format in our custom **R** script (ch. xxx) and **SPACECAP** (Gopalaswamy, 2012). To illustrate this format, the wolverine data are available as an encounter data **R** object named “**wcaps**” which has 3 columns and 115 rows, each representing a unique encounter event including the trap identity, the individual identity and the sample occasion index (**sample**). The first 10 rows of this matrix are as follows:

```
> wcaps
      trapid individual sample
```

3431	[1,]	1	2	127
3432	[2,]	1	2	128
3433	[3,]	1	2	129
3434	[4,]	1	18	130
3435	[5,]	2	3	106
3436	[6,]	2	18	104
3437	[7,]	5	5	73
3438	[8,]	5	5	89
3439	[9,]	6	18	117
3440	[10,]	6	18	118

3441 This “encounter data file” contains 1 row for each unique individual/trap en-
 3442 counter, and 3 variables (columns): `trapid` is an integer that runs from `1:ntraps`,
 3443 individual runs from `1:nind` and sample runs from `1:nperiods`. Often (as the case
 3444 here) “sample” will correspond to daily sample intervals. The variable `trapid` will
 3445 have to correspond to the row of a matrix containing the trap coordinates - a file
 3446 named `traplocs.csv` available in the **R** package `scrbook`.

3447 Note that these data do not represent a completely informative summary of the
 3448 data. For example, if no individuals were captured in a certain trap or during a
 3449 certain period, then this compact data format will have no record. Thus we will
 3450 need to know `ntraps` and `nperiods` when reformatting this SCR data format into a
 3451 2-d encounter frequency matrix or 3-d array. In addition, the encounter data file
 3452 does not provide information about which periods each trap was operated. This
 3453 additional information is also necessary as the trap-specific sample sizes must be
 3454 passed to **BUGS** as data. We provide this information in a 2nd data file - which
 3455 we call the “trap deployment” file (described below).

3456 The “encounter data file” `wcaps.csv` exists in the **R** package `scrbook` as a `.csv`
 3457 file that people can read into **R** and do some basic summary statistics on. For our
 3458 purposes we need to convert these data into the “individual x trap” array of binary
 3459 encounter frequencies, although more general models might require an encounter-
 3460 history formulation of the model which requires a full 3-d array. To obtain our `nind`
 3461 x `ntrap` encounter frequency matrix, we do this the hard way by first converting the
 3462 encounter data file into a 3-d array and then summarize to trap totals. We have a
 3463 handy function `SCR23darray.fn` which takes the compact encounter data file with
 3464 optional arguments `ntraps` and `nperiods`, and converts it to a 3-d array, and then
 3465 we use the **R** function `apply` to summarize over the “sample” period dimension (by
 3466 convention here, this is the 2nd dimension):

```

3467 SCR23darray.fn <- function(caps,ntraps=NULL,nperiods=NULL){
3468   nind<-max(caps[,2])
3469   if(is.null(ntraps)) ntraps<-max(caps[,1])
3470   if(is.null(nperiods)) nperiods<- max(caps[,3])
3471
3472   y<-array(0,c(nind,nperiods,ntraps))
3473   tmp<-cbind(caps[,2],caps[,3],caps[,1])
3474   y[tmp]<-1

```

```

3475 y
3476 }
3477
3478 # for the wolverine data do this:
3479
3480 Y3d <-SCR23darray.fn(wcaps,ntraps=37,nperiods=165)
3481 y <- apply(y3d,c(1,3),sum)

```

If `ntraps` and `nperiods` are not specified then they are assumed to be equal to the maximum value provided in the encounter data file. The 3-d array is necessary to fit certain types of models (e.g., behavioral response) and this is why we sometimes will require this maximally informative 3-d data format.

The other data file that we must have is the “trap deployment” file (henceforth “traps file”) which provides the additional information not contained in the encounter data file. The traps file has `nperiods + 3` columns. The first column is assumed to be a trap identifier, columns 2 and 3 are the easting and northing coordinates (assumed to be in a Euclidean coordinate system), and columns 4 to (`nperiods + 3`) are binary indicators of whether each trap was operational in each time period. The first 5 rows (out of 37) and 10 columns (out of 168) of the traps file for the wolverine data (“`wtraps.csv`” in the **R** package `scrbook` are:

```

3494      Trap Easting Northing 1 2 3 4 5 6 7 <- column names
3495      1   39040    19216 0 0 0 0 0 0 0
3496      2   41324    19772 1 1 1 1 1 1 1
3497      3   44957    12985 0 0 0 0 0 0 0
3498      4   41151    23220 0 0 0 0 0 0 0
3499      5   44240    17198 0 0 0 0 0 0 0

```

This tells us that trap 2 was operated in periods 1-7 but the other traps were not operational during those periods. To extract the relevant information to fit the model in **WinBUGS** we do this:

```

3503 traps<- read.csv("wtraps.csv")
3504 traplocs<- traps[,2:3]
3505 K<- apply(traps[,4:ncol(traps)],1,sum)

```

This results in a matrix `traplocs` which contains the coordinates of each trap and a vector `K` containing the number of days that each trap was operational. We now have all the information required to fit a basic SCR model in **WinBUGS**.

Summarizing these data files for the wolverine study, we see that 21 unique individuals were captured a total of 115 times. Most individuals were captured 1-6 times, with 4, 1, 4, 3, 1, and 2 individuals captured 1-6 times, respectively. In addition, 1 individual was captured each 8 and 14 times and 2 individuals each were captured 10 and 13 times. The number of unique traps that captured a particular individual ranged from 1-6, with 5, 10, 3, 1, 1, and 1 individual captured in each of 1-6 traps, respectively, for a total of 50 unique wolverine-trap encounters. These

numbers might be hard to get your mind around whereas some tabular summary is often more convenient. For that it seems natural to tabulate individuals by trap and total encounter frequencies. The spatial information in SCR data is based on multi-trap captures, and so, it is informative to understand how many unique traps each individual is captured in. At the same, it is useful to understand how many total captures we have of each individual because this is, in an intuitive sense, the effective sample size. So, we reproduce Table 1 from Royle et al. (2011c) which shows the trap and total encounter frequencies:

Table 4.1. Individual frequencies of capture for wolverines captured in camera traps in South-east Alaska in 2008. Rows index unique trap frequencies and columns represent total number of captures (e.g., we captured 4 individuals 1 time, necessarily in only 1 trap; we captured 3 individuals 3 times but in 2 different traps)

	No. of captures									
No. of traps	1	2	3	4	5	6	8	10	13	14
1	4	1	0	0	0	0	0	0	0	0
2	0	0	3	3	0	2	1	2	0	0
3	0	0	1	1	0	0	0	0	0	1
4	0	0	0	0	0	0	0	0	1	0
5	0	0	0	0	1	0	0	0	0	0
6	0	0	0	0	0	0	0	0	1	0

4.7.1 Fitting the model in WinBUGS

For illustrative purposes here we fit the simplest SCR model with the half-normal distance function although we revisit these data with more complex models in later chapters. The model is summarized by the following 3 components:

- (1) $y_{ij} | \mathbf{s}_i \sim \text{Bin}(K, z_i p_{ij})$
- (2) $p_{ij} = p_0 \exp(-\theta \|\mathbf{s}_i - x_j\|^2)$
- (3) $\mathbf{s}_i \sim \text{Unif}(\mathcal{S})$
- (4) $z_i \sim \text{Bern}(\psi)$

We assume customary flat priors on the structural (hyper-) parameters of the model, $\alpha_0 = \text{logit}(p_0)$, θ and ψ . It remains to define the state-space \mathcal{S} . For this, we nested the trap array (Fig. 4.3) in a rectangular state-space extending 20 km beyond the traps in each cardinal direction. We also considered larger state-spaces up to 50 km to evaluate that choice. The buffer of the state space should be larger enough so that individuals beyond the state-space boundary are not likely to be encountered. Thus some knowledge of typical space usage patterns of the species is useful. The coordinate system was scaled so that a unit distance was equal to 10km, producing a rectangular state-space of dimension 9.88x10.5 units ($\text{area} = 10374 \text{km}^2$) within which the trap array was nested. As a general rule, we recommend scaling the state-space so that it is defined near the origin $(x, y) = (0, 0)$. While the scaling of the

coordinate system is theoretically irrelevant, a poorly scaled coordinate system can produce Markov chains that mix poorly. We fitted this model in **WinBUGS** using data augmentation with $M = 300$ potential individuals, using 3 Markov chains each of 12000 total iterations, discarding the first 2000 as burn-in. [R commands for reading in the data and executing the analysis are as follows:

provide those commands here

The output follows (note, we have a parameter “sigma” which we discuss shortly)²:

All based on 3 chains, 12k iters, 2k burn, 30k total

Buffer = 10 km

	mean	sd	2.5%	25%	50%	75%	97.5%	Rhat	n.eff
psi	0.13	0.03	0.08	0.11	0.13	0.15	0.20	1	10000
sigma	0.65	0.06	0.55	0.61	0.64	0.68	0.76	1	1800
p0	0.06	0.01	0.04	0.05	0.06	0.06	0.08	1	20000
N	39.63	6.70	29.00	35.00	39.00	44.00	54.00	1	7100
D	5.92	1.00	4.33	5.22	5.82	6.57	8.06	1	7100
beta	1.23	0.21	0.85	1.08	1.22	1.36	1.66	1	1800
deviance	410.05	12.06	388.70	401.50	409.20	417.80	435.60	1	22000

Buffer = 15 km

n.sims = 30000 iterations saved

	mean	sd	2.5%	25%	50%	75%	97.5%	Rhat	n.eff
psi	0.16	0.04	0.10	0.14	0.16	0.19	0.25	1	3800
sigma	0.64	0.06	0.54	0.60	0.64	0.67	0.76	1	510
p0	0.06	0.01	0.04	0.05	0.06	0.06	0.08	1	17000
N	48.77	9.19	34.00	42.00	48.00	54.00	69.00	1	3300
D	5.78	1.09	4.03	4.98	5.69	6.40	8.18	1	3300
beta	1.25	0.21	0.86	1.10	1.24	1.39	1.70	1	510
deviance	411.00	12.16	389.50	402.40	410.30	418.70	437.00	1	5400

Buffer = 20 km

	mean	sd	2.5%	25%	50%	75%	97.5%	Rhat	n.eff
psi	0.20	0.05	0.12	0.17	0.20	0.23	0.30	1	16000
sigma	0.64	0.06	0.54	0.60	0.63	0.67	0.76	1	1200
p0	0.06	0.01	0.04	0.05	0.06	0.06	0.08	1	1900
N	59.84	11.89	40.00	51.00	59.00	67.00	86.00	1	20000
D	5.77	1.15	3.86	4.92	5.69	6.46	8.29	1	20000
beta	1.26	0.21	0.87	1.11	1.25	1.40	1.71	1	1200
deviance	411.01	12.36	389.10	402.30	410.20	418.80	437.50	1	1500

Buffer = 25 km

	mean	sd	2.5%	25%	50%	75%	97.5%	Rhat	n.eff
psi	0.24	0.05	0.15	0.20	0.24	0.28	0.36	1	3400

²Final as of 1/11/2012. output saved in wolv-buffer-study.txt

3585	sigma	0.64	0.05	0.54	0.60	0.63	0.67	0.75	1	3600
3586	p0	0.06	0.01	0.04	0.05	0.06	0.06	0.08	1	5000
3587	N	72.40	14.72	47.00	62.00	71.00	81.00	105.00	1	2700
3588	D	5.79	1.18	3.76	4.96	5.67	6.47	8.39	1	2700
3589	beta	1.26	0.21	0.88	1.12	1.25	1.40	1.71	1	3600
3590	deviance	411.35	12.23	389.70	402.70	410.55	419.20	437.20	1	30000
3591										
3592	Buffer = 30 km									
3593		mean	sd	2.5%	25%	50%	75%	97.5%	Rhat	n.eff
3594	psi	0.29	0.06	0.18	0.24	0.28	0.33	0.43	1	3100
3595	sigma	0.63	0.05	0.54	0.60	0.63	0.67	0.75	1	5600
3596	p0	0.06	0.01	0.04	0.05	0.06	0.06	0.08	1	11000
3597	N	86.42	17.98	56.00	74.00	85.00	97.00	126.02	1	3900
3598	D	5.82	1.21	3.77	4.98	5.72	6.53	8.49	1	3900
3599	beta	1.27	0.21	0.88	1.12	1.26	1.41	1.71	1	5600
3600	deviance	411.06	12.37	389.20	402.50	410.20	418.90	437.60	1	10000
3601										
3602	Buffer = 35 km									
3603		mean	sd	2.5%	25%	50%	75%	97.5%	Rhat	n.eff
3604	psi	0.34	0.08	0.21	0.29	0.34	0.39	0.50	1	30000
3605	sigma	0.63	0.05	0.54	0.60	0.63	0.67	0.75	1	4500
3606	p0	0.06	0.01	0.04	0.05	0.06	0.06	0.08	1	24000
3607	N	101.79	21.54	65.00	87.00	100.00	115.00	148.00	1	30000
3608	D	5.85	1.24	3.74	5.00	5.75	6.61	8.51	1	30000
3609	beta	1.27	0.21	0.89	1.12	1.25	1.40	1.70	1	4500
3610	deviance	411.10	12.20	389.50	402.40	410.30	418.90	437.20	1	22000
3611										
3612	Buffer = 40 km									
3613		mean	sd	2.5%	25%	50%	75%	97.5%	Rhat	n.eff
3614	psi	0.39	0.09	0.24	0.33	0.39	0.45	0.60	1.01	480
3615	sigma	0.64	0.05	0.54	0.60	0.63	0.67	0.75	1.01	410
3616	p0	0.06	0.01	0.04	0.05	0.06	0.06	0.08	1.00	21000
3617	N	118.05	26.14	75.00	100.00	116.00	133.00	178.00	1.01	450
3618	D	5.87	1.30	3.73	4.97	5.76	6.61	8.84	1.01	450
3619	beta	1.27	0.21	0.89	1.12	1.25	1.40	1.72	1.01	410
3620	deviance	411.37	12.35	389.30	402.60	410.60	419.30	437.50	1.00	9700
3621										
3622	Buffer = 45 km									
3623		mean	sd	2.5%	25%	50%	75%	97.5%	Rhat	n.eff
3624	psi	0.45	0.10	0.28	0.38	0.44	0.51	0.66	1	3600
3625	sigma	0.64	0.05	0.54	0.60	0.63	0.67	0.75	1	10000
3626	p0	0.06	0.01	0.04	0.05	0.06	0.06	0.08	1	8100
3627	N	134.43	28.68	85.00	114.00	132.00	153.00	196.00	1	3300
3628	D	5.83	1.24	3.68	4.94	5.72	6.63	8.50	1	3300
3629	beta	1.26	0.21	0.88	1.11	1.24	1.39	1.69	1	10000
3630	deviance	411.36	12.19	389.60	402.70	410.60	419.10	437.30	1	9400

3631										
3632	Buffer = 50 km									
3633		mean	sd	2.5%	25%	50%	75%	97.5%	Rhat	n.eff
3634	psi	0.51	0.11	0.31	0.43	0.50	0.57	0.74	1	3200
3635	sigma	0.63	0.05	0.54	0.60	0.63	0.67	0.75	1	4700
3636	p0	0.06	0.01	0.04	0.05	0.06	0.06	0.08	1	3300
3637	N	151.61	31.65	96.00	129.00	149.00	172.00	221.00	1	3400
3638	D	5.79	1.21	3.66	4.92	5.69	6.56	8.43	1	3400
3639	beta	1.27	0.21	0.89	1.12	1.25	1.40	1.70	1	4700
3640	deviance	410.81	12.18	389.20	402.30	410.10	418.50	436.70	1	30000
3641										
3642	Buffer = 55 km									
3643		mean	sd	2.5%	25%	50%	75%	97.5%	Rhat	n.eff
3644	psi	0.56	0.12	0.35	0.48	0.55	0.64	0.82	1.01	260
3645	sigma	0.64	0.05	0.54	0.60	0.63	0.67	0.76	1.00	1600
3646	p0	0.06	0.01	0.04	0.05	0.06	0.06	0.08	1.00	30000
3647	N	169.28	35.81	108.00	143.00	166.00	192.00	247.00	1.01	260
3648	D	5.73	1.21	3.66	4.84	5.62	6.50	8.36	1.01	260
3649	beta	1.25	0.21	0.88	1.11	1.24	1.39	1.69	1.00	1600
3650	deviance	411.28	12.38	389.40	402.60	410.50	419.10	437.50	1.00	26000

3651 We see that the estimated density is roughly consistent as we increase the state-
 3652 space buffer from 15 to 50 *km*. We do note that the data augmentation parameter
 3653 ψ (and, correspondingly, N) increase with the size of the state space in accordance
 3654 with the deterministic relationship $N = D * A$. However, density is constant more
 3655 or less as we increase the size of the state-space beyond a certain point. For the
 3656 10 *km* state-space buffer, we see a slight effect on the posterior distribution of D .
 3657 This is not a bug but rather a feature. As we noted above, the state-space is part
 3658 of the model.

3659 One thing we haven't talked about yet is that we can calibrate the desired size
 3660 of the state-space by looking at the estimated home range radius of the species. For
 3661 some models it is possible to convert the parameter θ directly into the home range
 3662 radius (section XXX XYZ). For the half-normal model we interpret the half-normal
 3663 scale parameter σ which is related to θ by $\theta = 1/(2\sigma^2)$ as the radius of a bivariate
 3664 normal movement model.

3665 4.7.2 Conclusion of Analysis

3666 Our point estimate of wolverine density from this study, using the posterior mean
 3667 from the state-space based on the 20 *km* buffer, is approximately 5.77 individu-
 3668 als/1000 *km*² with a 95% posterior interval of [3.86, 8.29]. Density is estimated
 3669 imprecisely which might not be surprising given the low sample size ($n = 21$ indi-
 3670 viduals!). This seems to be a basic feature of carnivore studies although it should
 3671 not (in our view) preclude the study of their populations nor attempts to estimate
 3672 density or vital rates.

It is worth thinking about this model, and these estimates, computed under a rectangular state space roughly centered over the trapping array (Fig. 4.3). Does it make sense to define the state-space to include, for example, ocean? What are the possible consequences of this? What can we do about it? There's no reason at all that the state space has to be a regular polygon – we defined it as such here strictly for convenience and for ease of implementation in **WinBUGS** where it enables us to specify the prior for the activity centers as uniform priors for each coordinate. While it would be possible to define a more realistic state-space using some general polygon, it might take some effort to implement that in the **BUGS** language (see chapter XYZXYZ³ for example of a simple case). Alternatively, we recommend using a discrete representation of the state-space – i.e., approximate \mathcal{S} by a grid of G points. We discuss this in the following section.

4.8 CONSTRUCTING DENSITY MAPS

One of the most useful aspects of SCR models is that they are parameterized in terms of individual locations - i.e., *where* each individual lives – and, thus, we can compute many useful or interesting summaries of the activity centers. For example, we can make a spatial density plot by tallying up the number of activity centers \mathbf{s}_i in boxes of arbitrary size and then producing a nice multi-color spatial plot of those which, we find, increases the acceptance probability of your manuscripts by a substantial amount. We discussed in chapter 2 the idea of estimating derived parameters from MCMC output. In SCR models, there are many derived parameters that are functions of the latent point locations $(\mathbf{s}_1, \dots, \mathbf{s}_N)$. In the present context, the number of individuals living in any well-defined polygon is a derived parameter. Specifically, let $B(x)$ indicate a box centered at x then

$$N(x) = \sum_i I(\mathbf{s}_i \in B(x))$$

is the population size of box $B(x)$, and $D(x) = N(x)/|B(x)|$ is the local density. These are just “derived parameters” (see chapter 2) which are estimated from MCMC output using the appropriate Monte Carlo average. One thing to be careful about, in the context of models in which N is unknown, is that, for each MCMC iteration m , we only tabulate those activity centers which correspond to individuals in the sampled population. i.e., for which the data augmentation variable $z_i = 1$. In this case, we take all of the output for MCMC iterations $m = 1, 2, \dots, \text{niter}$ and compute this summary:

$$N(x, m) = \sum_{z_{i,m}=1} I(s_{i,m} \in B(x))$$

³raccoon example or something?

3704 Thus, $N(x, 1), N(x, 2), \dots$, is the Markov chain for parameter $N(x)$. In what follows
 3705 we will provide a set of **R** commands for doing this calculations and making a basic
 3706 image plot from the MCMC output.

3707 **Step 1:** Define the center points of each box, $B(x)$, or point at which local density
 3708 will be estimated:

```
3709 xg<-seq(Xl,Xu,,50)
3710 yg<-seq(Yl,Yu,,50)
```

3711 **Step 2:** Extract the MCMC histories for the activity centers and the data aug-
 3712 mentation variables. Note that these are each $N \times \text{niter}$ matrices:

```
3713 Sxout<-out$sims.list$s[,1]
3714 Syout<-out$sims.list$s[,2]
3715 z<-out$sims.list$z
```

3716 **Step 3:** We associate each coordinate with the proper box using the **R** command
 3717 `cut()`. Note that we keep only the activity centers for which $z = 1$ (i.e., individuals
 3718 that belong to the population of size N):

```
3719 Sxout<-cut(Sxout[z==1],breaks=xg,include.lowest=TRUE)
3720 Syout<-cut(Syout[z==1],breaks=yg,include.lowest=TRUE)
```

3721 **Step 4:** Use the `table()` command to tally up how many activity centers are in
 3722 each $B(x)$:

```
3723 Dn<-table(Sxout,Syout)
```

3724 **Step 5:** Use the `image()` command to display the resulting matrix.

```
3725 image(xg,yg,Dn/nrow(z),col=terrain.colors(10))
```

3726 Praise the Lord! This map is somewhat useful or at least it looks pretty and will
 3727 facilitate the publication of your papers.

3728 It is worth emphasizing here that density maps will not usually appear uniform
 3729 despite that we have assumed that activity centers are uniformly distributed. This is
 3730 because the observed encounters of individuals provide direct information about the
 3731 location of the $i = 1, 2, \dots, n$ activity centers and thus their “estimated” locations
 3732 will be affected by the observations. In a limiting sense, were we to sample space
 3733 intensely enough, every individual would be captured a number of times and we
 3734 would have considerable information about all N point locations. Consequently,
 3735 the uniform prior would have almost no influence at all on the estimated density
 3736 surface in this limiting situation. Thus, in practice, the influence of the uniformity
 3737 assumption increases as the fraction of the population encountered decreases.

3738 **On the non-intuitiveness of `image()`** – the **R** function `image()` might not
 3739 be very intuitive to some – it plots $M[1,1]$ in the lower left corner. If you want $M[]$
 3740 to be plotted “as you look at it” then $M[1,1]$ should be in the upper left corner.
 3741 We have a function `rot()` which does that. If you do `image(rot(M))` then it puts
 3742 it on the monitor as if it was a map you were looking at. You can always specify
 3743 the x and y – labels explicitly as we did above.

3744 **Spatial dot plots** – Now here is a cruder version based on the “spatial
 3745 dot map” function `spatial.plot`. The useful functions in **R** are `image()` and
 3746 `image.scale()` which is a function we grabbed off the web somewhere. Use of
 3747 this function requires arguments of point locations and the resulting value to be
 3748 displayed. The function is defined and applied as follows:

```
3749 spatial.plot<- function(x,y){
3750   nc<-as.numeric(cut(y,20))
3751   plot(x,pch=" ")
3752   points(x,pch=20,col=topo.colors(20)[nc],cex=2)
3753   image.scale(y,col=topo.colors(20))
3754 }
3755 # To execute the function do this:
3756 spatial.plot(cbind(xg,yg), Dn/nrow(z))
```

3757 4.8.1 Example: Wolverine density map.

3758 We used the posterior output from the wolverine model fitted previous to compute
 3759 a relatively coarse version of a density map, using a 10×10 grid (Fig. 4.4) and
 3760 using a 30×30 grid (Fig. 4.5)⁴. In these figures density is expressed in units of
 3761 individuals per 1000 km^2 , while the area of the pixels is about 1037 km^2 and 115
 3762 km^2 , respectively. That calculation is based on⁵:

```
3763 > total.area<- (Yu-Yl)*(Xu-Xl)*1000
3764 > total.area/(10*10)
3765 [1] 1037.427
3766 > total.area/(30*30)
3767 [1] 115.2697
```

3768 A couple of things are worth noting: First is that as we move away from “where
 3769 the data live” - away from the trap array - we see that the density approaches
 3770 the mean density. This is a property of the estimator as long as the “detection
 3771 function” decreases sufficiently rapidly as a function of distance. Relatedly, it is
 3772 also a property of statistical smoothers such as splines, kernel smoothers, and re-
 3773 gression smoothers - predictions tend toward the global mean as the influence of

⁴Final as of 01/11/2012. Note: Not sure if we should use quantiles for color to make equal area slices. ??? Also should we use the same scale?

⁵This is wrong and needs fixed. Move decimal one place over. i.e., 100 instead of 1000.

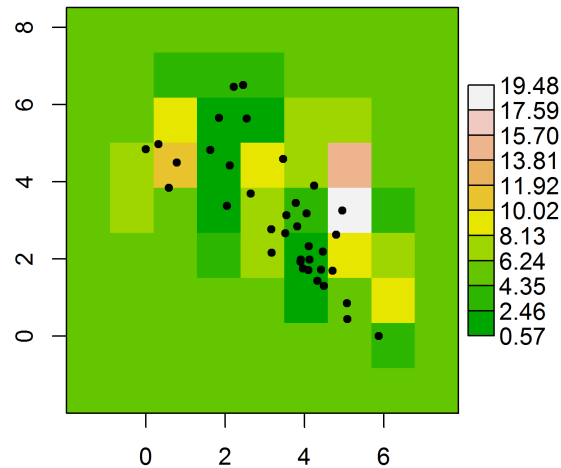


Figure 4.4. Needs a caption

data diminishes. Another way to think of it is that it is a consequence of the prior - which imposes uniformity, and as you get far away from the data, the predictions tend to the prior. The other thing to note about this map is that density is not 0 over water (although the coastline is not shown). This might be perplexing to some who are fairly certain that wolverines do not like water. However, there is nothing about the model that recognizes water from non-water and so the model predicts over water *as if* it were habitat similar to that within which the array is nested. But, all of this is ok as far as estimating density goes and, furthermore, we can compute valid estimates of N over any well-defined region which presumably wouldn't include water if we so choose.

4.9 DISCRETE STATE-SPACE

The SCR model developed previously in this chapter assumes that individual activity centers are distributed uniformly over the prescribed state-space. Clearly this will not always be a reasonable assumption. In chapter ?? we talk about developing models that allow explicitly for non-uniformity of the activity centers by modeling covariate effects on density. A simpler method of affecting the distribution of activity centers, which we address here, is to modify the shape of the state-space explicitly. For example, we might be able to classify the state-space into distinct

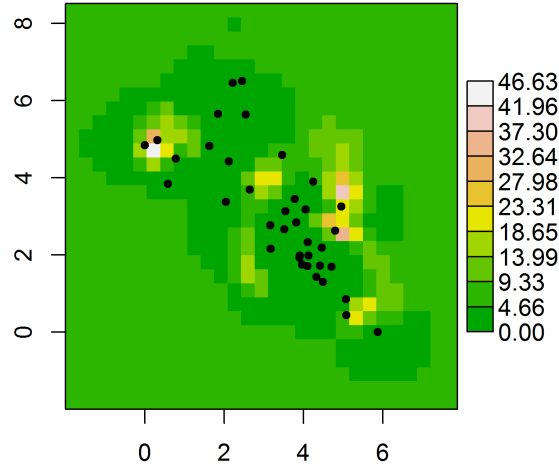


Figure 4.5. Needs a caption

blocks of habitat and non-habitat. In that case we can remove the non-habitat from the state-space and assume uniformity of the activity centers over the remaining portions judged to be suitable habitat. There are two ways to approach this: We can use a regular grid of points to represent the state-space, i.e., by the set of coordinates $\mathbf{s}_1, \dots, \mathbf{s}_G$, and assign a equal probabilities to each possible value, or we can retain the continuous formulation of the state-space but use basic polygon operations to induce constraints on the state-space We focus here on the formulation of our basic SCR model in terms of a discrete state-space but later on (chapter 7 and also Appendix XYZ) we demonstrate the latter approach based on using polygon operations to define an irregular state-space.

Use of a discrete state-space can be computationally expensive in **WinBUGS**. That said, it isn't too difficult to do the MCMC calculations in **R** which we discuss briefly in chapter 7. The **R** package **SPACECAP** (Gopalaswamy et al., 2011) arose from the **R** implementation developed for the application in Royle et al. (2009). As we will see in chapter 6, we must prescribe the state-space by a discrete mesh of points in order to do integrated likelihood and so if we are using a discrete state-space this can be accommodated directly in our code for obtaining MLEs.

While clipping out non-habitat seems like a good idea, its not obvious that we accomplish any biologically reasonable objective by doing so. We might prefer to do it when non-habitat represents a clear-cut restriction on the state-space such as

a reserve boundary or a lake, ocean or river. It makes sense in those situations. Unfortunately, having the capability to do this also causes people to start defining “habitat” vs. “non-habitat” based on their understanding of the system whereas it can’t be known whether the animal being studied has the same understanding. Moreover, differentiating of the landscape by habitat or habitat quality probably affects the geometry and morphology of home ranges much more than the plausible locations of activity centers. That is, a home range centroid could, in actual fact, occur in a walmart parking lot if there is pretty good habitat around walmart, so there is probably no sense to cut out the walmart lot and preclude it as the location for an activity center. It would generally be better to include some definition of habitat quality in the model for the detection probability (see chapter XYZ).

4.9.1 Evaluation of Coarseness of Discrete Approximation

The coarseness of the state-space should not really have much of an effect on estimates if the grain is sufficiently fine relative to typical animal home range sizes. Why is this? We have two analogies that can help us understand this. First is the relationship to Model M_h . As noted in section 4.3.4 above, we can think about SCR models as a type of finite mixture (Norris III and Pollock, 1996; Pledger, 2000) where we are fortunate to be able to obtain direct information about which “group” individuals belong to (group being location of activity center). In the standard finite mixture models we typically find that only 1 or a very small number of groups (e.g., 2 or 3 at the most) can explain really high levels of heterogeneity and are adequate for most data sets of small to moderate sample sizes. We therefore expect a similar effect in SCR models when we discretize the state-space. We can also think about discretizing the state-space as being related to numerical integration where we find (see chapter 6) that we don’t need a very fine grid of support points to evaluate the integral to a reasonable level of accuracy. We demonstrate this here by reanalyzing simulated data using a state-space defined by a different numbers of support points. We provide an R script called `simSCR0discrete.fn` in the **R** package `scrbook`. We note that for this comparison we generated the actual activity centers as a continuous random variable and thus the discrete state-space is, strictly speaking, an approximation to truth. That said, we regard all state-space specifications as approximations to truth because they are all, strictly speaking, models of some unknown truth. Thus the use of any specific discrete state-space is not intrinsically more “wrong” than any specific continuous representation.

We used **JAGS** from the `rjags` function to obtain the results for 6×6 , 9×9 , 12×12 , 15×15 , 20×20 , 25×25 and 30×30 state-space grids. We used 2000 burn, 12000 total iters with 3 chains, therefore a total of 30000 posterior samples. For **WinBUGS** we used 3 chains of 5k total with 1k burnin means 12k total posterior samples. Summary results for these analyses are shown in Table XYZ⁶.

⁶Andy to finish later.

3850 Table XYZ.

3851			Mean	SD	NaiveSE	Time-seriesSE	runtime
3852	6	N	109.7717	15.98959	0.0923160	0.377737	1239
3853	9	N	114.4621	16.72025	0.0965344	0.468659	1267
3854	12	N	115.4309	17.12403	0.098866	0.464830	1576
3855	15	N	114.7699	17.0242	0.0982894	0.425238	1638
3856	20	N	116.0370	17.10686	0.0987665	0.486867	1647
3857	25	N	116.3228	16.98323	0.0980527	0.465527	1661
3858	30	N	116.4252	17.4078	0.100504	0.533735	1806
3859	WinBUGS						
3860			Mean	SD	NaiveSE	Time-seriesSE	runtime
3861	6	N	111.67	16.61			2274
3862	9	N	114.23	17.99			4300
3863	12	N	115.98	17.38			7100
3864	15	N	115.38	17.94			13010

3866 Note: WinBUGS based on fewer samples too!

3868 To get SE and time-series SE do this:

3869 You can use `as.mcmc.list()` to convert to a coda object. Then use `summary`.

3870 The results in terms of the posterior summaries are, as we expect, very similar
 3871 using **WinBUGS**. However, it was interesting to note that **WinBUGS** runtime is
 3872 much worse (note the number of iterations is lower for **WinBUGS** yet the runtime
 3873 is much longer) and, furthermore, it seems to scale with the size of the discrete
 3874 state-space grid. While that was expected, it was unexpected that the runtime of
 3875 **JAGS** would seem relatively consistent as we increase the grid size. We suspect
 3876 that **WinBUGS** is evaluating the full-conditional for each activity center at all
 3877 G possible values whereas it may be that **JAGS** is evaluating the full-conditional
 3878 only at a subset of values or perhaps using previous calculations more effectively.

3879 While this might suggest that one should always use **JAGS** for this analysis, we
 3880 found in our analysis of the wolverine (next section) that **JAGS** could be extremely
 3881 sensitive to starting values, producing MCMC algorithms that sometimes simply
 3882 did not work.

3883 4.9.2 Analysis of the wolverine camera trapping data

3884 We reanalyzed the wolverine data using discrete state-space grids with points spaced
 3885 by 2, 4 and 8 km (depicted in Fig. 4.6). These were constructed from the 40 km
 3886 buffered state-space, and deleting the points over water (see Royle et al., 2011c).
 3887 Our interest in doing this was to evaluate the relative influence of grid resolution
 3888 on estimated density because the coarser grids will be more efficient from a compu-
 3889 tational stand-point and so we would prefer to use them, but perhaps not if there

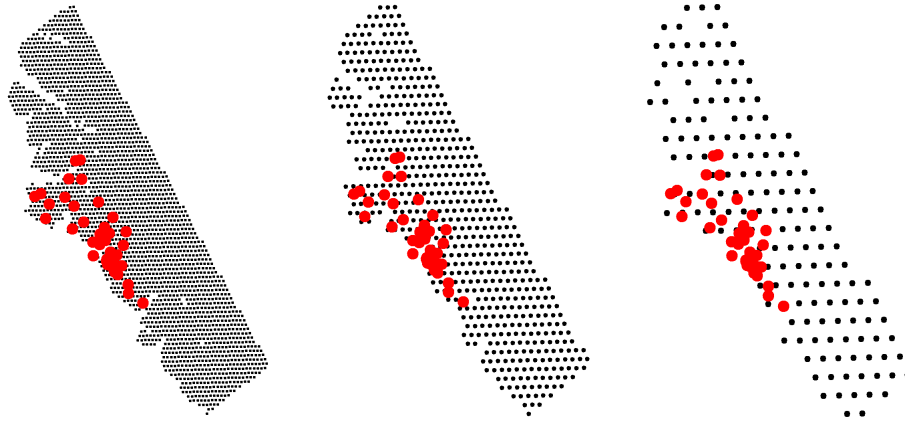


Figure 4.6. 2 km 4 km and 8km wolverine state-space grids extending about 40 km from the vicinity of the trap array.

3890 is a strong influence on estimated density.

3891 **Note:** Results from WinBUGS are given below based on short runs that took
 3892 a long long time. I am rerunning those. I will also show a density map for each
 3893 analysis.

3894 This took about 6 days in WinBUGS. Terrible mixing for the 2km and
 3895 8km. Why is this? We may never know!

```
3896 > print(out.2km,digits=2)
3897 Inference for Bugs model at "modelfile.txt", fit using WinBUGS,
3898 3 chains, each with 11000 iterations (first 1000 discarded)
3899 n.sims = 30000 iterations saved
3900
3901      mean    sd  2.5%  25%  50%  75%  97.5% Rhat n.eff
3902 psi    0.43  0.09  0.27  0.37  0.43  0.49   0.63 1.00   560
3903 sigma  0.62  0.05  0.54  0.59  0.62  0.65   0.73 1.01   160
3904 lam0   0.05  0.01  0.04  0.04  0.05  0.06   0.07 1.01   320
3905 p0     0.05  0.01  0.03  0.04  0.05  0.05   0.06 1.01   320
3906 N      86.56 16.94 57.00 75.00 85.00 97.00 124.00 1.00   510
3907 D       8.78  1.72  5.78  7.60  8.62  9.83  12.57 1.00   510
```

3908
 3909 For each parameter, n.eff is a crude measure of effective sample size,
 3910 and Rhat is the potential scale reduction factor (at convergence, Rhat=1).

```
3911 > print(out.4km,digits=2)
3912 Inference for Bugs model at "modelfile.txt", fit using WinBUGS,
```

```

3913 3 chains, each with 11000 iterations (first 1000 discarded)
3914 n.sims = 30000 iterations saved
3915      mean    sd  2.5%   25%   50%   75%  97.5% Rhat n.eff
3916 psi    0.45  0.09  0.28  0.38  0.44  0.50  0.64    1  1300
3917 sigma 0.61  0.04  0.53  0.58  0.61  0.64  0.71    1  1600
3918 lam0   0.05  0.01  0.04  0.05  0.05  0.06  0.07    1  2500
3919 p0     0.05  0.01  0.03  0.04  0.05  0.05  0.07    1  2500
3920 N      89.25 17.44 59.00 77.00 88.00 100.00 127.00    1  1100
3921 D       9.01  1.76  5.96  7.77  8.88  10.10  12.82    1  1100
3922
3923 For each parameter, n.eff is a crude measure of effective sample size,
3924 and Rhat is the potential scale reduction factor (at convergence, Rhat=1).
3925 > print(out.8km,digits=2)
3926 Inference for Bugs model at "modelfile.txt", fit using WinBUGS,
3927 3 chains, each with 11000 iterations (first 1000 discarded)
3928 n.sims = 30000 iterations saved
3929      mean    sd  2.5%   25%   50%   75%  97.5% Rhat n.eff
3930 psi    0.42  0.09  0.26  0.36  0.41  0.47  0.61  1.00   940
3931 sigma 0.68  0.05  0.59  0.64  0.67  0.71  0.77  1.01   220
3932 lam0   0.05  0.01  0.03  0.04  0.05  0.05  0.06  1.00   560
3933 p0     0.05  0.01  0.03  0.04  0.04  0.05  0.06  1.00   560
3934 N      83.18 16.14 56.00 72.00 82.00 93.00 119.00  1.00   700
3935 D       8.28  1.61  5.57  7.17  8.16  9.26  11.84  1.00   700
3936
3937 For each parameter, n.eff is a crude measure of effective sample size,
3938 and Rhat is the potential scale reduction factor (at convergence, Rhat=1).
3939
3940 We did the analysis in JAGS also. The results are shown below. Note: I am
3941 going to run these again but for longer to finalize the results.
3942
3943 ### 01/10/2012 -- need to rerun these JAGS runs but use more
3944 iterations and check results.
3945
3946 2km
3947 Iterations = 7001:13000
3948 Thinning interval = 1
3949 Number of chains = 3
3950 Sample size per chain = 6000
3951
3952      Mean      SD Naive SE Time-series SE
3953 N      86.28522 16.950626 1.263e-01    0.4878973
3954 lam0    0.04807  0.007512 5.599e-05    0.0002199
3955 p0      0.04581  0.006820 5.083e-05    0.0001996
3956 psi     0.28904  0.062117 4.630e-04    0.0017481
3957 sigma   0.62769  0.043596 3.249e-04    0.0018724

```

3958	4km				
3959		Mean	SD	Naive SE	Time-series SE
3960	N	85.53139	16.998966	1.267e-01	0.5181297
3961	lam0	0.04636	0.007542	5.621e-05	0.0002382
3962	p0	0.04425	0.006867	5.118e-05	0.0002172
3963	psi	0.28650	0.061922	4.615e-04	0.0018276
3964	sigma	0.64281	0.048321	3.602e-04	0.0022911
3965					
3966	8km				
3967		Mean	SD	Naive SE	Time-series SE
3968	N	83.97039	16.508146	1.230e-01	0.4548782
3969	lam0	0.04519	0.006919	5.157e-05	0.0001738
3970	p0	0.04319	0.006319	4.710e-05	0.0001589
3971	psi	0.28146	0.060653	4.521e-04	0.0016555
3972	sigma	0.66956	0.040989	3.055e-04	0.0015070

3973 4.9.3 SCR models as multi-state models

3974 While we invoke a discrete state-space artificially, by gridding the underlying con-
3975 tinuous state-space, sometimes the state-space is more naturally discrete. Consider
3976 a situation in which discrete patches of habitat are searched using some method
3977 and it might be convenient (or occur inadvertently) to associate samples to the
3978 patch level instead of recording observation locations. In this case we might use a
3979 model $\mathbf{s}_i \sim dcat(probs[])$ where $probs[]$ are the probabilities that an individual in-
3980 habits a particular patch. We consider such a case study in chapter XXPoissonXXX
3981 from Mollet et al. (2012) who obtained a population size estimate of a large grouse
3982 species known as the capracaille. Forest patches were searched for scat which was
3983 identified to individual by DNA analysis. Even when space is *not* naturally discrete,
3984 measurements are often made at a fairly coarse grain (e.g., meters or tens of meters
3985 along a stream), or associated with spatial quadrats for scat searches and therefore
3986 the state-space may be effectively discrete in many situations.

3987 This discrete formulation of SCR models suggests that SCR models are related
3988 to ordinary multi-state models (Kery and Schaub, 2011, ch. 9) which are also
3989 parameterized in terms of a discrete state variable which is often defined as a
3990 spatially-indexed state related either to location of capture or breeding location.
3991 While many multi-state models exist in which the state variable is not related to
3992 space, multi-state models have been extremely useful in development models of
3993 movements among geographic states and indeed this type of problem motivated
3994 their early developments by Arnason (1973, 1974) and Hestbeck (1991). We pursue
3995 this connection a little bit more in chapter XXX XYZ.

4.10 SUMMARY AND OUTLOOK

A point we tried to emphasize in this chapter is that the basic SCR model is not much more than an ordinary capture-recapture model for closed populations – it is simply that model but augmented with a set of “individual effects”, \mathbf{s}_i , which relate encounter probability to some sense of individual location. SCR models are therefore a type of individual covariate model (as introduced in chapter 3 – but with imperfect information about the individual covariate. In other words, they are GLMM type models when N is known or, when N is unknown, they are zero-inflated GLMMs (see Royle (2006)). Another class of capture-recapture models that SCR models are closely related to is so-called “Model M_h .” The effect of introducing a spatial location for individuals is that it induces heterogeneity in detection probability, as in Model M_h . However, unlike Model M_h , we obtain some information about the individual effect which is completely latent in Model M_h . If the state-space of the random effect \mathbf{s} is discrete then the SCR model resembles more closely the finite-mixture class of heterogeneity models (Norris III and Pollock, 1996) which parameterizes heterogeneity by assuming that individuals belong to discrete classes or groups (e.g., high, medium, low). In the context of SCR models we obtain some information about the “group membership” in the locations where individuals are captured. Given the direct relationship of SCR models with so many standard classes of models, we find that they are really quite easy to analyze using standard MCMC methods encased in black boxes such as **WinBUGS** or **JAGS** and possibly other packages. They are also easy to analyze using classical likelihood methods, which we address in chapter 6.

Formal consideration of the collection of individual locations $(\mathbf{s}_1, \dots, \mathbf{s}_N)$ in the model is fundamental to all of the models considered in this book. In statistical terminology, we think of the collection of points $\{\mathbf{s}_i\}$ as a realization of a point process and part of the promise, and ongoing challenge, of SCR models is to develop models that reflect interesting biological processes, for example interactions among points or temporal dynamics in point locations. Here we considered the simplest possible point process model – the points are independent and uniformly (“randomly”) distributed over space. Despite the simplicity of this assumption, it should suffice in many applications of SCR models although we do address generalizations of this model in later chapters. Moreover, even though the *prior* distribution on the point locations is uniform, the realized pattern may deviate markedly from uniformity as the observed encounter data provide information to impart deviations from uniformity. Thus, the estimated density map will typically appear distinctly non-uniform. As a general rule, information in the data will govern estimates of individual point locations so even fairly complex patterns of non-independence or non-uniformity will appear in the data. That is, we find in applications of the basic SCR model that this simple *a priori* model can effectively reflect or adapt to complex realizations of the underlying point process. For example, if individuals are highly territorial then the data should indicate this in the form of individuals not

being encountered in the same trap - the resulting posterior distribution of point locations should therefore reflect non-independence. Obviously the complexity of posterior estimates of the point pattern will depend on the quantity of data, both number of individuals and captures per individual. Because the point process is such an integral component of SCR models, the state-space of the point process plays an important role in developing SCR models. As we tried to emphasize in this chapter, the choice of the state-space is part of the model. It can have an influence on parameter estimates and other inferences such as model selection (see chapter 8). We emphasize however that this is not an arbitrary decision like “buffering” because the model induces an explicit interpretation of parameters and statistical effect on estimators.

We showed how to conduct inference about the underlying point process including calculation of density maps from posterior output. We can do other things we normally do with spatial point processes such as compute “K-functions” and test for “complete spatial randomness” (CSR) which we develop in chapter 8. Modifying and applying point process methods to SCR problems seems to us to be a fruitful area of research.

An obvious question that might be floating around in your mind is why should we ever go through all of this trouble when we could just use **MARK** or **CAPTURE** to get an estimate of N and apply 1/2 MMDM methods? The main reason is that these conventional methods are predicated on models that represent explicit misspecifications of both the observation and ecological process - they are wrong! Not just wrong, because of course all models are wrong, but they’re not even *plausible* models! Thus while we might be able to show adequate fit or whatever, we think as a conceptual and philosophical model one should not be using models that are not even plausible data-generating models - even if the plausible ones don’t fit! Perhaps more charitably, these ordinary non-spatial models are models of the wrong system. They do not account for trap identity. They don’t account for spatial organization or “clustering” of individual encounters in space. And, “density” is not a parameter of those models because density has no meaning absent an explicit representation of space. If we do define space explicitly, e.g., as a buffered minimum convex hull, then the normal models (M_0 , M_h , etc..) assume that individual capture-probability is not related to space, no matter how we define the buffer. Conversely, the SCR model is a model for trap-specific encounter data - how individuals are organized in space and interact with traps. SCR models provide a coherent framework for inference about density or population size and also, because of the formality of their derivation, can be extended and generalized to a large variety of different situations, as we demonstrate in subsequent chapters.

In the next few chapters we continue to work with this basic SCR design and model but consider some important extensions of the basic model. For example, we consider extensions to include covariates that vary by individual, trap, or over time (chapter 9), spatial covariates on density (chapter ??), open populations (chapter 11), model assessment and selection (chapter 8) and other topics. We also consider

4080 technical details of Bayesian (chapter 7) and maximum likelihood (chapter 6) esti-
4081 mation so that the interested reader can develop or extend their own methods to
4082 suit their needs.

5

4083

4084

4085

OTHER OBSERVATION MODELS

6

LIKELIHOOD ANALYSIS OF SPATIAL CAPTURE-RECAPTURE MODELS

In this book we mainly focus on Bayesian analysis of spatial capture-recapture models. And, in the previous chapters we learned how to fit some basic spatial capture-recapture models using a Bayesian formulation of the models analyzed in BUGS engines including **WinBUGS** and **JAGS**. Despite our focus on Bayesian analysis, it is instructive to develop the basic conceptual and methodological ideas behind classical analysis based on likelihood methods and frequentist inference. In fact, simple SCR models can be analyzed fairly easily using such methods. This has been the approach taken by Borchers and Efford (2008); Dawson and Efford (2009) and related papers.

This chapter provides some conceptual and technical footing for likelihood-based analysis of spatial capture-recapture models. We recognized earlier (chapt. 4) that SCR models are versions of binomial (or other) GLMs, but with random effects i.e., GLMMs. These models are routinely analyzed by likelihood methods. In particular, likelihood analysis is based on the integrated likelihood in which the random effects are removed by integration from the likelihood. In SCR models, the random effect, \mathbf{s} , i.e., the 2-dimensional coordinate, is a bivariate random effect.

In this chapter, we show that it is straightforward to compute the maximum likelihood estimates (MLE) for SCR models by integrated likelihood. We develop the MLE framework using **R**, and we also provide a basic introduction to an **R** package **secr** (Efford, 2011) which is based on the stand-alone package **DENSITY** (Efford et al., 2004). To set the context we analyze the SCR model here when N is known because, in that case, it is precisely a GLMM and does not pose any difficulty at all. We generalize the model to allow for unknown N using both conventional ideas based on the “joint likelihood” (e.g., Borchers et al., 2002) and also using

4114 a formulation based on data augmentation. We obtain the MLEs for the SCR
 4115 model from the wolverine camera trapping study (Magoun et al., 2011) analyzed
 4116 in previous chapters to compare/contrast the results.

6.1 LIKELIHOOD ANALYSIS

4117 We noted in chapter 4 that, with N known, the basic SCR model is a type of
 4118 binomial regression with a random effect. For such models we can easily obtain
 4119 maximum likelihood estimators of model parameters based on integrated likelihood.
 4120 The integrated likelihood is based on the marginal distribution of the data y in
 4121 which the random effects are removed by integration. Conceptually, our model is a
 4122 specification of the conditional-on- \mathbf{s} model $[y|\mathbf{s}, \theta]$ and we have a “prior distribution”
 4123 for \mathbf{s} , say $[\mathbf{s}]$, and the marginal distribution of the data y is

$$[y|\theta] = \int_{\mathbf{s}} [y|\mathbf{s}, \theta][\mathbf{s}]d\mathbf{s}.$$

4124 When viewed as a function of θ for purposes of estimation, the marginal distribu-
 4125 tion $[y|\theta]$ is often referred to as the *integrated likelihood*.

4126 It is worth analyzing the simplest SCR model with known- N in order to un-
 4127 derstand the underlying mechanics and basic concepts. These are directly relevant
 4128 to the manner in which many capture-recapture models are classically analyzed,
 4129 such as model Mh, and individual covariate models (see chapt. 3 and Royle and
 4130 Dorazio (2008, chapt. 6)). To develop integrated likelihood for SCR models, we
 4131 first identify the conditional likelihood.

4132 The observation model for each encounter observation y_{ij} , specified conditional
 4133 on \mathbf{s}_i , is

$$y_{ij}|\mathbf{s}_i \sim \text{Bin}(K, p_{\theta}(\mathbf{x}_j, \mathbf{s}_i)) \quad (6.1.1)$$

4134 where we have indicated the dependence of p_{ij} on \mathbf{s} and parameters θ explicitly.
 4135 For the random effect we have $\mathbf{s}_i \sim \text{Unif}(S)$. The joint distribution of the data
 4136 for individual i is the product of J such terms (i.e., contributions from each of J
 4137 traps).

$$[\mathbf{y}_i|\mathbf{s}_i, \theta] = \prod_{j=1}^J \text{Bin}(K, p_{\theta}(\mathbf{x}_j, \mathbf{s}_i))$$

4138 We note that this assumes that encounter of individual i in each trap is independent
 4139 of encounter in every other trap, conditional on \mathbf{s}_i , this is the fundamental property
 4140 of SCR0 or “multi-catch” traps.

4141 The so-called marginal likelihood is computed by removing \mathbf{s}_i , by integration
 4142 (hence also *integrated likelihood*), from the conditional-on- \mathbf{s} likelihood and regarding
 4143 the *marginal* distribution of the data as the likelihood. That is, we compute:

$$[y|\theta] = \int_S [\theta|\mathbf{y}_i|\mathbf{s}_i]g(\mathbf{s}_i)d\mathbf{s}_i$$

4144 In most SCR models, $g(\mathbf{s}) = 1/||\mathcal{S}||$ (but see chapt. ??).

4145 The joint likelihood for all N individuals, assuming independence of encounters
4146 among individuals, is the product of N such terms:

$$\mathcal{L}(\theta|\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N) = \prod_{i=1}^N [\mathbf{y}_i|\theta]$$

4147 We emphasize that two independence assumptions are explicit in this development:
4148 independence of trap-specific encounters within individuals and also independence
4149 among individuals. In particular, this would only be valid when individuals are not
4150 physically restrained or removed upon capture, and when traps do not “fill up”.

4151 The key operation for computing the likelihood is solving a 2-dimensional in-
4152 tegration problem. There are some general purpose **R** packages that implement
4153 a number of multi-dimensional integration routines including **adapt** (Genz et al.,
4154 2007) and **R2cuba** (Hahn et al., 2011). In practice, we won’t rely on these extrane-
4155 ous **R** packages (except see chapt. ?? for an application of **Rcuba**) but instead will
4156 use perhaps less efficient methods in which we replace the integral with a summa-
4157 tion over an equal area mesh of points on the state-space \mathcal{S} and explicitly evaluate
4158 the integrand at each point. We invoke the rectangular rule for integration here¹
4159 in which we evaluate the integrand on a regular grid of points of equal area and
4160 compute the average of the integrand over that grid of points. Let $u = 1, 2, \dots, nG$
4161 index a grid of nG points, \mathbf{s}_u , where the area of grid cell u is constant, say A . In
4162 this case, the integrand, i.e., the marginal pmf of \mathbf{y}_i , is approximated by

$$[\mathbf{y}_i|\theta] = \frac{1}{nG} \sum_{u=1}^{nG} [\mathbf{y}_i|\mathbf{s}_u, \theta] \quad (6.1.2)$$

4163 This is a specific case of the general expression that could be used for approxi-
4164 mating the integral for any arbitrary (bivariate or otherwise) distribution $g(\mathbf{s})$. The
4165 general case is

$$[y] = \frac{A}{nG} \sum_u [y|\mathbf{s}_u][\mathbf{s}_u]$$

4166 In the present context it happens that $[\mathbf{s}] = (1/A)$ and thus the grid-cell area
4167 cancels in the above expression to yield eq. 6.1.2, but we commonly apply this in
4168 the context of normal prior distributions, as we did for likelihood analysis of Model
4169 M_h in sec. 3.4). The rectangular rule for integration can be seen as an application
4170 of the Law of Total Probability for a discrete random variable \mathbf{s} , having nG unique
4171 values with equal probabilities $1/nG$.

¹e.g., http://en.wikipedia.org/wiki/Rectangle_method

6.1.1 Implementation (simulated data)

Here we will illustrate how to carryout this integration and optimization based on the integrated likelihood using simulated data (i.e., following that from Chapter 4). Using `simSCRO.fn` we simulate data for 100 individuals and a 25 trap array layed out in a 5×5 grid of unit spacing. The specific encounter model is the half-normal model. The 100 activity centers were simulated on a state-space defined by a 8×8 square within which the trap array was centered (thus the trap array is buffered by 2 units). Therefore, the density of individuals in this system is fixed at $100/64$.

In the following set of R commands we generate the data and then harvest the required data objects:

```

4182 data<-simSCRO.fn(discard0=FALSE,sd=2013)
4183 y<-data$Y
4184 traplocs<-data$traplocs
4185 nind<-nrow(y)
4186 X<-data$traplocs
4187 J<-nrow(X)
4188 K<-data$K
4189 Xl<-data$xlim[1]
4190 Yl<-data$ylim[1]
4191 Xu<-data$xlim[2]
4192 Yu<-data$ylim[2]

```

Now we need to define the integration grid, say **G**, which we do with the following set of **R** commands (here, `delta` is the grid spacing):

```

4195 delta<- .2
4196 xg<-seq(Xl+delta/2,Xu-delta/2,by=delta)
4197 yg<-seq(Yl+delta/2,Yu-delta/2,by=delta)
4198 npix<-length(xg)      # assumes xg and yg same dimension here
4199 area<- (Xu-Xl)*(Yu-Yl)/((npix)*(npix)) # dont need area for anything
4200 G<-cbind(rep(xg,npix),sort(rep(yg,npix)))
4201 nG<-nrow(G)

```

In this case, the integration grid is set up as a grid with spacing $\delta = 0.2$ which produces a 40×40 grid of points for evaluating the integrand if the state-space buffer is set at 2.

We next create an **R** function that defines the likelihood as a function of the data objects `y` and `X` which were created above but, in general, you would read these files into **R**, e.g., from a .csv file. In addition to these data objects, we need to have defined the quantities `G` and `nG` associated with the integration grid. However, instead of worrying about making all of these objects and keeping track of them we just put that code above into the likelihood function and pass δ as an additional (optional) argument and a few other things that we need such as the

boundary of the state-space over which the integration (summation) is being done. Here is one reasonably useful variation of a function for estimation based on the integrated likelihood:

```

4215 intlik1<-function(parm,y=y,delta=.2,X=traplocs,ssbuffer=2){
4216
4217   Xl<-min(X[,1]) - ssbuffer
4218   Xu<-max(X[,1]) + ssbuffer
4219   Yu<-max(X[,2]) + ssbuffer
4220   Yl<-min(X[,2]) - ssbuffer
4221
4222   xg<-seq(Xl+delta/2,Xu-delta/2,,length=npix)
4223   yg<-seq(Yl+delta/2,Yu-delta/2,,length=npix)
4224   npix<-length(xg)
4225
4226   G<-cbind(rep(xg,npix),sort(rep(yg,npix)))
4227   nG<-nrow(G)
4228   D<- e2dist(X,G)
4229
4230   alpha<-parm[1]
4231   theta<-parm[2]
4232   probcap<- plogis(alpha)*exp(-theta*D*D)
4233   Pm<-matrix(NA,nrow=nrow(probcap),ncol=ncol(probcap))
4234           # all zero encounter histories
4235   n0<-sum(apply(y,1,sum)==0)
4236           # encounter histories with at least 1 detection
4237   ymat<-y[apply(y,1,sum)>0,]
4238   ymat<-rbind(ymat,rep(0,ncol(ymat)))
4239   lik.marg<-rep(NA,nrow(ymat))
4240   for(i in 1:nrow(ymat)){
4241     Pm[1:length(Pm)]<- (dbinom(rep(ymat[i,],nG),K,probcap[1:length(Pm)],log=TRUE))
4242     lik.cond<- exp(colSums(Pm))
4243     lik.marg[i]<- sum( lik.cond*(1/nG))
4244   }
4245   nv<-c(rep(1,length(lik.marg)-1),n0)
4246   -1*( sum(nv*log(lik.marg)) )
4247 }

```

The function accepts as input the encounter history matrix, y , the trap locations, X , and the state-space buffer. This allows us to vary the state-space buffer and easily evaluate the sensitivity of the MLE to the size of the state-space. Note that we have a peculiar handling of the encounter history matrix y . In particular, we remove the all-zero encounter histories from the matrix and tack-on a single all-zero encounter history as the last row which then gets weighted by the number of such encounter histories (n_0). This is a bit long-winded and strictly unnecessary when N is known, but we did it this way because the extension to the unknown- N case

is now transparent (as we demonstrate in the following section). The matrix Pm holds the log-likelihood contributions of each encounter frequency for each possible state-space location of the individual. The log contributions are summed up and the result exponentiated on the next line, producing `lik.cond`, the conditional-on-s likelihood (Eq. 6.1.1 above). The marginal likelihood (`lik.marg`) sums up the conditional elements weighted by $\Pr(\mathbf{s})$ (Eq. 6.1.2 above). This is a fairly primitive function which doesn't allow much flexibility in the data structure. For example, it assumes that K , the number of replicates, is constant for each trap. Further, it assumes that the state-space is a square. We generalize this to some extent later in this chapter.

Here is the **R** command for maximizing the likelihood and saving the results into an object called `frog`. The output is a list of the following structure and these specific estimates are produced using the simulated data set:

```
# should take 15-30 seconds

> starting.values <- c(-2, 2)
> frog<-nlm(intlik1,starting.values,y=y,delta=.1,X=traplocs,ssbuffer=2,hessian=TRUE)
> frog

$minimum
[1] 297.1896

$estimate
[1] -2.504824  2.373343

$gradient
[1] -2.069654e-05  1.968754e-05

$hessian
      [,1]      [,2]
[1,] 48.67898 -19.25750
[2,] -19.25750 13.34114

$code
[1] 1

$iterations
[1] 11
```

Details about this output can be found on the help page for `nlm`. We note briefly that `frog$minimum` is the negative log-likelihood value at the MLEs, which are stored in the `frog$estimate` component of the list. The hessian is the observed Fisher information matrix, which can be inverted to obtain the variance-covariance matrix using the commands:

```
4299 > solve(frog$hessian)
```

4300 It is worth drawing attention to the fact that the estimates are different than
 4301 the Bayesian estimates reported previously in chapt. 4. How can that be?! There
 4302 are several reasons for this. First Bayesian inference is based on the posterior
 4303 distribution and it is not generally the case that the MLE should correspond to any
 4304 particular value of the posterior distribution. If the prior distributions in a Bayesian
 4305 analysis are uniform, then the (multivariate) mode of the posterior is the MLE,
 4306 but note that Bayesians almost always report posterior *means* and so there will
 4307 typically be a discrepancy there. Secondly, we have implemented an approximation
 4308 to the integral here and there might be a slight bit of error induced by that. We
 4309 will evaluate that shortly. Third, the Bayesian analysis by MCMC is subject to
 4310 some amount of Monte Carlo error which the analyst should always be aware of
 4311 in practical situations. All of these different explanations are likely responsible for
 4312 some of the discrepancy. Accounting for these, we see general consistency between
 4313 the two estimates.

4314 To compute the integrated likelihood we used a discrete representation of the
 4315 state-space so that the integral could be approximated as a summation over possible
 4316 values of \mathbf{s} with each value being weighted by its probability of occurring, which
 4317 is $1/nG$ under the assumption that \mathbf{s} is uniform on the state-space \mathcal{S} . Recall in
 4318 chapt. 4 we used a discrete state-space in developing a Bayesian analysis of the
 4319 model in order to be able to modify the state-space in a flexible manner. In that
 4320 case, we could use the discretized state-space as the integration grid and just feed
 4321 it into our integrated likelihood routine.

4322 In summary, we note that, for the basic SCR model, integrated likelihood is
 4323 a really easy calculation when N is known. Even for N unknown it is not too
 4324 difficult, and we will do that shortly. However, if you can solve the known- N
 4325 problem then you should be able to do a real analysis, for example by considering
 4326 different values of N and computing the results for each value and then making a
 4327 plot of the log-likelihood or AIC and choosing the value of N that produces the
 4328 best log likelihood or AIC. As a homework problem we suggest that the reader take
 4329 the code given above and try to estimate N without modifying the code by just
 4330 repeatedly calling that code for different values of N and trying to deduce the best
 4331 value. Nevertheless, we will formalize the unknown- N problem shortly.

4332 The software package **DENSITY** (Efford et al., 2004) implements certain types
 4333 of SCR models using integrated likelihood methods. **DENSITY** has been made
 4334 into an **R** package called **secr** (Efford, 2011) and we provide an analysis of some
 4335 data using **secr** shortly along with a discussion of its capabilities.

6.2 MLE WHEN N IS UNKNOWN

4336 Here we build on the previous introduction to integrated likelihood but we consider
 4337 now the case in which N is unknown. We will see that adapting the analysis based

on the N -known model is really straightforward for the more general problem. The main distinction is that we don't observe the all-zero encounter history so we have to make sure we compute the probability for that encounter history which we do by tacking a row of zeros onto the encounter history matrix. In addition, we include the number of such all-zero encounter histories as an unknown parameter of the model. Call that unknown quantity n_0 . In addition, we have to be sure to include a combinatorial term to account for the fact that of the n observed individuals there are $\binom{N}{n}$ ways to realize a sample of size n . The combinatorial term involves the unknown n_0 and thus it must be included in the likelihood.

Operationally then, things proceed much as before: We compute the marginal probability of each observed \mathbf{y}_i , i.e., by removing the latent \mathbf{s}_i by integration. In addition, we compute the marginal probability of the “all-zero” encounter history \mathbf{y}_{n+1} , and make sure to weight it n_0 times. We accomplish this by “padding” the data set with a single encounter history having $y_{n+1,j} = 0$ for all traps $j = 1, 2, \dots, J$. Then we be sure to include the combinatorial term in the likelihood or log-likelihood computation. We demonstrate this shortly.

To analyze a specific case, we'll read in our fake data set (simulated using the parameters given above). To set some things up in our workspace we do this:

```

4356 data<-simSCR0.fn(discard0=TRUE,sd=2013)
4357 y<-data$Y
4358 nind<-nrow(y)
4359 X<-data$traplocs
4360 J<-nrow(X)
4361 K<-data$K
4362 Xl<-data$xlim[1]
4363 Yl<-data$ylim[1]
4364 Xu<-data$xlim[2]
4365 Yu<-data$ylim[2]

```

Recall that these data were generated with $N = 100$, on an 8×8 unit state-space representing the trap locations (\mathbf{X}) buffered by 2 units. As before, the likelihood is defined in the **R** workspace as an **R** function which takes an argument being the unknown parameters of the model and additional arguments as prescribed. In particular, we provide the encounter history matrix \mathbf{y} , the trap locations `traplocs`, the spacing of the integration grid (δ) and the state-space buffer. Here is the new likelihood function:

```

4373 intlik2<-function(parm,y=y,delta=.3,X=traplocs,ssbuffer=2){
4374
4375   Xl<-min(X[,1]) -ssbuffer
4376   Xu<-max(X[,1])+ ssbuffer
4377   Yu<-max(X[,2])+ ssbuffer
4378   Yl<-min(X[,2])- ssbuffer

```

```

4379
4380 #delta<- (Xu-Xl)/npix
4381 xg<-seq(Xl+delta/2,Xu-delta/2,delta)
4382 yg<-seq(Yl+delta/2,Yu-delta/2,delta)
4383 npix.x<-length(xg)
4384 npix.y<-length(yg)
4385 area<- (Xu-Xl)*(Yu-Yl)/((npix.x)*(npix.y))
4386 G<-cbind(rep(xg,npix.y),sort(rep(yg,npix.x)))
4387 nG<-nrow(G)
4388 D<- e2dist(X,G)
4389
4390 alpha<-parm[1]
4391 theta<-parm[2]
4392 n0<-exp(parm[3])
4393 probcap<- plogis(alpha)*exp(-theta*D*D)
4394 Pm<-matrix(NA,nrow=nrow(probcap),ncol=ncol(probcap))
4395 ymat<-rbind(y,rep(0,ncol(y)))
4396
4397 lik.marg<-rep(NA,nrow(ymat))
4398 for(i in 1:nrow(ymat)){
4399   Pm[1:length(Pm)]<- (dbinom(rep(ymat[i,],nG),K,probcap[1:length(Pm)],log=TRUE))
4400   lik.cond<- exp(colSums(Pm))
4401   lik.marg[i]<- sum( lik.cond*(1/nG) )
4402 }
4403 nv<-c(rep(1,length(lik.marg)-1),n0)
4404 part1<- lgamma(nrow(y)+n0+1) - lgamma(n0+1)
4405 part2<- sum(nv*log(lik.marg))
4406 -1*(part1+ part2)
4407 }

```

4408 To execute this function for the data that we created with `simSCRO.fn`, we
 4409 execute the following command (saving the result in our friend `frog`). This re-
 4410 sults in the usual output, including the parameter estimates, the gradient, and the
 4411 numerical Hessian which is useful for obtaining asymptotic standard errors (see
 4412 below):

```

4413 > frog<-nlm(intlik2,c(-2.5,2,log(4)),hessian=TRUE,y=y,X=X,delta=.2,ssbuffer=2)
4414 There were 50 or more warnings (use warnings() to see the first 50)
4415 >
4416 >
4417 > frog
4418 $minimum
4419 [1] 113.5004
4420

```

```

4421 $estimate
4422 [1] -2.538334  2.466515  4.232810
4423
4424 [. Additional output deleted .]

```

While this produces some **R** warnings, these happen to be harmless in this case, and we will see from the **nlm** output that the algorithm performed satisfactory in minimizing the objective function. The estimate of population size for the state-space (using the default state-space buffer) is

```

4429 > nrow(y)+exp(4.2328)
4430 [1] 110.9099

```

Which differs from the data-generating value ($N = 100$) as we might expect. We usually will present an estimate of uncertainty associated with this MLE which we can obtain by inverting the Hessian. Note that $Var(\hat{N}) = n + Var(\hat{n}_0)$. Since we have parameterized the model in terms of $\log(n_0)$ we use a delta approximation to obtain the variance on the scale of n_0 as follows:

```

4436 > (exp(4.2328)^2)*solve(frog$hessian)[3,3]
4437 [1] 260.2033
4438 > sqrt(260)
4439 [1] 16.12452

```

Therefore, the asymptotic “Wald-type” confidence interval for N is $110.91 + / - 1.96 \times 16.125 = (79.305, 142.515)$. To report this in terms of density, we scale appropriately by the area of the prescribed state-space which is 64 units of area (i.e., an 8×8 square).

6.2.1 Exercises

1. Run the analysis with different state-space buffers and comment on the result.
2. Conduct a brief simulation study using this code by simulating 100 data sets and obtain the MLEs for each data set. Do things seem to be working as you expect?
3. Further extensions: It should be straightforward to generalize the integrated likelihood function to accommodate many different situations. For examples, if we want to include more covariates in the model we can just add stuff to the object **probcap**, and add the relevant parameters to the argument that gets passed to the main function. For the simulated data, make up a covariate by generating a Bernoulli covariate (“trap type” perhaps baited or not baited) randomly and try to modify the likelihood to accommodate that.
4. We would probably be interested in devising the integrated likelihood for the full 3-d encounter history array so that we could include temporally varying covariates. This is not difficult but naturally will slow down the execution substantially. The interested reader should try to expand the capabilities of this basic **R** function.

4459 6.2.2 Integrated Likelihood using the model under data augmentation

4460 Note that this likelihood analysis is based on the standard likelihood in which N
 4461 (or n_0) is an explicit parameter. This is usually called the “joint likelihood” or
 4462 “unconditional likelihood”. We could also express the joint likelihood using data
 4463 augmentation, replacing the parameter N with ψ (e.g., see sec. 7.1.6 Royle and
 4464 Dorazio, 2008, for an example). We don’t go into detail here, but we note that the
 4465 likelihood under data augmentation is a zero-inflated binomial mixture precisely
 4466 an occupancy type model (Royle, 2006). Thus, while it is possible to carryout
 4467 likelihood analysis of models under data augmentation, we primarily advocate data
 4468 augmentation for Bayesian analysis.

4469 6.2.3 Extensions

4470 We have only considered basic SCR models with no additional covariates. However,
 4471 in practice, we are interested in other types of covariate effects including “behavioral
 4472 response”, sex-specificity of parameters, and potentially other effects. Some of these
 4473 can be added directly to the likelihood if the covariate is fixed and known for all
 4474 individuals captured or not. An example is a behavioral response, which amounts
 4475 to having a covariate $x_{ik} = 1$ if individual i was captured prior to occasion k and
 4476 $x_{ik} = 0$ otherwise. For uncaptured individuals, $x_{ik} = 0$ for all k . Royle et al.
 4477 (2011c) called this a global behavioral response because the covariate is defined
 4478 for all traps, no matter the trap in which an individual was captured. We could
 4479 also define a *local* behavioral response which occurs at the level of the trap, i.e.,
 4480 $x_{ijk} = 1$ if individual i was captured in trap j prior to occasion k , etc.. Trap-
 4481 specific covariates such as trap type or status, or time-specific covariates such as
 4482 date, are easily accommodated as well. As an example, Kéry et al. (2010) develop
 4483 a model for the European wildcat in which traps are either baited or not (a trap-
 4484 specific covariate with only 2 values), and also encounter probability varies over time
 4485 in the form of a quadratic seasonal response. We consider models with behavioral
 4486 response or fixed covariates in chapter XXXX, although the integrated likelihood
 4487 routines we provided above can be modified directly for such cases, which we leave
 4488 to the interested reader.

4489 Sex-specificity is more difficult to deal with since sex is not known for uncaptured
 4490 individuals (and sometimes not even for all captured individuals). To analyze
 4491 such models, we do Bayesian analysis of the joint likelihood facilitated by the use of
 4492 data augmentation (Gardner et al., 2010; R.E. et al., 2012). For covariates that are
 4493 not fixed and known for all individuals, it is somewhat more challenging to do MLE
 4494 for these based on the joint likelihood as we have developed above. Instead it is
 4495 more conventional to use what is colloquially referred to as the “Huggins-Alho” type
 4496 model which is one of the approaches taken in the software package `secr` (Efford,
 4497 2011, see sec. 6.5). This idea is motivated by thinking about unequal probability
 4498 sampling methods known as Horvitz-Thompson sampling (e.g., see Overton and

4499 Stehman, 1995). We don't use that method anywhere in this book because it rep-
 4500 represents a paradigm shift in the inference framework which is done historically only
 4501 for convenience (i.e., ease of constructing an estimator) and not for philosophical
 4502 or theoretical reasons.

6.3 CLASSICAL MODEL SELECTION AND ASSESSMENT

4503 In most analyses, one is interested in choosing from among various potential mod-
 4504 els. A good thing about classical analysis based on likelihood is we can apply AIC
 4505 methods without difficulty (Burnham and Anderson, 2002). There are two distinct
 4506 contexts for model-selection that we think are relevant to SCR models. First is se-
 4507 lecting among models that represent distinct biological hypotheses (e.g., covariates
 4508 affecting encounter probability or density), and AIC is convenient for assessing the
 4509 relative merits of these different models although if there are only a few models
 4510 it is not objectionable to use hypothesis tests or confidence intervals to determine
 4511 importance of effects. The second context is selecting among various detection func-
 4512 tions. Indeed, when distance is used as a covariate (e.g., distance sampling), AIC
 4513 is usually applied to some large and arbitrary selection of distance functions with
 4514 no biological motivation. As a general rule, we don't recommend this given there
 4515 is hardly ever (if at all) a rational subject-matter based reason motivating specific
 4516 distance functions. As a result, we believe that doing too much model selection
 4517 will invariably lead to over-fitting and thus over-statement of precision. This is the
 4518 main reason that we haven't loaded you down with a basket of models for detection
 4519 probability so far, although we discuss many possibilities in chapter XYZ.

4520 Goodness-of-fit: For many standard capture-recapture models, it is possible to
 4521 identify goodness-of-fit statistics based on the multinomial likelihood and evaluate
 4522 model adequacy using formal statistical tests. Similar strategies can be applied
 4523 to SCR models using expected cell-frequencies based on the marginal distribution
 4524 of the observations. Also, because computing MLEs is somewhat more efficient in
 4525 many cases compared to Bayesian analysis, it is also sometimes easy to use boot-
 4526 strap methods². Bayesian goodness-of-fit is almost always addressed with Bayesian
 4527 p-values or some other posterior predictive check (chapter 2 REF XXX and see
 4528 chapter 8). Royle et al. (2011b) suggested checking model fit by decomposing
 4529 fit into two components: an evaluation of the encounter process model based on
 4530 expected encounter frequencies computed *conditional* on \mathbf{s} , and then independent
 4531 evaluation of the "spatial randomness" hypothesis. We discuss this in chapter 8.

6.4 LIKELIHOOD ANALYSIS OF THE WOLVERINE CAMERA TRAPPING DATA

4532 Here we compute the MLEs for the wolverine data using an expanded version of
 4533 the function we developed in the previous section. To accommodate that each trap

²I could use some references in the context of SCR models for this stuff

might be operational a variable number of nights, we provided an additional argument to the likelihood function (allowing for a vector K), which requires also a modification to the construction of the likelihood. In addition, we had to accommodate that the state-space is a general rectangle, and we included a line in the code to compute the state-space area which we apply below for computing density. The more general function (`intlik3`) is given in the **R** package. It has a general purpose wrapper named `scr`³ which has other capabilities too.

The data were read into our R session and manipulated using the following commands. Note that we use the utility **R** function `SCR23darray.fn` which we defined in chapt. 4.

```

4544 > wcaps<-source("wcaps.R")$value
4545 > wtraps<-source("wtraps.R")$value
4546 > K.wolv<-apply(wtraps[,4:ncol(wtraps)],1,sum)
4547 >
4548 > xx<-SCR23darray.fn(wcaps,ntraps=37,nperiods=165)
4549 > y.wolv<- apply(xx,c(1,3),sum)
4550 > traplocs.wolv<-wtraps[,2:3]
4551 > traplocs.wolv<-traplocs.wolv/10000
4552 >
4553 > frog<-nlm(intlik3,c(-1.5,1.2,log(4)),hessian=TRUE,y=y.wolv,K=K.wolv,X=traplocs.wolv,delt
4554 There were 23 warnings (use warnings() to see them)
4555 > frog
4556
4557 $minimum
4558 [1] 220.4355
4559
4560 $estimate
4561 [1] -2.817570  1.255112  3.599040
4562
4563 $gradient
4564 [1] -6.274309e-06  2.146722e-05 -1.045566e-05
4565
4566 $hessian
4567           [,1]      [,2]      [,3]
4568 [1,]  37.687931 -11.852236  4.688911
4569 [2,] -11.852236  30.846144 -9.199113
4570 [3,]  4.688911 -9.199113 13.050428
4571
4572 $code
4573 [1] 1
4574

```

³Not written yet

```

4575 $iterations
4576 [1] 12
4577
4578 > exp(3.599)*sqrt(solve(frog$hessian)[3,3])
4579 [1] 11.41059
4580 >
4581

```

4582 We obtained the MLEs for a state-space buffer of 2 (standardized units) and
 4583 for integration grid with spacing $\delta = .3, .2, .1, .05$. The MLEs for these 4 cases
 4584 including the relative runtime are given in Table 6.1.

Table 6.1. Run time and MLEs for different integration grid resolutions.

δ	Estimates			
	runtime	α_0	θ	$\log(n_0)$
0.30	9.9	-2.819786	1.258468	3.569731
0.20	32.3	-2.817610	1.254757	3.583690
0.10	115.1	-2.817570	1.255112	3.599040
0.05	407.3	-2.817559	1.255281	3.607158

4585 We see the results change only slightly as the fineness of the integration grid
 4586 increases. Conversely, the runtime on the platform of the day for the 4 cases
 4587 increases rapidly. As we have suggested previously these runtimes could be regarded
 4588 in relative terms, across platforms, for gaging the decrease in speed as the fineness
 4589 of the integration grid increases. The effect of this is that we anticipate some
 4590 numerical error in approximating the integral on a mesh of points, and that error
 4591 increases as the coarseness of the mesh increases.

4592 We studied the effect of the state-space buffer on the MLEs, using a fixed $\delta = .2$
 4593 for all analyses. We used state-space buffers of 1 to 4 units stepped by .5. This
 4594 produced the following results, given here are the state-space buffer, area of the
 4595 state-space, the MLE of N for the prescribed state-space and the corresponding
 4596 MLE of density:

	ssbuff	Ass	Nhat	Dhat
4597 [1,]	1.0	66.98212	37.73338	0.5633352
4598 [2,]	1.5	84.36242	46.21008	0.5477567
4599 [3,]	2.0	103.74272	57.00617	0.5494956
4600 [4,]	2.5	125.12302	69.03616	0.5517463
4601 [5,]	3.0	148.50332	82.17550	0.5533580
4602 [6,]	3.5	173.88362	96.44018	0.5546249
4603 [7,]	4.0	201.26392	111.83524	0.5556646

4605 The estimates of D stabilize rapidly and the incremental difference is within the
 4606 numerical error associated with approximating the integral. The results suggest

that wolverine density is around 0.55 individuals per 100 km^2 (recall that a state-space unit is $10 \times 10 km$). This is about 5.5 individuals per thousand km^2 which compares closely with 5.77 reported in sec. 4.7 based on Bayesian analysis of the model.

6.4.1 Restricted state-space

In sec. 4.9 we used a discrete representation of the state-space in order to have control over its extent and shape, for example so that we could clip out “non-habitat”. Clearly that formulation of the model is relevant to the use of integrated likelihood in the sense that such a representation of the state-space underlies the computation of the integral. Thus, for example, we could easily compute the MLE of parameters under some model with a restricted state-space merely by creating the required state-space at whatever grid resolution is desired, and then feed that state-space into the likelihood evaluation above. The **R** function `scr` which comes with the **R** package for this book accommodates an arbitrary state-space fashioned in this manner, as well as state-spaces created by polygons or GIS shapefiles which we demonstrate here. To reformulate the MLE routine to accommodate shapefiles we make use of the **R** libraries `sp` and `maptools`.

XYZ-lookup-XYZ reported in Royle et al. (2011c) based on a clipped state-space as described in section XYZ (XYZ chapter 4 XYZ).

TO BE COMPLETED

6.4.2 Exercises

1. Compute the 95% confidence interval for wolverine density, somehow. Comment on the practical implication of this level of precision.
2. Compute the AIC of this model and modify `intlik3` to consider alternative link functions (at least one additional) and compare the AIC of the different models and the estimates. Comment.

6.5 PROGRAM DENSITY AND THE R PACKAGE SECR

DENSITY is a software program developed by Efford (2004) for fitting spatial capture-recapture models based mostly on classical maximum likelihood estimation and related inference methods. Efford (2011) has also released an **R** package named `secr`, that contains much of the functionality of **DENSITY** but also incorporates new models and features. Here, we will focus on `secr` as it will continue to be developed, contains more functionality and is based in **R**.

To install and run models in `secr`, you must download the package and load it in **R**.

```

4641 > install.packages(secr)
4642 > library(secr)

```

4643 **secr** allows the user to simulate data and fit a suite of models with various detection
 4644 functions and covariate responses. **secr** uses the standard **R** model specification
 4645 framework using tildes. E.g., the model command is **secr.fit** and is generally
 4646 written as

```

4647 > secr.fit(capturedata, model = list(D~1, g0~1, sigma~1), buffer = 20000)

```

4648 where we have **g0~1** indicating the intercept model. Possible predictors for detection
 4649 probability include both pre-defined variables (e.g., **t** and **b** corresponding to “time”
 4650 and “behavior”), and user-defined covariates of several kinds. For example, to
 4651 include a behavioral response, this would be written as **g0~b**. The discussion of
 4652 covariates is developed more in chapter XX(8)⁴

4653 Before we can fit the models, the data must first be packaged properly for **secr**.
 4654 Two input files are required: trap layout (location and identification information
 4655 for each trap) and capture data (e.g., sampling session, animal identification, trap
 4656 day, and trap location). **secr** requires that you specify the trap type, the two
 4657 most common for camera trapping/hair snares are proximity detectors and count
 4658 detectors. The ‘proximity’ detector type allows, at most, one detection of each
 4659 individual at a particular detector on any occasion. The count detector designation
 4660 allows repeat encounters of each individual at a particular detector on any occasion.
 4661 There are other detector types that one can select such as: ‘polygon’ detector type
 4662 which allows for a trap to be a sampled polygon, e.g., scat surveys, and ‘signal’
 4663 detector which allows for traps that have a strength indicator, e.g., acoustic arrays.
 4664 The detector types single and multi can be confusing as multi seems like it would
 4665 appropriate for something like a camera trap, but instead these two designations
 4666 refer to traps that retain individuals, thus precluding the ability for animals to be
 4667 captured in other traps during the sampling occasion. The single type indicates
 4668 trap that can only catch one animal at a time, while multi indicates traps that may
 4669 catch more than one animal at a time. For a full review of the detector types, one
 4670 should look at the help manual, which can be accessed in **R** after installing the
 4671 **secr** package by using the command:

```

4672 > RShowDoc("secr-manual", package = "secr")

```

4673 As with all of the **scr** models, **secr** fits a detection function relating the proba-
 4674 bility of detection to the distance of a detector from an individual activity center.
 4675 **secr** allows the user to specify one of a variety of detection functions including the
 4676 commonly used half-normal, hazard rate, and exponential. There are 12 different
 4677 functions, but some are only available for simulating data, and one should take
 4678 caution when using different detection functions as the interpretation of the pa-
 4679 rameters, such as sigma, may not be consistent across formulations. The different

⁴Beth: does secr fit a local trap-specific response or just a global behavioral response?

detection functions are defined in the secr manual and can be found by calling the help function for the detection function:

```
> ?detectfn
```

It is useful to note that **secr** requires the buffer distance to be defined in meters and density will be returned as number of animals per hectare. Thus to make comparisons between **secr** and other models, we will often have to convert the density to the same units. Also, note that sigma is returned in units of meters.

5

6.5.1 Analysis using the secr package

To demonstrate the use of the **secr** package, we will show how to do the same analysis on the wolverine study as shown in section 4.6. To use the **secr** package, the data need to be formatted in a similar but slightly different manner than we use in **WinBUGS**⁶. After installing the **secr** package, we first have to read in the trap locations and other related information, such as if the trap is operational during a sampling occasion. The **secr** package reads in the trap data through a command called “**read.traps**”, which requires the detector type as input. The detector type is important because it will determine the likelihood that **secr** will use to fit the model. Here, we have selected proximity since individuals are captured at most once in each trap during each sampling occasion.

```
> traps= read.csv(wtraps.csv)
> #name the first 3 columns to match the secr nomenclature
> colnames(traps)[1:3]<- c("trapID","x", "y")
> trapfile <- read.traps(data = traps, detector = "proximity")
```

After reading in the data, we now need to create the encounter matrix or array. The **secr** package does this through the use of the **make.capthist** command, where we provide the capture histories in raw data format (each line contains the session, identification number, occasion, and trap id for only 1 individual). This is the format that was shown in the data input file “**wcaps**”, and we only need a line or two to organize the data into the order that the **make.capthist** command wants. In creating the capture history, we provide also the trapfile with the trap information, and the format (e.g., here **fmt= ‘‘trapID’’**) so that **secr** knows how to match the encounters to the trap, and finally, we provide the number of occasions:⁷

```
> wolv.dat <- wcaps[,c(2, 3, 1)]
```

⁵One question: SECR only ever reports sigma. What exactly is sigma? It is a scale parameter of a detection function and all detection functions have a scale parameter. But in what sense is this sigma parameter related to home range diameter? Efford doesn't explain this, does he? In some sections in chapter 4 or possibly 6 we get into this issue.

⁶Elaborate on this point – and how is this different than introduced in chapter 4?

⁷Beth: Do you need to update this?

```

4714         #NEED TO UPDATE THIS WHEN I GET THE FILES,
4715         ### I JUST GUESSED AT THE CODE, BUT WOULD LIKE TO TRY IT.
4716 > wolv.dat <- cbind(rep(1, dim(wolv.dat)[1]), wolv.dat)
4717 > colnames(wolv.dat) <- c("Session", "ID", "Occasion", "trapID")
4718
4719 > wolvcapt=make.caphist(wolv.dat, trapfile, fmt = "trapID", noccasions = 165)

4720     Calling the secr.fit command, will run the model. We are using the basic model
4721     (SCR0), so we do not need to make any specifications in the command line except
4722     for the providing the buffer size (in m). To specify different models, you can change
4723     the default  $D^1$ ,  $g0^1$ ,  $\sigma^1$ , which the interested reader can do with very little
4724     difficulty.

4725 > wolv.secr=secr.fit(wolvcapt, model = list(D~1, g0~1, sigma~1), buffer = 20000)
4726
4727 > wolv.secr
4728
4729 secr.fit( caphist = wolvcapt, buffer = 20000, binomN = 1 )
4730 secr 2.0.0, 18:26:39 05 Jul 2011
4731
4732 Detector type      proximity
4733 Detector number    37
4734 Average spacing    4415.693 m
4735 x-range            593498 652294 m
4736 y-range            6296796 6361803 m
4737 N animals          : 21
4738 N detections       : 115
4739 N occasions        : 165
4740 Mask area          : 1037069 ha
4741
4742 Model              : D~1 g0~1 sigma~1
4743 Fixed (real)       : none
4744 Detection fn        : halfnormal
4745 Distribution        : poisson
4746 N parameters       : 3
4747 Log likelihood     : -746.754
4748 AIC                 : 1499.508
4749 AICc                : 1500.920
4750
4751 Beta parameters (coefficients)
4752      beta      SE.beta      lcl      ucl
4753 D      -9.749576 0.23027860 -10.200913 -9.298238
4754 g0      -4.275736 0.15846104  -4.586313 -3.965158
4755 sigma   8.699202 0.07868944   8.544973  8.853430
4756
4757 Variance-covariance matrix of beta parameters
4758      D      g0      sigma

```

```

4759 D      0.053028233  0.000546922 -0.005226926
4760 g0      0.000546922  0.025109900 -0.005885213
4761 sigma -0.005226926 -0.005885213  0.006192027
4762
4763 Fitted (real) parameters evaluated at base levels of covariates
4764      link      estimate SE.estimate      lcl      ucl
4765 D      log 5.831941e-05 1.360973e-05 3.713638e-05 9.158548e-05
4766 g0     logit 1.371121e-02 2.142902e-03 1.008756e-02 1.861207e-02
4767 sigma  log 5.998124e+03 4.727205e+02 5.140849e+03 6.998355e+03

```

Under the fitted (real) parameters, we find D , the density, given in units of individuals/hectare (1 hectare = 10000 m^2). To convert this into individuals/1000 km^2 , we multiply by 100000, thus our density estimate is 5.83 individuals/1000 km^2 . σ is given in units of meters, to convert to kilometers, we divide by 1000, which puts sigma at 5.99 km . Both of these estimates are very similar to those provided in section 4.6 for the buffer size equal to 20 km XXXX How similar? XXXXX.

As an exercise, run this analysis for 30 and 40 km buffers and compare those found in section 4.6 under **WinBUGS**. NOTE: The function `secr.fit` will return a warning when the buffer size appears to be too small. This is useful particularly with the different units being used between programs and packages.

6.6 SUMMARY AND OUTLOOK

In this chapter, we showed that classical analysis of SCR models based on likelihood methods is a relatively simple proposition. Analysis is based on the so-called integrated likelihood in which the individual activity centers (random effects) are removed from the conditional-on-s likelihood by integration. We showed how to construct the integrated likelihood and fit some simple models in the **R** programming language. In addition, likelihood analysis for some broad classes of SCR models can be accomplished in the software package **DENSITY** or the **R** library `secr` which we provided an illustration of here. In later chapters we provide more detailed analyses of SCR data using the `secr` package.

To compute the integrated likelihood we have to precisely describe the state-space of the underlying point process. In practice, this leads to a “buffer” around the trap array. We note that this is not really a “buffer strip” in the sense of Wilson and Anderson (1985) which is a feature of the analysis but it is somewhat more general here. In particular, it establishes the support of the integrand which we generally require to compute any integral. It might be that the integrand itself is finite even if the support is infinity but that may or may not be the case depending on the choice of detection function. As a practical matter then, it will typically be the case that, while estimates of N increase with the size of the buffer, estimates of density stabilize. This is not a feature of the classical methods based on using model M_0 or model M_h and buffering the trap array.

Why or why not use likelihood inference exclusively? For certain specific models, it is probably more computationally efficient to produce MLEs. However, **BUGS** is extremely flexible in terms of describing models, although it sometimes can be quite slow. We can devise models in the **BUGS** language easily that we cannot fit in **secr**. E.g., random individual effects of various types (see next chapter), we can handle missing covariates in complete generality and seamlessly, and impose arbitrary distributions on random variables. Moreover, models can easily be adapted to include auxiliary data types. For example, we might have camera trapping and genetic data and we can describe the models directly in **BUGS** and fit a joint model. For the MLE we have to write a custom new piece of code for each model or hope someone has done it for us. Later we consider open population models which are straightforward to develop in **BUGS** but, so far, there is no available platform for doing MLE although we imagine one could develop this. Another thing that is more conceptual here is non-CSR point processes (see chapter XXXX) and generating predictions of how many individuals have home range centers in any particular polygon. Basic benefits of Bayesian analysis have been discussed elsewhere (XXXXXXXX Chapter 2? BPA book? Link and Barker?) and we believe these are compelling. On the other hand, likelihood analysis makes it easy to do model-selection by AIC and in some cases compute standard errors or carry-out goodness-of-fit evaluations.

In summary, basic SCR models are easy to implement by either likelihood or Bayesian methods but we feel that the typical user will realize much more flexibility in model development using existing platforms for Bayesian analysis. While these tend to be slow (sometimes excruciatingly slow), this will probably not be an impediment in most problems, especially at some near point in the future. Since we spent a lot of time here talking about specific technical details on how to implement likelihood analysis of SCR models, we provided a corresponding treatment in the next chapter on how to devise MCMC algorithms for SCR models. This is a bit more tedious and requires more coding, but is not technically challenging (except perhaps to develop highly efficient algorithms which we don't excel at).

7

MCMC FOR SPATIAL CAPTURE-RECAPTURE

7.1 INTRODUCTION

In this chapter we will dive a little deeper into Markov chain Monte Carlo (MCMC) sampling. We will construct custom MCMC samplers in R, starting with easy-to-code GLMs and GLMMs and moving on to simple SCR models. We will also demonstrate some tricks and simple extensions to the 'spatial null model'. Finally, we will illustrate some alternative ready-to-use software packages for MCMC sampling. We will NOT provide exhaustive background information on the theory and justification of MCMC sampling there are entire books dedicated to that subject and we refer you to Robert and Casella (2004) and Robert and Casella (2010). Rather we aim to provide you with enough background and technical know-how to start building your own MCMC samplers for SCR models in R.

7.1.1 Why build your own MCMC algorithm?

The standard program we have used so far to run MCMC analyses is WinBUGS (Gilks et al., 1994). The wonderful thing about WinBUGS is that it will automatically use the most appropriate and efficient form of MCMC sampling for the model specified by the user.

The fact that we have such a Swiss Army knife type of MCMC machine begs the question: Why would anyone want to build their own MCMC algorithm? For one, there are a limited number of distributions and functions implemented in WinBUGS. While OpenBUGS provides more options, some more complex models may be impossible to build within these programs. A very simple example from spatial capture-recapture that can give you a headache in WinBUGS is when your

state-space is an irregular-shaped polygon, rather than an ideal rectangle that can be characterized by four pairs of coordinates. It is easy to restrict activity centers to any arbitrary polygon in R using an ESRI shapefile (and we will show you an example in a little bit), but you cannot use a shape file in a BUGS model.

Sometimes implementing an MCMC algorithm in R may be faster than in WinBUGS - especially if you want to run simulation studies where you have hundreds or more simulated data sets, several years' worth of data or other large models, this can be a big advantage.

Finally, building your own MCMC algorithm is a great exercise to understand how MCMC sampling works. So while using the BUGS language requires you to understand the structure of your model, building an MCMC algorithm requires you to think about the relationship between your data, priors and posteriors, and how these can be efficiently analyzed and characterized. Not to mention that, if you are an R junkie, it can actually be fun. However, if you don't think you will ever sit down and write your own MCMC sampler, consider skipping this chapter - apart from coding it will not cover anything SCR-related that is not covered by other, more model-oriented chapters as well.

7.2 MCMC AND POSTERIOR DISTRIBUTIONS

As mentioned in Chapter 2, MCMC is a class of simulation methods for drawing (correlated) random numbers from a target distribution, which in Bayesian inference is the posterior distribution. As a reminder, the posterior distribution is a probability distribution for an unknown parameter, say θ , given a set of observed data and its prior probability distribution (the probability distribution we assign to a parameter before we observe data). The great benefit of computing the posterior distribution of θ is that it can be used to make probability statements about θ , such as the probability that θ is equal to some value, or the probability that θ falls within some range of values. As an example, suppose we conducted a Bayesian analysis to estimate detection probability of some species at a study site (p), and we obtained a posterior distribution of $\text{beta}(20,10)$ for the parameter p . The following R commands demonstrate how we make inferences based upon summaries of the posterior distribution. Fig 1 shows the posterior along with the summary statistics.

```
> (post.median <- qbeta(0.5, 20, 10))
[1] 0.6704151
> (post.95ci <- qbeta(c(0.025, 0.975), 20, 10))
[1] 0.4916766 0.8206164
```

Thus, we can state that there is a 95% probability that θ lies between 0.49 and 0.82.

The posterior distribution summarizes all we know about a parameter and thus, is the central object of interest in Bayesian analysis. Unfortunately, in many if not

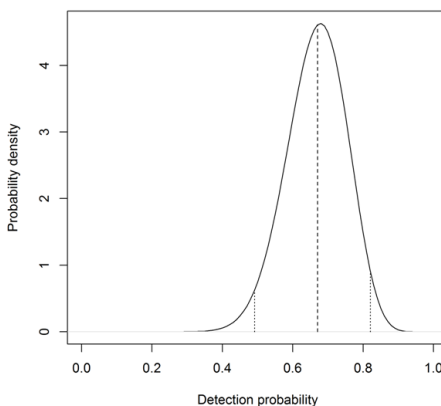


Figure 7.1. Probability density plot of a hypothetical posterior distribution of $\text{beta}(20,10)$; dashed lines indicate mean and upper and lower 95% interval

most practical applications, it is nearly impossible to directly compute the posterior.
Recall Bayes theorem:

$$p(\theta|y) = p(y|\theta) * p(\theta) / p(y), \quad (7.2.1)$$

where θ is the parameter of interest, y is the observed data, $p(\theta|y)$ is the posterior, $p(y|\theta)$ the likelihood of the data conditional on θ , $p(\theta)$ the prior probability of θ , and, finally, $p(y)$ is the marginal probability of the data, which can also be written as

$$p(y) = \int p(y|\theta) * p(\theta) d\theta$$

This marginal probability is a normalizing constant that ensures that the posterior integrates to 1. You read in Chapter 2 that this integral is often hard or impossible to evaluate, unless you are dealing with a really simple model. For example, consider that you have a Normal model, with a set of n observations, y that come from a Normal distribution:

$$y \sim \text{Normal}(\mu, \sigma),$$

where σ is known and our objective is to obtain an estimate of μ using Bayesian statistics. To fully specify the model in a Bayesian framework, we first have to define a prior distribution for μ . Recall from Chapter 2 that for certain data models, certain priors lead to conjugacy i.e. if you choose the right prior for your

parameter, your posterior distribution will be of a known parametric form. The conjugate prior for the mean of a normal model is also a Normal distribution:

$$\mu \sim \text{Normal}(\mu_0, \sigma_0^2)$$

If μ_0 and σ_0^2 are fixed, the posterior for μ has the following form (for the algebraic proof, see XXX):

$$\mu|y \sim \text{Normal}(\mu_n, \sigma_n^2) \quad (7.2.2)$$

where

$$\mu_n = (\text{sig}^2 / \text{sig}^2 + n * \text{sig}0^2) * \text{mu}0 + (n * \text{sig}0^2 / \text{sig}^2 + n * \text{sig}0^2) * y - \text{bar}$$

And

$$\text{sign}^2 = \text{sig}^2 * \text{sig}0^2 / (\text{sig}^2 + n * \text{sig}0^2)$$

We can directly obtain estimates of interest from this Normal posterior distribution, such as the mean μ -hat and its variance; we do not need to apply MCMC, since we can recognize the posterior as a parametric distribution, including the normalizing constant $p(y)$. But generally we will be interested in more complex models with several, say n , parameters. In this case, computing $p(y)$ from Eq. 7.2.1 requires n -dimensional integration, which is can be difficult or impossible. Thus, the posterior distribution is generally only known up to a constant of proportionality:

$$p(\theta|y) \propto p(y|\theta) * p(\theta)$$

The power of MCMC is that it allows us to approximate the posterior using simulation without evaluating the high dimensional integrals and to directly sample from the posterior, even when the posterior distribution is unknown! The price is that MCMC is computationally expensive. Although MCMC first appeared in the scientific literature in 1949 (Metropolis and Ulam, 1949), widespread use did not occur until the 1980s when computational power and speed increased (Gelfand and Smith, 1990). It is safe to say that the advent of practical MCMC methods is the primary reason why Bayesian inference has become so popular during the past three decades. In a nutshell, MCMC lets us generate sequential draws of θ (the parameter(s) of interest) from distributions approximating the unknown posterior over T iterations. The distribution of the draw at t depends on the value drawn at $t-1$; hence, the draws form a Markov chain.¹ As T goes to infinity, the Markov chain converges to the desired distribution. In our case the posterior distribution for $\theta=y$. Thus, once the Markov chain has reached its stationary distribution, the generated samples can be used to characterize the posterior distribution, $p(\theta|y)$, and point estimates of θ , its standard error and confidence bounds, can be obtained directly from this approximation of the posterior. In practice, although we know that

¹In case you are not familiar with Markov chains, for t random samples $\theta(1), \dots, \theta(t)$ from a Markov chain the distribution of $\theta(t)$ depends only on the most recent value, $\theta(t-1)$.

a Markov chain will eventually converge, we can only generate a limited number of samples a process that depending on the model can be quite time consuming. Assessing whether our Markov chain has indeed converged is an important part of MCMC sampling and we will speak about some common diagnostics in Section XX.

7.3 TYPES OF MCMC SAMPLING

There are several MCMC algorithms, the most popular being Gibbs sampling and Metropolis-Hastings sampling. We will be dealing with these two classes in more detail and use them to construct the MCMC algorithms for SCR models. Also, we will briefly review alternative techniques that are applicable in some situations.

7.3.1 Gibbs sampling

Gibbs sampling was named after the physicist J.W. Gibbs by Geman and Geman (1984), who applied the algorithm to a Gibbs distribution². The roots of Gibbs sampling can be traced back to work of Metropolis et al. (1953), and it is actually closely related to Metropolis sampling (see Chapter 11.5 in Gelman et al. (2004), for the link between the two samplers). We will focus on the technical aspects of this algorithm, but if you find yourself hungry for more background, Casella and George (1992) provide a more in-depth introduction to the Gibbs sampler.

In Chapter 2 you already heard about the basic principles of Gibbs sampling³. But as a refresher, let's go back to our simple example from above to understand the motivation and functioning of Gibbs sampling. Recall that for a Normal model with known variance and a Normal prior for μ , the posterior distribution of $\mu|y$ is also Normal. Conversely, with a fixed (known) μ , but unknown variance, the conjugate prior for σ^2 is an Inverse-Gamma distribution with shape and scale parameters a and b :

$$\sigma^2 \sim \text{Inv-Gamma}(a, b),$$

With fixed a and b , the posterior $p(\text{sig}|\mu, y)$ is also an Inverse Gamma distribution, namely:

$$\text{sig}|\mu, y \sim \text{InvGamma}(an, bn), \quad (7.3.1)$$

where $an = n/2 + a$ and $bn = 1/2\sigma(y_i - \mu)^2 + b$. However, what if we know neither μ nor sig , which is probably the more common case? The joint posterior distribution of μ and sig now has the general structure

$$p(\mu, \text{sig}|y) = \frac{p(y|\mu) * p(\mu) * p(\text{sig})}{\int p(y|\mu) * p(\mu) * p(\text{sig}) d\mu d\text{sig}}$$

²a distribution from physics we are not going to worry about, since it has no immediate connection with Gibbs sampling other than giving its name

³maybe we should think out chapter 2 and concentrate that material here?

4967 Or

$$p(\mu, \sigma|y) \propto p(y|\mu) * p(\mu) * p(\sigma)$$

4968 This cannot easily be reduced to a distribution we recognize. However, we can
 4969 condition μ on σ (i.e., we treat σ as fixed) and remove all terms from the
 4970 joint posterior distribution that do not involve μ to construct the full conditional
 4971 distribution,

$$p(\mu|\sigma, y) \propto p(y|\mu) * p(\mu)$$

4972 The full conditional of μ again takes the form of the Normal distribution shown
 4973 in Eq. ??; similarly, $p(\sigma|\mu, y)$ takes the form of the Inverse Gamma distribution
 4974 shown in Eq. Eq. 7.3.1 both distribution we can easily sample from. And this is
 4975 precisely what we do when using Gibbs sampling we break down high-dimensional
 4976 problems into convenient one-dimensional problems by constructing the full con-
 4977 ditional distributions for each model parameter separately; and we sample from
 4978 these full conditionals, which, if we choose conjugate priors, are known parametric
 4979 distributions. Let's put the concept of Gibbs sampling into the MCMC framework
 4980 of generating successive samples, using our simple Normal model with unknown μ
 4981 and σ and conjugate priors as an example. These are the steps you need to build
 4982 a Gibbs sampler:

4983 **Step 0:** Begin with some initial values for θ , $\theta(0)$. In our example, we have to
 4984 specify initial values for μ and σ , for example by drawing a random number from
 4985 some uniform distribution, or by setting them close to what we think they might
 4986 be. (Note: This step is required in any MCMC sampling chains have to start from
 4987 somewhere. We will get back to these technical details a little later.)

4988 **Step 1:** Draw $\theta_1(1)$ from the conditional distribution $p(\theta_1(1)|\theta_2(0), \dots, \theta_d(0))$ Here,
 4989 θ_1 is μ , which we draw from the Normal distribution in Eq. ?? using $\sigma(0)$ as
 4990 value for σ .

4991 Step 2: Draw $\theta_2(1)$ from the conditional distribution $p(\theta_2(1)|\theta_1(1), \theta_3(0), \dots, \theta_d(0))$
 4992 Here, θ_2 is σ , which we draw from the Inverse Gamma distribution of Eq. 7.3.1,
 4993 using $\mu(1)$ as value for μ ...

4994 **Step d:** Draw $\theta_d(1)$ from the conditional distribution $p(\theta_d(1)|\theta_1(1), \dots, \theta_{d-1}(1))$

4995 In our example we have no additional parameters, so we only need step 0 through
 4996 to 2. Repeat Steps 1 to d for K = a large number of samples. In terms of R coding,
 4997 this means we have to write Gibbs updaters for μ and σ and embed them into
 4998 a loop over K iterations. The final code in the form of an R function is shown in
 4999 Panel 1.

5000 Andy will build the panel environment here soon.

5001

5002 Panel 1: R-code for a Gibbs sampler for a Normal model with unknown μ

```

5003 and sig and conjugate (Normal and Inverse Gamma, respectively) priors
5004 for both parameters.
5005
5006 Normal.Gibbs<-function(y=y,mu0=mu0, sig0=sig0, a=a,b=b,niter=niter) {
5007
5008   ybar<-mean(y)
5009   n<-length(y)
5010   mu<-runif(1) #mean initial value
5011   sig<-runif(1) #sd initial value
5012   an<-n/2 + a
5013
5014   out<-matrix(nrow=niter, ncol=2)
5015   colnames(out)<-c('mu', 'sig')
5016
5017   for (i in 1:niter) {
5018
5019     #update mu
5020     mun<- (sig/(sig+n*sig0))*mu0 + (n*sig0/(sig+n* sig0))*ybar
5021     sign <- (sig*sig0)/ (sig+n*sig0)
5022     mu<-rnorm(1,mun, sqrt(sign))
5023
5024     #update sig
5025     bn<- 0.5 * (sum((y-mu)^2)) +b
5026     sig<-1/rgamma(1,shape=an, rate=bn)
5027     out[i,<-c(mu,sqrt(sig))
5028
5029   }
5030   return(out)
5031 }

```

5032 This is it! You can use the code `NormalGibbs.R` in the **R** package `scrbook` to
 5033 simulate some data, $y \sim \text{Normal}(5, 0.5)$ and run your first Gibbs sampler. Your
 5034 output will be a table with two columns, one per parameter, and K rows, one per
 5035 iteration. For this 2-parameter example you can visualize the joint posterior by
 5036 plotting samples of μ against samples of σ (Fig. 2 XXX):

```

5037 plot(out[,1], out[,2])

```

5038 The marginal distribution of each parameter is approximated by just examining the
 5039 samples of this particular parameter you can visualize it by plotting a histogram
 5040 of the samples (Fig. 3 a, b XXX):

```

5041 par(mfrow=c(1,2))
5042 hist(out[,1]); hist (out[,2])

```

Finally, recall an important characteristic of Markov chains, namely, that the chain has to have converged (reached its stationary distribution) for samples to come from the posterior distribution. In practice, that means you have to throw out some of the initial samples called the burn-in. We will talk about this in more when we talk about convergence diagnostics. For now, you can use the `plot(out[,1])` or `plot(out[,2])` command to make a time series plot of the samples of each parameter and visually assess how many of the initial samples you should discard. Figure 3 c and d shows plots for the estimates of mu and sigma from our simulated data set; you see that in this simple example the Markov chain apparently reaches its stationary distribution very quickly the chains look 'grassy' seemingly from the start. It is hard to discern a burn-in phase visually (but we will see examples further on where the burn-in is clearer) and you may just discard the first 500 draws to be sure you only use samples from the posterior distribution. The mean of the remaining samples are your estimates of mu and sig:

```
> summary(mod[501:10000,])
      mu          sig
Min.   : 4.936    Min.   : 0.4569
1st Qu.: 4.984    1st Qu.: 0.4889
Median : 4.994    Median : 0.4961
Mean   : 4.994    Mean   : 0.4964
3rd Qu.: 5.005    3rd Qu.: 0.5037
Max.   : 5.062    Max.   : 0.5356
```

7.3.2 Metropolis-Hastings sampling

Although it is applicable to a wide range of problems, the limitations of Gibbs sampling are immediately obvious what if we do not want to use conjugate priors (or what if we cannot recognize the full conditional distribution as a parametric distribution, or simply do not want to worry about these issues)? The most general solution is to use the Metropolis-Hastings (MH) algorithm, which also goes back to the work by Metropolis et al. (1953). You saw the basics of this algorithm in Chapter 2. In a nutshell, because we do not recognize the posterior $p(\theta|y)$ as a parametric distribution, the MH algorithm generates samples from a known proposal distribution, say $h(\theta)$, that depends on θ at $t-1$. The t^{th} sample is accepted or rejected based on its joint posterior probability density compared to the density of the sample at $t-1$. The original Metropolis algorithm requires $h(\theta)$ to be symmetric so that $h(\theta^t|\theta^{t-1}) = h(\theta^{t-1}|\theta^t)$; but a later development of the algorithm by Hastings (1970) lifted this condition. Using a symmetric proposal distribution makes life a little easier and we are going to limit our coverage of the Metropolis-Hastings sampler to this specific case. Specifically, we are going to use a Normal proposal distribution, which is also referred to as 'random walk Metropolis-Hastings

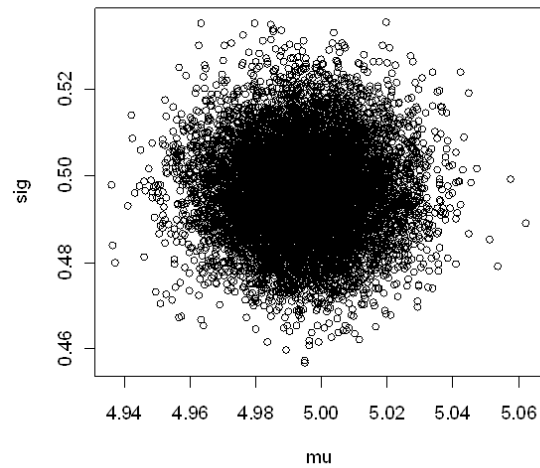


Figure 7.2. Joint posterior distribution of μ and σ from a Normal Model

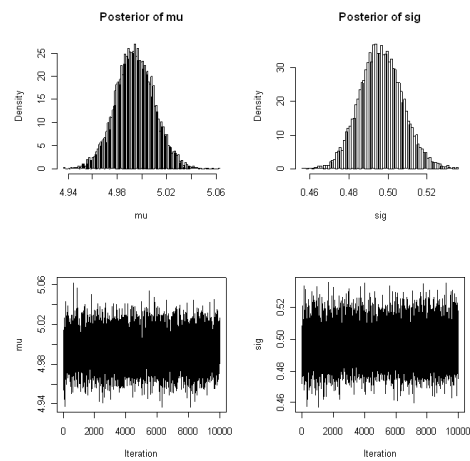


Figure 7.3. Plots of the posterior distributions of μ (a) and σ (b) from a Normal model and time series plots of μ (c) and σ (d).

sampling'. It is worth knowing that there are alternative formulations of the algorithm. For example, in the independent M-H, θ^t does not depend on θ^{t-1} , while the Langevin algorithm (Roberts and Rosenthal, 1998) aims at avoiding the random walk by favoring moves towards regions of higher posterior probability density. The interested reader should look up these algorithms in Robert and Casella (2004) or Robert and Casella (2010).

Building a MH sampler can be broken down into several steps. We are going to demonstrate these steps using a different but still simple and common model the logit-normal or logistic regression model. For simplicity, assume that

$$y \sim \text{Bern}(\exp(\theta)/(1 + \exp(\theta)))$$

and

$$\theta \sim \text{Normal}(\mu_0, \sigma)$$

The following steps are required to set up a random walk MH algorithm:

Step 0: Choose initial values, $\theta(0)$.

Step 1: Generate a proposed value of θ at t from $h(\theta^t - \theta^{t-1})$. We often use a Normal proposal distribution, so we draw θ_1 from $\text{Normal}(\theta^t, \text{sig}^2)$, where sig^2 is the variance of the Normal proposal distribution, a tuning parameter that we have to set.

Step 2: Calculate the ratio of posterior densities for the proposed and the original value for θ :

$$r = p(\theta^t|y)/p(\theta^{t-1}|y)$$

In our example,

$$r = \text{Bern}(y|\theta^t) * \text{Normal}(\theta^t|\mu_0, \sigma_0) / \text{Bernoulli}(y|\theta^{t-1}) * \text{Normal}(\theta^{t-1}|\mu_0, \text{sig}^2)$$

Step 3: Set

```
\begin{eqnarray*}
\theta(t) &= & \theta(t) \text{ with probability } \min(r,1) //
&= & \theta(t-1) \text{ otherwise }
\end{eqnarray*}
```

We can do that by drawing a random number u from a $\text{Unif}(0, 1)$ and accept θ^t if $u < r$. Repeat for $t = 1, 2, \dots$ a large number of samples. The **R** code for this MH sampler is provided in Panel 2 XXXX.

Panel 2: R code to run a Metropolis sampler on a simple Logit-Normal model.

```
Logreg.MH<-function(y=y, mu0=mu0, sig0=sig0, niter=niter) {
```

```

5113 out<-c()
5114
5115 theta<-runif(1, -3,3) #initial value
5116
5117 for (iter in 1:niter){
5118   theta.cand<-rnorm(1, theta, 0.2)
5119
5120   loglike<-sum(dbinom(y, 1, exp(theta)/(1+exp(theta)), log=TRUE))
5121   logprior <- dnorm(theta,mu0 ,sig0, log=TRUE)
5122   loglike.cand<-sum(dbinom(y, 1, exp(theta.cand)/(1+exp(theta.cand)), log=TRUE))
5123   logprior.cand <- dnorm(theta.cand, mu0, sig0, log=TRUE)
5124
5125   if (runif(1)<exp((loglike.cand+logprior.cand)-(loglike+logprior))){
5126     theta<-theta.cand
5127   }
5128   out[iter]<-theta
5129 }
5130
5131 return(out)
5132 }

```

5133 The reason we sum the logs of the likelihood and the prior, rather than multiply-
 5134 ing the original values, is simply computational. The product of small probabilities
 5135 can be numbers very close to 0, which computers do not handle well. Thus we add
 5136 the logarithms, sum, and exponentiate to achieve the desired result. Similarly, in
 5137 case you have forgotten some elementary math, $x/y = \exp(\log(x) - \log(y))$, with
 5138 the latter being favored for computational reasons.

5139 Comparing MH sampling to Gibbs sampling, where all draws from the con-
 5140 ditional distribution are used, in the MH algorithm we discard a portion of the
 5141 candidate values, which inherently makes in less efficient than Gibbs sampling the
 5142 price you pay for its increased generality. In Step 1 of the MH sampler we had to
 5143 choose a variance for the Normal proposal distribution. Choice of the parameters
 5144 that define our candidate distribution is also referred to as 'tuning', and it is im-
 5145 portant since adequate tuning will make your algorithm more efficient, i.e. your
 5146 Markov chain will converge faster. The variance should be chosen so that (a) each
 5147 step of drawing a new proposal value for θ can cover a reasonable distance in the
 5148 parameter space, as otherwise, the random walk moves too slowly; and (b) proposal
 5149 values are not rejected too often, as otherwise the random walk will 'get stuck' at
 5150 specific values for too long. As a rule of thumb, your candidate value should be ac-
 5151 cepted in about 40% of all cases. Acceptance rates of 20-80% are probably ok, but
 5152 anything below or above may well render your algorithm inefficient (this does not
 5153 mean that it will give you wrong results only that you will need more iterations to
 5154 converge to the posterior distribution). In practice, tuning will require some 'trial-
 5155 and-error' and some common sense. Or, one can use an adaptive phase, where the
 5156 tuning parameter is automatically adjusted until it reaches a user-defined accep-

5157 tance rate, at which point the adaptive phase ends and the actual Markov chain
 5158 begins. This is computationally a little more advanced. Link and Barker (2009)
 5159 discuss this in more detail. It is important the samples drawn during the adaptive
 5160 phase are discarded. You can easily check acceptance rates for the parameters you
 5161 monitor (that are part of your output) using the `rejectionRate()` function of the
 5162 package `coda` (we will talk more about this package a little later on). Do not let
 5163 the term 'rejection rate' confuse you; it is simply $1 - \text{acceptance rate}$. There may
 5164 be parameters for example, individual values of a random effect or latent variables
 5165 that you do not want to save, though, and in our next example we will show you a
 5166 way to monitor their acceptance rates with a few extra lines of code.

5167 7.3.3 Metropolis-within-Gibbs

5168 One weakness of the MH sampler is that formulating the joint posterior when
 5169 evaluating whether to accept or reject the candidate values for θ becomes increas-
 5170 ingly complex or inefficient as the number of parameters in a model increases. It
 5171 is probably going to sound like MCMC sampling is too good to be true but in
 5172 these cases you can simply combine MH sampling and Gibbs sampling. You can
 5173 use Gibbs sampling to break down your high-dimensional parameter space into
 5174 easy-to-handle one-dimensional conditional distributions and use MH sampling for
 5175 these conditional distributions. Better yet if you have some conjugacy in your
 5176 model, you can use the more efficient Gibbs sampling for these parameters and
 5177 one-dimensional MH for all the others. You have already seen the basics of how to
 5178 build both types of algorithms, so we can jump straight into an example here and
 5179 build a Metropolis-within-Gibbs algorithm.

7.4 GLMMS POISSON REGRESSION WITH A RANDOM EFFECT

5180 Let's assume a model that gets us closer to the problem we ultimately want to
 5181 deal with a GLMM. Here, we assume we have Poisson counts, y , from i plots
 5182 in j different study sites, and we believe that the counts are influenced by some
 5183 plot-specific covariate, x , but that there is also a random site effect. So our model
 5184 is:

$$y_{ij} \sim \text{Poisson}(l_{amij})$$

5185

$$l_{amij} = \exp(a_j + b * x_i)$$

5186 Let's use Normal priors on a and b ,

$$a_j \sim \text{Normal}(m_{ua}, s_{iga})$$

5187 and

$$b \sim \text{Normal}(m_{ub}, s_{igb})$$

5188 . ⁴ Since we want to estimate the random effect in this model, we do not specify
 5189 μ_a and σ_a , but instead, estimate them as well, so we have to specify hyperpriors
 5190 for these parameters:

$$\begin{aligned}\mu_a &\sim \text{Normal}(\mu_0, \text{sig}_0) \\ \sigma_a &\sim \text{InvGamma}(a_0, b_0)\end{aligned}$$

5191 With the model fully specified, we can compile the full conditionals, breaking
 5192 the multi-dimensional parameter space into one-dimensional components:

```
5193 \begin{eqnarray*}
5194 p(a1|a2,a3,aj,b,y) &\& \propto & p(yi1|a1,b) * p(a1|mua, siga) \\
5195 &\& \propto & \text{Poisson}(yi1| \exp(a1 + b*x[j=1])) * \text{Normal}(a1|mua, siga) \\
5196 \end{eqnarray*}
5197 \begin{eqnarray*}
5198 p(a2|a1,a3,aj,b,y) &\& \propto & p(yi2|a2,b) * p(a2|mua, siga) \\
5199 &\& \propto & \text{Poisson}(yi2|\exp(a2 + b*x[j=1])) * \text{Normal}(a2|mua, siga) \\
5200 \end{eqnarray*}
5201 \text{and so on for all elements of a.}
5202 \begin{eqnarray*}
5203 p(b|a,y) &\& \propto & p(y|a,b) * p(b) \\
5204 &\& \propto & \text{Poisson}(y|\exp(a + b*x)) * \text{Normal}(b|mub, sigb) \\
5205 \end{eqnarray*}
```

5206 Finally, we need to update the hyperparameters for a:

$$p(mua|a) \propto p(a|mua, siga) * p(mua)$$

5207

$$p(siga|a) \propto p(a|mua, siga) * p(siga)$$

5208 Since we assumed a to come from a Normal distribution, the choice of priors for
 5209 mua Normal and siga Inverse Gamma leads to the same conjugacy we observed
 5210 in our initial Normal model, so that both hyperparameters can be updated using
 5211 Gibbs sampling.

5212 Now let' build the updating steps for these full conditionals. Again, for the
 5213 MH steps that update a and b we use Normal proposal distributions with standard
 5214 deviations sigha and sighb.

5215 First, we set the initial values a(0) and b(0). Then, starting with a1, we draw
 5216 a1(1) from Normal (a1(0), sigha), calculate the conditional posterior density of
 5217 a1(0) and a1(1) and compare their ratios,

$$r = \text{Poisson}(y(j=1)|\exp(a1(1)+b*x)) * \text{Normal}(a1(1)|mua, siga) / \text{Poisson}(y(j=1)|\exp(a1(0)+b*x)) * \text{Normal}(a1(0)|mua, siga)$$

5218 and accept a1(1) with probability min(r,1). We repeat this for all a's.

⁴Why is b a hyperparameter?

5219 For b, we draw $b(1)$ from $\text{Normal}(b(0), \text{sigbh})$, compare the posterior densities
 5220 of $b(0)$ and $b(1)$,

$$r = \text{Poisson}(y|\exp(a+b(1)*x)) * \text{Normal}(b(1)|\text{mub}, \text{sigb}) / \text{Poisson}(y|\exp(a+b(0)*x)) * \text{Normal}(b(0)|\text{mub}, \text{sigb}),$$

5221 and accept $b(1)$ with probability $\min(r, 1)$.

5222 For mua and sig , we sample directly from the full conditional distributions (Eq
 5223 XX and Eq XX):

$$\text{mua}(1) \sim \text{Normal}(\text{mun}, \text{sign})$$

5224 where $\text{mun} = (\text{sig}(0)/\text{sig}(0) + n_a * \text{sig}) * \text{mu0} + (n_a * \text{sig0}/\text{sig}(0) + n_a * \text{sig0}) * \text{abar}(1)$ and $\text{sign} = \text{sig}(0) * \text{sig0}/(\text{sig}(0) + n * \text{sig0})$. Here, abar is the
 5225 current mean of the vector a , which we updated before, and n_a is the length of
 5226 a . For sig we use $\text{sig}(1) \sim \text{InvGamma}(an, bn)$, where $an = n_a/2 + a0$, and
 5227 $bn = 1/2 \sum (\text{a}(1) - \text{mua}(1))^2 + b0$.
 5228

5229 We repeat these steps over K iterations of the MCMC algorithm. In this example
 5230 we may not want to save each value for a , but are only interested in their mean and
 5231 standard deviation. Since these two parameters will change as soon as the value for
 5232 one element in a changes, their acceptance rates will always be close to 1 and are
 5233 not representative of how well your algorithm performs. To monitor the acceptance
 5234 rates of parameters you do not want to save, you simply need to add a few lines
 5235 of code into your updater to see how often the individual parameters are accepted.
 5236 The full code for the MCMC algorithm of our Poisson GLMM in Panel 3 shows
 5237 one way how to monitor acceptance of individual a 's.

5238 Panel 3: R code for the Metropolis-within-Gibbs sampler for
 5239 a Poisson regression with random intercepts.

```
5240
5241 Pois.reg<-function(y=y,site=site,mu0=mu0,sig0=sig0,a0=a0,b0=b0,
5242                   mub=mub, sigb=sigb, niter=niter){
5243
5244   lev<-length(unique(site))      #number of sites
5245   a<-runif(lev,-5,5) #initial values a
5246   b<-runif(1,0,5) #initial value b
5247   mua<-mean(a)
5248   siga<-sd(a)
5249
5250   out<-matrix(nrow=niter, ncol=3)
5251   colnames(out)<-c('mua','siga','b')
5252
5253   for (iter in 1:niter) {
5254
5255     #update a
5256     aUps<-0 #initiate counter for acceptance rate of a
5257     for (j in 1:lev) { #loop over sites
5258       a.cand<-rnorm(1, a[j], 0.1) #update intercepts a one at a time
```

```

5259 loglike<- sum(dpois (y[site==j], exp(a[j] + b*x[site==j]), log=TRUE))
5260 logprior<- dnorm(a[j], mua,siga, log=TRUE)
5261 loglike.cand<- sum(dpois (y[site==j], exp(a.cand + b *x[site==j]), log=TRUE))
5262 logprior.cand<- dnorm(a.cand, mua,siga, log=TRUE)
5263 if (runif(1)< exp((loglike.cand+logprior.cand) (loglike+logprior))) {
5264   a[j]<-a.cand
5265   aUps<-aUps+1
5266 }
5267 }
5268
5269 if(iter %% 100 == 0) { #this lets you check the acceptance rate of a at every 100th iteration
5270   cat("   Acceptance rates\n")
5271   cat("     a =", aUps/lev, "\n")
5272 }
5273
5274 #update b
5275 b.cand<-rnorm(1, b, 0.1)
5276 avec<-rep(a, times=c(rep(10,10)))
5277 loglike<- sum(dpois (y, exp(avec + b*x), log=TRUE))
5278 logprior<- dnorm(b, mub,sigb, log=TRUE)
5279 loglike.cand<- sum(dpois (y, exp(avec + b.cand *x), log=TRUE))
5280 logprior.cand<- dunif(b.cand, mub,sigb, log=TRUE)
5281 if (runif(1)< exp((loglike.cand+logprior.cand) (loglike+logprior) )) {
5282   b<-b.cand
5283 }
5284
5285 #update mua using Gibbs sampling
5286 abar<-mean(a)
5287 mun<- (siga/(siga+lev*sig0))*mu0 + (lev*sig0/(siga+lev* sig0))*abar
5288 sign <- (siga*sig0)/ (siga+lev*sig0)
5289 mua<-rnorm(1,mun, sqrt(sign))
5290
5291 #update siga using Gibbs sampling
5292 a0n<-lev/2 + a0
5293 b0n<- 0.5 * (sum((a-mua)^2)) +b0
5294 siga<-1/rgamma(1,shape=a0n, rate=b0n)
5295
5296 out[iter,]<-c(mua, sqrt(siga), b)
5297
5298 }
5299
5300 return(out)
5301 }

```

7.4.1 Rejection sampling and slice sampling

While MH and Gibbs sampling are probably the most widely applied algorithms for posterior approximation, there are other options that work under certain circumstances and may be more efficient when applicable. WinBUGS applies these algorithms and we want you to be aware that there is more out there to approximate posterior distributions than Gibbs and MH. One alternative algorithm is rejection sampling. Rejection sampling is not an MCMC method, since each draw is independent of the others. The method can be used when the posterior $p(\theta|y)$ is not a known parametric distribution but can be expressed in closed form. Then, we can use a so-called envelope function, say, $g(\theta)$, that we can easily sample from, with the restriction that $p(\theta|y) < M * g(\theta)$. We then sample a candidate value for θ from $g(\theta)$, calculate $r = p(\theta|y) / M * g(\theta)$ and keep the sample with the probability r . M is a constant that has to be picked so that $r \in [0,1]$, for example by evaluating both $p(\theta|y)$ and $g(\theta)$ at n points and looking at their ratios. Rejection sampling only works well if $g(\theta)$ is similar to $p(\theta|y)$, and packages like WinBUGS use adaptive rejection sampling (Gilks and Wild, 1992), where a complex algorithm is used to fit an adequate and efficient $g(\theta)$ based on the first few draws. Though efficient in some situations, rejection sampling does not work well with high-dimensional problems, since it becomes increasingly hard to define a reasonable envelope function. For an example of rejection sampling in the context of SCR models, see Chapter 9. Another alternative is slice sampling (Neal, 2003). In slice sampling, we sample uniformly from the area under the plot of $p(\theta|y)$. Considering a single univariate θ . Let's define an auxiliary variable, $U \sim \text{Uniform}(0, p(\theta|y))$. Then, θ can be sampled from the vertical slice of $p(\theta|y)$ at U (Figure 4):

$$\theta|U \sim \text{Unif}(B),$$

where $B = \{\theta : U < p(\theta|y)\}$

5

Slice sampling can be applied in many situations; however, implementing an efficient slice sampling procedure can be complicated. We refer the interested reader to chapter 7 of Robert and Casella (2010) for a simple example. Both rejection sampling and slice sampling can be applied on one-dimensional conditional distributions within a Gibbs sampling setup.

7.5 MCMC FOR CLOSED CAPTURE-RECAPTURE MODEL MH

6

⁵there are supposed to be equations in the caption of figure 4 but it kept causing errors

⁶Andy could move material from chapter 3 to here.

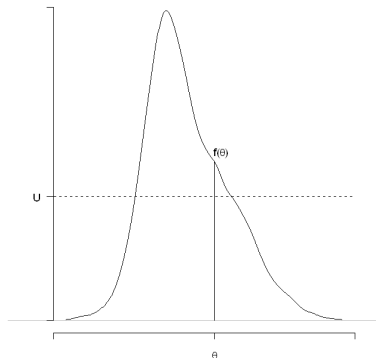


Figure 7.4. Slice sampling. For...

7.6 MCMC ALGORITHM FOR THE BASIC SPATIAL CAPTURE-RECAPTURE MODEL

By now you have seen how to build MCMC algorithms for some basic generalized linear models. Now, we'll walk you through the steps of building your own MCMC sampler for the basic SCR model (i.e. without any individual, site or time specific covariates) with both a Poisson and a binomial encounter process. As usual, we will have to go through two general steps before we write the MCMC algorithm:

- (1) Identify your model with all its components (including priors)
- (2) Recognize and express the full conditional distributions for all parameters

It is worthwhile to go through all of step 1 for an SCR model, but you have probably seen enough of step 2 in our previous examples to get the essence of how to express a full conditional distribution. Therefore, we will exemplify step 2 for some parameters and tie these examples directly to the respective R code.

Step 1 Identify your model

Recall the components of the basic SCR model with a Poisson encounter process from Chapter 3: We assume that individuals i , or rather, their activity centers s_i , are uniformly distributed across our state space S ,

$$s_i \sim U(S)$$

and that the number of times individual i encounters trap j , y_{ij} , is a random Poisson variable with mean λ_{mij} ,

$$y_{ij} \sim \text{Poisson}(\lambda_{mij})$$

The tie between individual location, movement and trap encounter rates is made by the assumption that λ_{mij} , is a decreasing function of the distance between s_i

5354 and j , D_{ij} , of the half-normal form

$$Lam_{ij} = lam0 * \exp(-D_{ij}^2/2 * sig^2),$$

5355 where $lam0$ is the baseline trap encounter rate at $D_{ij} = 0$ and sig controls the
5356 shape of the half-normal function.

5357 In order to estimate the number of s_i in S , N , we use data augmentation (sect.
5358 3.XYZ) and create M - n all-0 encounter histories, where n is the number of individ-
5359 uals we observed and M is an arbitrary number that is larger than N . We estimate
5360 N by summing over the auxiliary data augmentation variables, z_i , which is 1 if the
5361 individual is part of the population and 0 if not, and assume that z_i is a random
5362 Bernoulli variable,

$$z_i \sim \text{Bern}(\psi)$$

5363 To link the two model components, we modify our trap encounter model to

$$Lam_{ij} = lam0 * \exp(-D_{ij}^2/2 * sig^2) * z_i.$$

5364 The model has the following structural parameters, for which we need to specify
5365 priors ψ the Uniform (0,1) is required as part of the data augmentation procedure
5366 and in general is a natural choice of an uninformative prior for a probability; note
5367 that this is equivalent to a Beta(1,1) prior, which will come in handy later. s_i since
5368 s_i is a pair of coordinates it is two-dimensional and we use a uniform prior limited
5369 by the extent of our state-space over both dimensions. σ we can conceive several
5370 priors for sigma but let's assume an improper prior one that is Uniform over (-Inf,
5371 Inf). We will see why this is convenient when we construct the full conditionals for
5372 sigma. λ_0 analogous, we will use a Uniform (-Inf, Inf) improper prior for sigma.
5373 The parameter that is the objective of our modeling, N , is a derived parameter that
5374 we can simply obtain by summing all z 's:

$$N = \text{sum}(z)$$

5375 **Step 2 - Construct the full conditionals** Having completed step 1, let's
5376 look at the full conditional distributions for some of these parameters. We find
5377 that with improper priors, full conditionals are proportional only to the likelihood
5378 of the observations; for example, take the movement parameter sigma:

$$Sig|s, lam0, z, ypropto[y|s, lam0, z, sig] * [sig]$$

5379 Since the improper prior implies that $[sig] \propto 1$, we can reduce this further to

$$Sig|s, lam0, z, ypropto[y|s, lam0, z, sig]$$

5380 The R code to update sigma is shown in Panel 4. ⁷

⁷ Somewhere in chapter 2 i added a comment about rejecting parameters outside of the parameter space as being an ok thing to do. Richard said he read something in Robert and Casellas book on that. Hopefully he can remember where and we can cite it back in Ch 2 and again here. It could be mentioned in a sentence or two up in the MCMC section.

Panel 4: R code to update sigma within an MCMC algorithm for an SCR model when using an improper prior

```

5381 sig.cand <- rnorm(1, sigma, 0.1) #draw candidate value
5382 if(sig.cand>0){ #automatically reject sig.cand that are <0
5383     lam.cand <- lam0*exp(-(D*D)/(2*sig.cand*sig.cand))
5384     ll<- sum(dpois(y, lam*z, log=TRUE))
5385     llcand <- sum(dpois(y, lam.cand*z, log=TRUE))
5386     if(runif(1) < exp( llcand - ll) ){
5387         ll<-llcand
5388         lam<-lam.cand
5389         sigma<-sig.cand
5390     }
5391 }
5392 }
5393 }
5394 }
5395 }
5396 }

```

These steps are analogous for lam0 and si and we will use MH steps for all of these parameters. Similar to the random intercepts in our Poisson GLMM, we update each si individually. Note that to be fully correct, the full conditional for si contains both the likelihood and prior component, since we did not specify an improper, but a Uniform prior on si. However, with a Uniform distribution the probability density of any value is 1/(upper limit - lower limit) = constant. Thus, the prior components are identical for both the current and the candidate value and can be ignored (formally, when you calculate the ratio of posterior densities, r, the identical prior component appears both in the numerator and denominator, so that they cancel each other out).

We still have to update zi. The full conditional for zi is

$$z_i|y, \sigma, \lambda_0, \text{sprpto}[y|z, \sigma, \lambda_0, s] * [z_i]$$

and since $z_i \sim \text{Bernoulli}(\psi_i)$, the term has to be taken into account when updating zi. The R code for updating zi is shown in Panel 5.

Panel 5: R code to update z

```

5410 zUps <- 0 #set counter to monitor acceptance rate
5411 for(i in 1:M) {
5412     if(seen[i]) #no need to update seen individuals, since their z =1
5413         next
5414     zcand <- ifelse(z[i]==0, 1, 0)
5415     llz <- sum(dpois(y[i,], lam[i,]*z[i], log=TRUE))
5416     llcand <- sum(dpois(y[i,], lam[i,]*zcand, log=TRUE))
5417     prior <- dbinom(z[i], 1, psi, log=TRUE)
5418     prior.cand <- dbinom(zcand, 1, psi, log=TRUE)
5419 }
5420 }
5421 }

```

```

5422         if(runif(1) < exp( (llcand+prior.cand) - (llz+prior) )) {
5423             z[i] <- zcand
5424             zUps <- zUps+1
5425         }
5426     }

```

5427 ψ itself is a hyperparameter of the model, with an uninformative prior distribu-
5428 tion of Unif(0,1) or Beta(1,1), so that

$$Psi|z \propto [z|psi] * Beta(1, 1)$$

5429 The Beta distribution is the conjugate prior to the Binomial and Bernoulli distribu-
5430 tions (remember that $z \sim Bernoulli(psi)$). The general form of a full conditional
5431 of a Beta-Binomial model with $y_i \sim Bernoulli(p)$ and $p \sim Beta(a, b)$ is

$$p(p|y) \propto Beta(a + sum(yi), b + n - sum(yi))$$

5432 In our case, this means we update psi as follows:

```

5433 si<-rbeta(1, 1+sum(z), 1 + M-sum(z))

```

5434 These are all the building blocks you need to write the MCMC algorithm for
5435 the spatial null model with a Poisson encounter process. You can find the full R
5436 code (SCR0pois.R) in the online supplementary material.

5437 7.6.1 SCR model with binomial encounter process

5438 The equivalent SCR model with a binomial encounter process is very similar. Here,
5439 each individual i can only be detected once at any given trap j during a sampling
5440 occasion k . Thus

$$y_{ij} \sim Binomial(p_{ij}, K)$$

5441 Where p_{ij} is some function of distance between s_i and trap location x_j . Here we
5442 use:

$$p_{ij} = 1 - \exp(-lam_{ij})$$

5443 Recall from Chapter 2 that this is the complementary log-log (cloglog) link func-
5444 tion, which constrains p_{ij} to fall between 0 and 1. For our MCMC algorithm that
5445 means that, instead of using a Poisson likelihood, $Poisson(y|sigma, lam0, s, z)$, we
5446 use a Binomial likelihood, $Binomial(y, K|sigma, lam0, s, z)$, in all the conditional
5447 distributions. As an example, Panel 6 shows the updating step for $lam0$ under
5448 a binomial encounter model. The full MCMC code for the binomial SCR can be
5449 found in the online supplements.

5450 Panel 6: MCMC updater for $lam0$ in a SCR model with Binomial encounter
5451 process and cloglog link function on detection. Here, `pmat =`

```

5452 1-exp(-lam).
5453
5454     lam0.cand <- rnorm(1, lam0, 0.1)
5455     if(lam0.cand > 0){ #automatically reject lam0.cand that are <0
5456         lam.cand <- lam0.cand*exp(-(D*D)/(2*sigma*sigma))
5457         p.cand <- 1-exp(-lam.cand)
5458         ll<- sum(dbinom(y, K, pmat *z, log=TRUE))
5459         llcand <- sum(dbinom(y, K, p.cand *z, log=TRUE))
5460         if(runif(1) < exp( llcand - ll) ){
5461             ll<-llcand
5462             pmat<-p.cand
5463             lam0<- lam0.cand
5464         }
5465     }

```

Another possibility is to model variation in the individual and site specific detection probability, p_{ij} , directly, without any transformation, such that

```

5468 pij<-p0 * exp(-Dij2/(2*sig^2))

```

and $p_0 = \{0,1\}$. This formulation is analogous to how detection probability is modeled in distance sampling under a half-normal detection function; however, in distance sampling p_0 - detection of an individual on the transect line - is assumed to be 1 (Buckland, 2001). Under this formulation the updater for lam_0 (equivalent to p_0 in Eq XX) becomes:

```

5474     lam0.cand <- rnorm(1, lam0, 0.1)
5475     if(lam0.cand > 0 & lam0.cand < 1 ){ #automatically reject lam0.cand that are not
5476         lam.cand <- lam0.cand*exp(-(D*D)/(2*sigma*sigma))
5477         ll<- sum(dbinom(y, K, lam *z, log=TRUE)) #no transformation needed
5478         llcand <- sum(dbinom(y, K, lam.cand *z, log=TRUE))
5479         if(runif(1) < exp( llcand - ll) ){
5480             ll<-llcand
5481             lam<-lam.cand
5482             lam0<- lam0.cand
5483         }
5484     }

```

5485 7.6.2 Looking at model output

Now that you have an MCMC algorithm to analyze spatial capture-recapture data with, let's run an actual analysis so we can look at the output. As an example, we will use the bear data ...⁸ You can use the same script provided back in

⁸Does this data set come up before Ch6? If not, introduce data here. Or, Andy, would you rather use simulated data?

Chapter XX to read in the data and build the augmented encounter history array; then source the MCMC code for the binomial encounter model algorithm with the cloglog link and run 5000 iterations. This should take approximately 25 minutes.

```
> source('SCR0binom.txt')
> mod0<-SCR.0(y=bigTrap, X=trapmat, M=M, xl=xl, xu=xu, yl=yl, yu=yu, K=8, niter=5000)
```

Before, we used simple R commands to look at model results. However, there is a specific R package to summarize MCMC simulation output and perform some convergence diagnostics package coda (Plummer et al., 2006). Download and install coda, then convert your model output to an mcmc object

```
> chain<-mcmc(mod0)
```

Markov chain time series plots

Start by looking at time series plots of your Markov chains using `plot(chain)`. This command produces a time series plot and marginal posterior density plots for each monitored parameter, similar to what we did before using the `hist()` and `plot()` commands (Fig. 5). Time series plots will tell you several things: First, the way the chains move through the parameter space gives you an idea of whether your MH steps are well tuned. If chains were constant over many iterations you would probably need to decrease the tuning parameter of the (Normal) proposal distribution. If a chain moves along some gradient to a stationary state very slowly, you may want to increase the tuning parameter so that the parameter space is explored more efficiently.

Second, you will be able to see if your chains converged and how many initial simulations you have to discard as burn-in. In the case of the chains shown in Figure 5, we would probably consider the first 750 - 1000 iterations as burn-in, as afterwards the chains seem to be fairly stationary.

A word of caution about chain convergence

Since we do not know what the stationary posterior distribution of our Markov chain should look like (this is the whole point of doing an MCMC approximation), we effectively have no means to assess whether it has converged to this desired distribution or not. As mentioned before, the only certainty is that a Markov chain will *eventually* converge to its stationary distribution, but no-one can tell us how long this will take. Also, you only now the part of your posterior distribution that the Markov chain has explored so far for all you know the chain could be stuck in a local maximum, while other maxima remain completely undiscovered. Acknowledging that there is truly nothing we can do to ever proof convergence of our MCMC chains, there are several things we can do to increase the degree of confidence we have about the convergence of our chains. One option, and that advocated by what we will loosely call the WinBUGS community, is to run several Markov chains and to start them off at different initial values that are overdispersed relative to the posterior distribution. Such initial values help to explore different

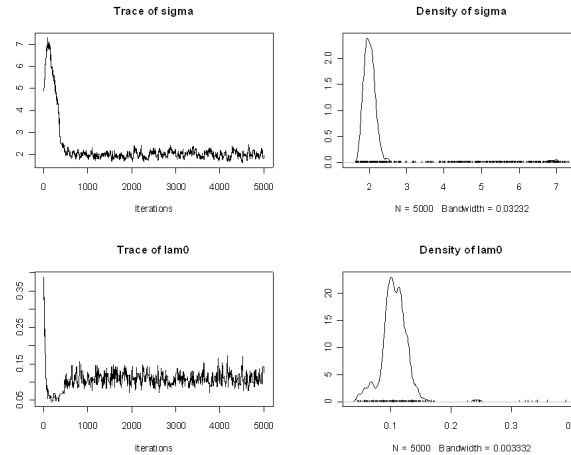


Figure 7.5. Time series and posterior density plots for sigma and lam0.

5529 areas of the parameter space simultaneously; if after a while all chains oscillate
 5530 around the same average value, chances are good that they indeed converged to
 5531 the posterior distribution. Gelman and Rubin came up with a diagnostic statistic
 5532 that essentially compares within-chain and between-chain variance to check for
 5533 convergence of multiple chains (Gelman et al., 2004). Of course, running several
 5534 parallel chains is computationally expensive. Extra computational demands are not
 5535 the only and by no means the major concern some people voice when it comes to
 5536 running several parallel MCMC chains to assess convergence. Again, consider the
 5537 fact that we do not know anything about the true form of the posterior distribution
 5538 we are trying to approximate. How do we, then, know how to pick overdispersed
 5539 initial values? We don't all we can do is pick overdispersed values relative to our
 5540 expectations of what the posterior should look like. To use a quote from the home
 5541 page of Charlie Geyer, a Bayesian statistician from the University of Minnesota,
 5542 "If you don't know any good starting points [...], then restarting the sampler at
 5543 many bad starting points is [...] part of the problem, not part of the solution."
 5544 (<http://users.stat.umn.edu/~charlie/mcmc/diag.html>). His suggestion is that your
 5545 only chance to discover a potential problem with your MCMC sampler is to run it
 5546 for a very long time. But again, there is no way of knowing how long is long enough.
 5547 It is up to you to decide, which school of thoughts appeals more to you one long
 5548 versus several parallel Markov chains. Irrespectively, part of developing an MCMC
 5549 sampler should be to make sure (within reasonable limits) that you are not missing
 5550 regions of high posterior density because of the way you specify your starting values.
 5551 Once you have explored the behavior of your chain under a reasonable range of

starting values, you may feel comfortable enough to run only one long chain. The fact that convergence cannot be proven does not mean that you should not look for potential problems in your MCMC sampler. Some problems are easily detected using simple plots, such as the time series plots we discussed above. If the overall trajectory of your chain at the end of your simulations is still upward or downward, your chain clearly has not converged and you need to run your model much longer. If you run several parallel chains and their stationary distributions look different, you may be looking at a multi-modal posterior or a problem with your sampler. With these words of caution, let's get back to looking at our model output.

7.6.3 Posterior density plots

The `plot()` command also produces posterior density plots and it is worthwhile to look at those carefully. For parameters with priors that have bounds (e.g. Uniform over some interval), you will be able to see if your choice of the prior is truncating the posterior distribution. In the context of SCR models, this will mostly involve our choice of M , the size of the augmented data set. If the posterior of N has a lot of mass concentrated close to M (or equivalently the posterior of ψ has a lot of mass concentrated close to 1), as in the example in Figure 6, we have to re-run the analysis with a larger M . A flat posterior plot shows you that the parameter essentially cannot be identified there may not be enough information in your data to estimate model parameters and you may have to consider a simpler model. Finally, posterior density plots will show you if the posterior distribution is symmetrical or skewed if the distribution has a heavy tail, using the mean as a point estimate of your parameter of interest may be biased and you may want to opt for the median or mode instead.

7.6.4 Serial autocorrelation and effective sample size

Even when we can be relatively confident that our chains have converged, the subsequent samples generated from a Markov chain are not iid samples from the posterior distribution, due to the correlation amongst samples introduced by the Markov process. As a consequence, the variance of the mean cannot simply be derived with the standard variance estimator, which takes into account the sample size (here, number of iterations). Rather, the sample size has to be adjusted to account for the autocorrelation in subsequent samples (see Chapter 8 in Robert and Casella (2010) for more details). This adjusted sample size is referred to as the effective sample size. Checking the degree of autocorrelation in your Markov chains and estimating the effective sample size your chain has generated should be part of evaluating your model output. If you use WinBUGS through the R2WinBUGS package, the `print()` command will automatically return the effective sample size for all monitored parameters. In the coda package there are several functions you

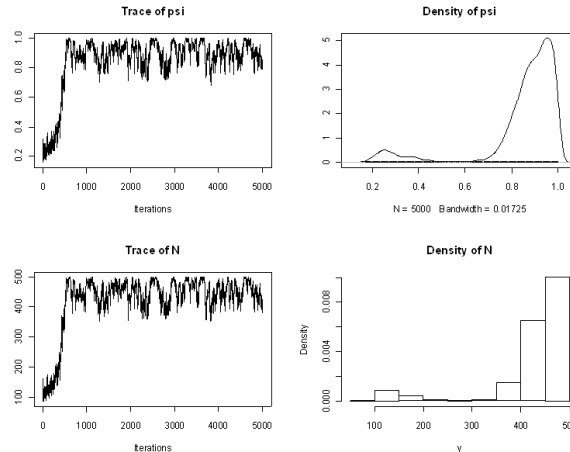


Figure 7.6. Time series and posterior density plots of ψ and N for the bear data set truncated by the upper limit of M (500).

5590 can use to do so. `effectiveSize()` will directly give you an estimate of the effective
5591 sample size for your parameters:

```
5592 > effectiveSize(chain)
5593     sigma     lam0     psi      N
5594 3.930303 78.259159 30.436348 32.047392
```

5595 Alternatively, you can use the `autocorr.diag()` function, which will show you the
5596 degree of autocorrelation for different lag values (which you can specify within the
5597 function call, we use the defaults below):

```
5598 > autocorr.diag(mcmc(mod))
5599     sigma     lam0     psi      N
5600 Lag 0  1.0000000 1.0000000 1.0000000 1.0000000
5601 Lag 1  0.9979948 0.9494134 0.9847503 0.9774201
5602 Lag 5  0.9915567 0.8038168 0.9111951 0.9113525
5603 Lag 10 0.9836016 0.6714021 0.8462108 0.8509803
5604 Lag 50 0.8985337 0.1983780 0.6138516 0.6233994
```

5605 Whichever function you use, if you find that your supposedly long Markov chain
5606 has not generated enough pseudo-iid samples, you should consider a longer run. In
5607 the present case we see that autocorrelation is especially high for the parameter
5608 `sigma` and our effective sample size for this parameter is 4!⁹ This means we would

⁹Anyone have any idea how the autocorrelation in `sigma` could be reduced?

have to run the model for much longer to obtain a reasonable effective sample size. Unfortunately, with many SCR models we observe high degrees of serial autocorrelation, which means we have to run long chains to obtain enough samples that can be considered iid, in order to obtain reasonable estimates of our parameters and their variances. What exactly constitutes a reasonable effective sample size is hard to say, but as a rule of thumb you should probably aim at several hundreds of these pseudo-iid samples. A more meaningful measure of whether you've run your chain for enough iterations is the time-series or Monte Carlo error the 'noise' introduced into your samples by the stochastic MCMC process which we introduced in Chapter 2. The MC error decreases with increasing sample size and its magnitude can thus be controlled by adjusting the length of the Markov chain. As a rule of thumb, the MC error should be 1% or less of the parameter estimate. Once you have reached this level, the estimates of the mean, standard error and 95% quantiles should no longer change significantly with additional iterations. For highly correlated samples, it will take more iterations to reduce the MC error. In coda, the MC error is given as part of the summary results (see below). Another option to deal with the serial autocorrelation of samples is to 'thin' Markov chains by some rate r and save only every r -th iteration. But as discussed in Chapter 2, this is not efficient and should only be applied if needed for practical reasons (e.g. a large number of parameters and iterations may force you to thin your samples so you object storing the model output does not become unmanageably large). For now, let's continue using this small set of samples to continue looking at the output.

7.6.5 Summary results

Now that we checked that our chains apparently have converged and pretending that we have generated enough samples from the posterior distribution, we can look at the actual parameter estimates. The `summary()` function will return two sets of results: the mean parameter estimates, with their standard deviation, the naive standard error - i.e. your regular standard error calculated for K (= number of iterations) samples without accounting for serial autocorrelation - and the corrected MC error (Time-series SE), which accounts for autocorrelation. In WinBUGS, this latter value is referred to as MC error and is only given in the log output within BUGS itself. You should adjust the `summary()` call by removing the burn-in from calculating parameter summary statistics. To do so, use the `window()` command, which lets you specify at which iteration to start 'counting'. In contrast to WinBUGS, which requires you to set the burn-in length before you run the model, this command gives us full flexibility to make decisions about the burn-in after we have seen the trajectories of our Markov chains. For our example, `summary(window(chain, start=1001))` returns the following output:

```
Iterations = 1001:5000
Thinning interval = 1
```

```

5649 Number of chains = 1
5650 Sample size per chain = 4000
5651
5652 1. Empirical mean and standard deviation for each variable,
5653    plus standard error of the mean:

```

	Mean	SD	Naive SE	Time-series SE
sigma	1.9986	0.13805	0.0021827	0.016091
lam0	0.1096	0.01523	0.0002407	0.001401
psi	0.6113	0.09148	0.0014465	0.010734
N	489.8535	71.79695	1.1352094	8.431119

```

5660
5661 2. Quantiles for each variable:

```

	2.5%	25%	50%	75%	97.5%
sigma	1.75780	1.89847	1.9900	2.0944	2.2772
lam0	0.08357	0.09824	0.1087	0.1192	0.1427
psi	0.45110	0.54838	0.6052	0.6639	0.8192
N	366.00000	440.00000	485.0000	530.0000	654.0000

5668 Looking at the MC errors, we see that in spite of the high autocorrelation, the
 5669 MC error for sigma is below the 1Our algorithm gives us a posterior distribution of
 5670 N, but we are usually interested in the density, D. Density itself is not a parameter
 5671 of our model, but we can derive a posterior distribution for D by dividing each
 5672 value of N (N at each iteration) by the area of the state-space (here 3032.719 km²)
 5673 and we can use summary statistics of this distribution to characterize D:

```

5674 > summary(window(chain[,4]/ 3032.719, start=1001))
5675 Iterations = 1001:5000
5676 Thinning interval = 1
5677 Number of chains = 1
5678 Sample size per chain = 4000
5679
5680 1. Empirical mean and standard deviation for each variable,
5681    plus standard error of the mean:

```

	Mean	SD	Naive SE	Time-series SE
	0.1615229	0.0236741	0.0003743	0.0027801

```

5685
5686 2. Quantiles for each variable:
5687
5688 2.5% 25% 50% 75% 97.5%
5689 0.1207 0.1451 0.1599 0.1748 0.2156

```

If we compare our mean density of 0.16/km² (and other parameters) with results from the same model run in secr and WinBUGS in Chapter XX, we see that estimates are almost identical (Table 1).

7.6.6 Other useful commands

While inspecting the time series plot gives you a first idea of how well you tuned your MH algorithm, use `rejectionRate()` to obtain the rejection rates (1 - acceptance rates) of the parameters that are written to your output:

```
> rejectionRate(chain)
      sigma      lam0      psi      N
0.44108822 0.77675535 0.00000000 0.01940388
```

Recall that rejection rates should lie between 0.2 and 0.8, so our tuning seems to have been appropriate here. Psi is never rejected since we update it with Gibbs sampling, where all candidate values are kept. And since N is the sum of all z, all it takes for N to change from one iteration to the next are small changes in the z-vector, so the rejection rate of N is always low. If you have run several parallel chains, you can combine them into a single mcmc object using the `mcmc.list()` command on the individual chains (note that each chain has to be converted to an mcmc object before combining them with `mcmc.list()`). You can then easily obtain the Gelman-Rubin diagnostic (Gelman et al., 2004), in WinBUGS called R-hat, using `gelman.diag()`, which will indicate if all chains have converged to the same stationary distribution. For details on these and other functions, see the coda manual, which can be found together with the package on the CRAN mirror.

7.7 MANIPULATING THE STATE-SPACE

So far, we have constrained the location of the activity centers to fall within the outermost coordinates of our rectangular state space by posing upper and lower bounds for x and y. But what if S has an irregular shape maybe there is a large water body we would like to remove from S, because we know our terrestrial study species does not occur there. Or the study takes place in a clearly defined area such as an island. As mentioned before, this situation is difficult to handle in WinBUGS. In some simple cases we can adjust the state space by setting `SXi` to be some function of `SYi` or vice versa. In this manner, we can cut off corners of the rectangle to approximate the actual state space. In R, we are much more flexible, as we can use the actual state-space polygon to constrain out si.¹⁰ To illustrate that, let's look at a camera trapping study of Florida panthers (*Puma concolor coryi*) conducted in the Picayune Strand Restoration Project (PSRP) area, southwest Florida (Fig. 7), by XXX, and financed by XXX. In the 1960ies the PSRP area was slated for

¹⁰ Have to check if we can use panther stuff for the book; otherwise, use raccoon example.

housing development, but then bought back by the State of Florida and is currently being restored to its original hydrology and vegetation. In an effort to estimate the density of the local Florida panther population, 98 camera traps were operated in the area for 21 months between 2005 and 2007. Florida panthers are wide-ranging animals and in order to account for their wide movements, the state-space was defined as the trapping grid buffered by 15 km around its outermost coordinates. However, the resulting rectangle contained some ocean in its southwestern corner (Fig. 7). In order to precisely describe the state-space, the ocean has to be removed. You can create a precise state-space polygon in ArcGIS and read it into R, or create the polygon directly within R. In the present case we intersected two shape files one of the state of Florida and one of the rectangle defined by a strip of 15 km around the camera-trapping grid. While you will most likely have to obtain the shapefile describing the landscape of and around your trapping grid (coastlines, water bodies etc.) from some external source, a polygon shapefile buffering your outermost trapping grid coordinates can easily be written in R.

If `xmin`, `xmax`, `ymin` and `ymax`, mark the outermost `x` and `y` coordinates of your trapping grid and `b` is the distance you want to buffer with, load the package `shapefiles` (Stabler, 2006) and use:

```

5743 x1= xmin-b
5744 xu= xmax+b
5745 y1= ymin-b
5746 yu= ymax+b
5747
5748 dd <- data.frame(Id=c(1,1,1,1,1),X=c(x1,xu,xu,x1,x1),Y=c(y1,y1,yu,yu,y1)) #create data fra
5749 ddTable <- data.frame(Id=c(1),Name=c("Item1"))
5750 ddShapefile <- convert.to.shapefile(dd, ddTable, "Id", 5) #convert #to shapefile, type pol
5751 write.shapefile(ddShapefile, 'c:/', arcgis=T) # save to location of #choice

```

You can read shapefiles into R loading the package `maptools` (Lewin-Koh et al., 2011) and using the function `readShapeSpatial()`. Make sure you read in shapefiles in UTM format, so that units of the trap array, the movement parameter `sigma` and the state-space are all identical. Intersection of polygons can be done in R also, using the package `rgeos` (Bivand and Rundel, 2011) and the function `gIntersect()`. The area of your single - polygon can be extracted directly from the state-space object `SSp`:

```

5759 > area <- SSp@polygons[[1]]@Polygons[[1]]@area /1000000

```

Note that dividing by 1000000 will return the area in km² if your coordinates describing the polygon are in UTM. If your state-space consists of several disjunct polygons, you will have to sum the areas of all polygons to obtain the size of the state-space. To include this polygon into our MCMC sampler we need one last spatial R package `sp` (Pebesma and Bivand, 2011), which has a function, `over()`,

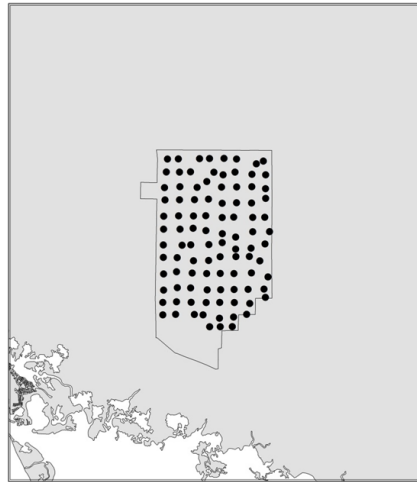


Figure 7.7. Rectangular state-space for a Florida panther camera trapping study in the PSRP area (grey outline, red block inset map of Florida) contain some ocean (white) that needs to be removed from the state-space.

```

5765 which allows us to check if a pair of coordinates falls within a polygon or not. All
5766 we have to do is embed this new check into the updating steps for s:

5767         Scand <- as.matrix(cbind(rnorm(M, S[,1], 2),
5768                                 rnorm(M, S[,2], 2)))          #draw candidate value
5769
5770 Scoord<-SpatialPoints(Scand*1000)      #convert to spatial points on UTM (m) scale
5771 SinPoly<-over(Scoord,SSp) # check if scand is within the polygon
5772
5773         for(i in 1:M) {
5774 if(is.na(SinPoly[i])==FALSE) { #if scand falls within polygon, continue update
5775   [rest of the updating step remains the same]

```

```

5776 Note that it is much more time-efficient to draw all M candidate values for s and
5777 check once if they fall within the state-space, rather than running the over() com-
5778 mand for every individual pair of coordinates. To make sure that our initial values
5779 for s also fall within the polygon of S, we use the function runifpoint() from the
5780 package spatstat (Baddeley and Turner, 2005), which generates random uniform
5781 points within a specified polygon. You'll find this modified MCMC algorithm in
5782 the online supplementary material (SCR0poisSSp). Finally, observe that we are
5783 converting candidate coordinates of S back to meters to match the UTM polygon.
5784 In all previous examples, for both the trap locations and the activity centers we

```

5785 have used UTM coordinates divided by 1000 to estimate sigma on a km scale. This
5786 is adequate for wide ranging individuals like bears. In other cases you may center
5787 all coordinates on 0. No matter what kind of transformation you use on your co-
5788 ordinates, make sure to always convert candidate values for S back to the original
5789 scale (UTM) before running the `over()` command.

7.8 MCMC SOFTWARE PACKAGES

5790 Throughout most of this book we will use WinBUGS and, occasionally, JAGS to
5791 run MCMC analyses. Here, we will briefly discuss the main pros and cons of these
5792 two programs as well as WinBUGS successor OpenBUGS. You can find scripts to
5793 simulate data and run the basic SCR model in all three programs in the online
5794 supplementary material (`simSCR0poisBUGS`).

7.8.1 WinBUGS

5795 In a nutshell, WinBUGS (and the other programs) do everything that we just went
5796 through in this chapter (and quite a bit more). Looking through your model, Win-
5797 BUGS determines which parameters it can use standard Gibbs sampling for (i.e.
5798 for conjugate full conditional distributions). Then, it determines, in the following
5799 hierarchy, whether to use adaptive rejection sampling, slice sampling or in the
5800 'worst' case Metropolis-Hastings sampling for the other full conditionals (Spiegel-
5801 halter et al., 2003). If it uses MH sampling, it will automatically tune the updater
5802 so that it works efficiently. While WinBUGS is a convenient piece of software that
5803 is still widely used, its major drawback is that it is no longer being developed, i.e.
5804 no new functions or distributions are added and no bugs are fixed.
5805

7.8.2 OpenBUGS

5806 OpenBUGS is essentially the successor of WinBUGS. While the latter is no longer
5807 worked on, OpenBUGS is constantly developed further. The name 'OpenBUGS'
5808 refers to the software being open source, so users do not need to download a license
5809 key, like they have to for WinBUGS (although the license key for WinBUGS is free
5810 and valid for life).

5812 Compared to WinBUGS, OpenBUGS has a lot more built-in functions. The
5813 method of how to determine the right updater for each model parameter has
5814 changed and the user can manually control the MCMC algorithm used to update
5815 model parameters. Several other changes have been implemented in OpenBUGS
5816 and a detailed list of differences between the two BUGS versions, can be found at
5817 <http://www.openbugs.info/w/OpenVsWin>

5818 While OpenBUGS is a useful program for a lot of MCMC sampling applications,
5819 for reasons we do not understand, simple SCR models do not converge in Open-
5820 BUGS. It is therefore advisable that you check any OpenBUGS SCR model results

against result from WinBUGS. Also, currently, the R package BRugs (Thomas et al., 2006) necessary for running OpenBUGS through R has problems with 64-bit machines, so you may have to use the 32-bit version of R and OpenBUGS in order to make it work. The BUGS project site at <http://www.openbugs.info> provides a lot of information on and download links for OpenBUGS.

There is an extensive help archive for both WinBUGS and OpenBUGS and you can subscribe to a mailing list, where people pose and answer questions of how to use these programs at <http://www.mrc-bsu.cam.ac.uk/bugs/overview/list.shtml>

7.8.3 JAGS Just Another Gibbs Sampler

JAGS, currently at Version 3.1.0, is another free program for analysis of Bayesian hierarchical models using MCMC simulation. Originally, JAGS was the only program using the BUGS language that would run on operating systems other than the 32 bit Windows platforms. By now, there are OpenBUGS versions for Linux or Macintosh machines. JAGS 'only' generates samples from the posterior distribution; analysis of the output is done in R either by running JAGS through R using either the packages `rjags` (Plummer, 2011) or `R2jags` (Su and Yajima, 2011), or by using `coda` on your JAGS output. The program, manuals and `rjags` can be downloaded at <http://sourceforge.net/projects/mcmc-jags/files/> When run from within R using the package `rjags` or `R2jags`, writing a JAGS model is virtually identical to writing a WinBUGS model. However, some functions may have slightly different names and you can look up available functions and their use in the JAGS manual. One potential downside is that JAGS can be very particular when it comes to initial values. These may have to be set as close to truth as possible for the model to start. Although JAGS lets you run several parallel Markov chains, this characteristic interferes with the idea of using overdispersed initial values for the different chains. Also, we have occasionally experienced JAGS to crash and take the R GUI with it. Only re-installing both JAGS and `rjags` seemed to solve this problem. On the plus side, JAGS usually runs a little faster than WinBUGS, sometimes considerably faster (see section 4.XYZ), is constantly being developed and improved and it has a variety of functions that are not available in WinBUGS. For example, JAGS allows you to supply observed data for some deterministic functions of unobserved variables. In BUGS we cannot supply data to logical nodes. Another useful feature is that the adaptive phase of the model (the burn-in) is run separately from the sampling from the stationary Markov chains. This allows you to easily add more iterations to the adaptive phase if necessary without the need to start from 0. There are other, more subtle differences and there is an entire manual section on differences between JAGS and OpenBUGS. For questions and problems there is a JAGS forum online at <http://sourceforge.net/projects/mcmc-jags/forums/forum/610037>.

¹¹

¹¹As we make progress on the book, let's be sure to add linkages to places where we use JAGS in examples.

7.9 SUMMARY AND OUTLOOK

While there are a number of flexible and extremely useful software packages to perform MCMC simulations, it sometimes is more efficient to develop your own MCMC algorithm. Building an MCMC code follows three basic steps: Identify your model including priors and express full conditional distributions for each model parameter. If full conditionals are parametric distributions, use Gibbs sampling to draw candidate parameter values from this distributions; otherwise use Metropolis-Hastings sampling to draw candidate values from a proposal distribution and accept or reject them based on their posterior probability densities. These custom-made MCMC algorithms give you more modeling flexibility than existing software packages, especially when it comes to handling the state-space: In BUGS (and JAGS for that matter) we define a continuous rectangular state-space using the corner coordinates to constrain the Uniform priors on the activity centers s . But what if a continuous rectangle isn't an adequate description of the state-space? In this chapter we saw that in R it only takes a few lines of code to use any arbitrary polygon shapefile as the state-space, which is especially useful when you are dealing with coastlines or large bodies of water that need removing from the state-space. Another example is the SCR R package SPACECAP (Gopalaswamy et al., 2011) that was developed because implementation of an SCR model with a discrete state-space was inefficient in WinBUGS. Another situations in which using BUGS/JAGS becomes increasingly complicated or inefficient is when using point processes other than the Uniform Poisson point process which underlies the basic SCR model (see Chapter X). In the Chapters 9 and XX you will see examples of different point processes, implemented using custom-made MCMC algorithms.¹² Finally, the Chapters XX and XX deal with unmarked or partially marked populations using hand-made MCMC algorithms to handle the (partially) latent individual encounter histories. While some of these models can be written in BUGS/JAGS,¹³ they are painstakingly slow; others cannot be implemented in BUGS/JAGS at all. In conclusion, while you can certainly get by using BUGS/JAGS for standard SCR models, knowing how to write your own MCMC sampler allows you to tailor these models to your specific needs.

¹²Richard, Beth expand on that?

¹³the Poisson one for partially marked we wrote in BUGS and it should work with a known number of marked; the Bernoulli in JAGS with the `dsum()` function should work for the fully unknown; maybe some others? I don't remember. We may have to try writing the others before saying that they don't work in BUGS/JAGS; they are certainly much faster in R, though.

8

5890

5891

5892

GOODNESS OF FIT AND STUFF

9

5893
5894
5895

COVARIATE MODELS

STATE-SPACE COVARIATES

Underlying all spatial capture recapture models is a point process model describing the distribution of individual activity centers (\mathbf{s}_i) within the state space (\mathcal{S}). So far we have focused our discussion on the homogeneous binomial point process, $\mathbf{s}_i \sim \text{Uniform}(\mathcal{S}), i = 1, 2, \dots, N$, where N is the size of the population. This is a model of “spatial-randomness”¹ because the intensity of the activity centers is constant across the study area and the activity centers are distributed independently of each other.

The spatial-randomness assumption is often viewed as restrictive because ecological processes such as territoriality and habitat selection can result in non-random distributions of organisms. We have argued, however, that this assumption is less restrictive than may be recognized because the homogeneous point process actually allows for infinite possible configurations of activity centers. Furthermore, given enough data, the uniform prior will have very little influence on the estimated locations of activity centers. Nonetheless, the homogeneous point process model does not allow one to model population density using covariates—a central objective of much ecological research. For example, a homogeneous point process model may result in a density surface map indicating that individuals were more abundant in one habitat than another, but it does not do so explicitly. A more direct approach would be to model density using covariates as is done in generalized linear models (GLMs).

In this chapter we will present a method for fitting inhomogeneous binomial point process models using covariates in much the same way as is done with GLMs. The covariates we consider differ from those covered in previous chapters, which were typically attributes of the animal (*e.g.* sex, age) and were used to model movement or encounter rate. In contrast, here we wish to model covariates that

¹The phrase “complete spatial-randomness” is reserved for the homogeneous Poisson point process

are defined for all points in \mathcal{S} , which we will refer to as state-space, or density, covariates. These may include continuous covariates such as elevation, or discrete covariates such as habitat type.

Borchers and Efford (2008) were the first to propose an inhomogeneous point process model for SCR models, and our approach is similar to theirs with the exception that we will use a binomial rather than a Poisson model because the binomial model is easily integrated into our data augmentation scheme and is consistent with the objective of determining how a *fixed* number of activity centers are distributed with respect to covariates.

The method we use to accommodate inhomogeneous binomial point process models within our MCMC algorithm is simple—we replace the uniform prior with a prior describing the distribution of the N activity centers conditional on the covariates. Development of this prior, which does not have a standard form, is a central component of this chapter. First we will begin with a review of homogeneous point process models.

10.1 HOMOGENEOUS POINT PROCESS REVISITED

The homogeneous Poisson point process is *the* model of “complete spatial randomness” and is often used in ecology as a null model to test for departures from randomness. Given its central role in the analysis of point processes, it is helpful to compare it with the binomial model that we use in our SCR models. The primary descriptor of the homogeneous point process model is the “intensity” parameter, μ which describes the expected number of points in an infinitesimally small area. The intensity parameter can also be used to determine the expected number of points in any region of the state-space \mathcal{S} . To denote this, we say that the expected number of points in region $B \in \mathcal{S}$ is $n(B) = A(B)\mu$ where $A(B)$ is the area of region B . In words, the expected number of points in B is simply the area of B multiplied by the intensity parameter. One property of the Poisson model is that if we divide the entire state-space into $k = 1, \dots, K$ disjunct regions, the counts $\mathbf{n}(\mathbf{B})$ are independent and identically distributed, (*i.i.d.*). This is one of the distinctions between the Poisson model and the binomial model, for which the counts $n(B_k)$ are not *i.i.d.* as we will explain shortly. This difference is also related to another distinction between the two models, namely that the binomial model conditions on the number of points to be simulated N ; whereas under the Poisson model N is random. Here is some simple **R** code to illustrate this point.

```

5957 mu <- 4                                # intensity
5958 Np <- rpois(1, mu)                     # Np is random
5959 PPP <- cbind(runif(Np), runif(Np)) # Poisson point process
5960
5961 Nb <- 4                                # Nb is fixed
5962 BPP <- cbind(runif(Nb), runif(Nb)) # Binomial point process

```


5963 Note that in both models, the N points are independent of one another and
 5964 distributed uniformly throughout \mathcal{S} . Thus, the intensity at any point $x \in \mathcal{S}$ is
 5965 $\mu = 1/A(\mathcal{S})$ where $A(\mathcal{S})$ denotes the area of the state-space. In the **R** code above,
 5966 the area of the state-space is 1 unit, and thus the intensity is $\mu = 1/1$.

5967 Although the Poisson model is typically described in terms of μ , the binomial
 5968 model is not; rather, it is more common to consider a discrete state space, such
 5969 as a grid with K pixels. Under the binomial model, the number of points
 5970 in each region is $n(B_k) \sim \text{Bin}(N, p_k)$ where $p_k = A(B_k)/A(\mathcal{S})$, ie p_k is simply the
 5971 fraction of the state-space area in B_k . This discrete space representation of the
 5972 binomial point process is shown in Fig. 10.1. The state-space in this case is the
 5973 unit square, and thus the probability of a point falling in each of the 25 disjunct
 5974 regions is $p_k = 1/25$ and thus the expected counts are simply $\mathbb{E}(n(B_k)) = Np_k$. In
 5975 the figure $N = 50$ and thus we would expect 2 points per pixel, which happens to be
 5976 the empirical mean of the data in Fig. 10.1. Note also that these counts are not in-
 5977 dependent realizations from a binomial distribution since $\sum_k n(B_k) = N$. Instead,
 5978 the model for the entire vector is $\mathbf{n}(\mathbf{B}) \sim \text{Multinomial}(N, \pi = (p_1, p_2, \dots, p_K))$
 5979 (Illian, 2008b). The dependence among counts has virtually no practical conse-
 5980 quence when the number of pixels is large. For example, if we have 100 pixels,
 5981 the number of counts in one pixels tells you very little about the expected count
 5982 in another pixel. However, if there are only 2 pixels, then clearly the number of
 5983 points in one pixel tells you exactly how many will occur in the remaining pixel. To
 5984 gain familiarity with the multinomial distribution and the discrete representation of
 5985 space, use the `rmultinom` function in **R** to simulate counts similar to those shown
 5986 in Fig. 10.1, for example using a command such as:

```
5987 n.B_k <- rmultinom(1, size=50, prob=rep(1/25, 25))
5988 matrix(n.B_k, 5, 5)
```

5989 The discrete space representation of the binomial point process is of practical
 5990 importance when fitting SCR models because spatial covariates are almost always
 5991 represented in a discrete format, often called “rasters” in GIS-speak. In such cases,
 5992 we often need to change our definition of the prior for an activity center from
 5993 $s_i \sim \text{Uniform}(\mathcal{S})$ to $s_i \sim \text{Multinomial}(1, \pi)$. In the latter case, the activity
 5994 center is simply defined as an integer representing pixel “id”. Note also that the
 5995 multinomial distribution with an index of 1 (i.e. `size=1` in `rmultinom`) is referred
 5996 to as the categorical distribution, which we will frequently use in the BUGS language.

10.2 INHOMOGENEOUS BINOMIAL POINT PROCESS

5997 As with the homogeneous model, the inhomogeneous binomial point process model
 5998 is developed conditional on N . The primary distinction is that the uniform distri-
 5999 bution is replaced with another distribution allowing for the intensity parameter to
 6000 vary spatially. To arrive at this new distribution, define $\mu(x, \beta)$ to be a function of

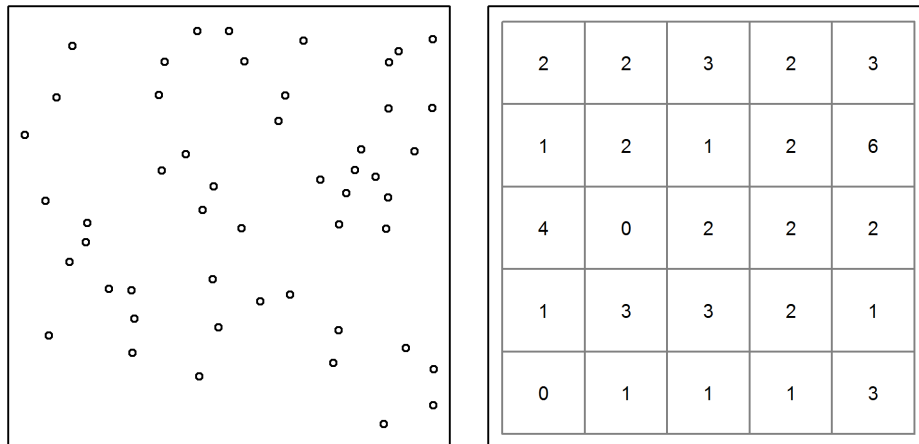


Figure 10.1. Homogeneous binomial point process with $N=50$ points represented in continuous and discrete space.

6001 spatially-referenced covariates (β) available at all points of the state space. To be
 6002 concise we will subsequently drop the vector of coefficients from our notation, and
 6003 simply use $\mu(x)$. Since an intensity must be strictly positive, it is natural to model
 6004 $\mu(x)$ using the log-link.

$$\log(\mu(x)) = \sum_{j=1}^J \beta_j v_j(x), \quad x \in \mathcal{S}$$

6005 where β_j is the regression coefficient for covariate $v_j(x)$. To be clear, $v(x)$ is the
 6006 value of any covariate, such as habitat type or elevation, at location x . This equa-
 6007 tion should look familiar because it is the standard linear model used in log-linear
 6008 GLMs. Note, however, that we have no need for an intercept because it would be
 6009 confounded with N . This should be intuitive since an intercept would represent the
 6010 expected value of N when $\beta = 0$, but we already have a parameter in the model for
 6011 expected abundance, namely $\mathbb{E}[N] = \psi M$. Thus an intercept would be redundant,
 6012 and without it we are still able to achieve our goal of describing the distribution of
 6013 N activity centers as a function of spatial covariates.

6014 Now that we have a model of the intensity parameter $\mu(x)$, we need to develop
 6015 the associated probability density function to use in place of the uniform prior.
 6016 Remembering that the integral of a pdf must be unity, we can create a pdf by
 6017 dividing $\mu(x)$ by a normalizing constant, which in this case is the integral of $\mu(x)$
 6018 evaluated over the entire state-space. **ANDY, is there a better justification for this?.**

6019 The probability density function is therefore

$$f(x) = \frac{\mu(x)}{\int_{x \in S} \mu(x) dx} \quad (10.2.1)$$

6020 Substituting this distribution for the uniform prior allows us to fit inhomogeneous
 6021 binomial point process models to spatial capture-recapture data. We can also use
 6022 this distribution to obtain the expected number of individuals in any given region.
 6023 Specifically, the proportion of N expected to occur in any region B when hetero-
 6024 geneity in density is present is $p(B) = \int_B f(x) dx$. These are also the multinomial
 6025 cell probabilities if the regions are disjoint and compose the entire state-space.

6026 As a practical matter, note that the integral in the denominator of $f(x)$ is
 6027 evaluated over space, and since we almost always regard space as two-dimensional,
 6028 this is a two-dimensional integral that can be approximated using the methods
 6029 discussed in refChXXX. These methods include Monte Carlo integration, Gaussian
 6030 quadrature, etc... Alternatively, if our state-space covariates are in raster format,
 6031 *i.e* they are in discrete space, the integral can be replaced with a sum over all pixels,
 6032 which is much more efficient computationally.

6033 We now have all the tools needed to fit inhomogeneous point process (IPP)
 6034 models. Before doing so, we note that the IPP for the activity centers results in
 6035 another IPP for the observation process, $\lambda(x)$. As a reminder, $\lambda(x)$ is the expected
 6036 number of captures for a trap at point x . As was true for the homogeneous model,
 6037 this intensity function is a product of the point process intensity and the encounter
 6038 rate function, $\lambda(x) = \mu(x)g(x)$.

6039 In the next section we walk through a few examples, building up from the
 6040 simplest case where we actually observe the activity centers as though they were
 6041 data. In the second example, we fit our new model to simulated data in which
 6042 density is a function of a single continuous covariate. Example three shows an
 6043 analysis in discrete space using both **secr** (Efford, 2011) and **JAGS** (Plummer,
 6044 2003). In the last example, we model the intensity of activity centers for a real
 6045 dataset collected on jaguars (*Panthera onca*) in Argentina.

10.3 EXAMPLES

6046 10.3.1 Simulation and analysis of inhomogeneous point processes

6047 In SCR models, the point process is not directly observed, but in other contexts it
 6048 is. Examples include the locations of disease outbreaks, the locations of trees in a
 6049 forest, or the locations of radio-tracked animals. Indeed Eq. 10.2.1 has been used
 6050 extensively in the radio-telemetry literature to model so-called “resource selection
 6051 functions” (??). In such cases where the point locations are directly observed,
 6052 estimating the parameters β is straight-forward as demonstrated in the following
 6053 example. This example also illustrates the fundamental process that we will later
 6054 embed in our MCMC algorithm used to fit SCR models.

Suppose we knew the locations of 100 animals' activity centers. To estimate the intensity surface $\mu(x)$ underlying these points, we need to derive the likelihood for our data under this model. Given the pdf $f(x)$ (Eq. 10.2.1) and assuming that the points are mutually independent of one another, we may write the likelihood as the product of R such terms, where $R = 100$ is the sample size in this case, *ie* the observed number of activity centers.

$$\mathcal{L}(\beta|\mathbf{x}_i) = \prod_{i=1}^R f(x_i)$$

Having defined the likelihood we could choose a prior and obtain the posterior for β using Bayesian methods, or we can find the maximum likelihood estimates (MLEs) using standard numerical methods as is demonstrated below.

First, let's simulate some data. Simulating data under an inhomogeneous point process model is often accomplished using indirect methods such as rejection sampling. Rejection sampling proceeds by simulating data from a standard distribution and then accepting or rejecting each sample using probabilities defined by the distribution of interest. For more information, readers should consult an accessible text such as Robert and Casella (2004). In our example, we simulate from a uniform distribution and then accept or reject using the (scaled) probability density function $f(x)$. Note that we first define a spatial covariate (elevation) that is a simple function of the spatial coordinates increasing from the southwest to the northeast of our state-space.²

The following **R** commands demonstrate the use of rejection sampling to simulate an inhomogeneous point process for the covariate depicted in Fig. 10.3.1. The code uses the **cuhre** function in the **R2Cuba** package to integrate the intensity function over space (Hahn et al., 2011).

```
# spatial covariate (with mean 0)
elev.fn <- function(x) x[1]+x[2]-1
# intensity function
mu <- function(x, beta) exp(beta*elev.fn(x=x))

# Simulate PP using rejection sampling
set.seed(300225)
N <- 100
count <- 1
s <- matrix(NA, N, 2)
beta <- 2 # parameter of interest
int.mu <- R2Cuba::cuhre(2, 1, mu, beta=beta)$value
elev.min <- elev.fn(c(0,0)) #elev.fn(cbind(0,0))
elev.max <- elev.fn(c(1,1)) #elev.fn(cbind(1,1))
```

²Such functional forms of covariates are rarely available, which is why continuous spatial covariates are more often measured on a discrete grid.

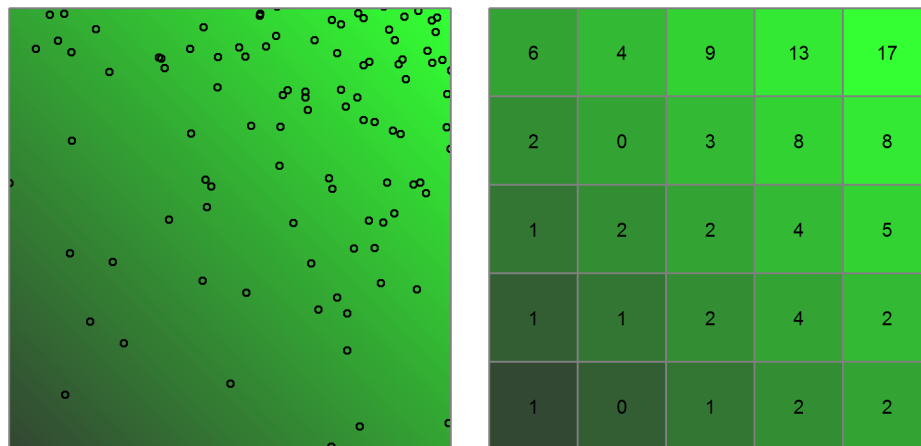


Figure 10.2. An example of a spatial covariate, say elevation, and a realization of an inhomogeneous binomial point process with $N=100$ and $\mu(x) = \exp(\beta \text{Elev})$ where $\beta = 2$.

```

6092 Q <- max(c(exp(beta*elev.min) / int.mu, #2d(beta),
6093            exp(beta*elev.max) / int.mu)) #2d(beta))
6094 while(count <= 100) {
6095   x.c <- runif(1, 0, 1); y.c <- runif(1, 0, 1)
6096   s.cand <- c(x.c,y.c)
6097   pr <- exp(beta*elev.fn(s.cand)) / int.mu #2d(beta)
6098   if(runif(1) < pr/Q) {
6099     s[count,] <- s.cand
6100     count <- count+1
6101   }
6102 }

```

6103 The simulated data are shown in Fig 10.3.1. High elevations are represented by
6104 light green and low elevations by dark green. The activity centers of one hundred
6105 animals are shown as points, and it is clear that these simulated animals prefer the
6106 high elevations. Perhaps they are mountain goats. The underlying model describing
6107 this preference is $\log(\mu(x)) = \exp(\beta \times \text{Elevation}(x))$ where $\beta = 2$ is the parameter
6108 to be estimated and $\text{Elevation}(x)$ is a function of the coordinates at x , as displayed
6109 on the map.

6110 Given these points, we will now estimate β by minimizing the negative-log-
6111 likelihood using R's `optim` function.

```

6112 # Negative log-likelihood

```

```

6113 nll <- function(beta) {
6114     int.mu <- cuhre(2, 1, mu, beta=beta)$value
6115     -sum(beta*elev.fn(s) - log(int.mu))
6116 }
6117 starting.value <- 0
6118 fm <- optim(starting.value, nll, method="Brent",
6119             lower=-5, upper=5, hessian=TRUE)
6120 c(Est=fm$par, SE=sqrt(1/fm$hessian)) # estimates and SEs

```

6121 Maximizing the likelihood took a small fraction of a second, and we obtained
 6122 an estimate of $\hat{\beta} = 1.99$. We could plug in this estimate to our linear model at each
 6123 point in the state-space to obtain the MLE for the intensity surface.

6124 This example demonstrates that if we had the data we wish we had, *i.e.* if we
 6125 knew the coordinates of the activity centers, we could easily estimate the parameters
 6126 governing the underlying point process. Unfortunately, in SCR models, the activity
 6127 centers cannot be directly observed, but spatial re-captures, that is captures of
 6128 individuals at multiple locations in space, provide us with the information needed
 6129 to estimate these latent parameters.

6130 10.3.2 Fitting inhomogeneous point process SCR models

6131 Continuous space

6132 One of the nice things about hierarchical models is that they allow us to break a
 6133 problem up into a series of simple conditional relationships. Thus, we can simply
 6134 add the methods described above into our existing MCMC algorithm to simulate
 6135 the posteriors of β conditional on the simulated values of \mathbf{s}_i . To demonstrate, we
 6136 will continue with the previous example. Specifically, we will overlay a grid of
 6137 traps upon the map shown in Fig. 10.3.1. We will then simulate capture histories
 6138 conditional upon the activity centers shown on the map. Then, we will attempt to
 6139 estimate the activity center locations as though we did not know where they were,
 6140 as is the case in real applications.

6141 Here is some **R** code to simulate the encounter histories under a Poisson ob-
 6142 servation model, which would be appropriate if animals could be detected multiple
 6143 times at a trap during a single occasion.

```

6144 # Create trap locations
6145 xsp <- seq(-0.8, 0.8, by=0.2)
6146 len <- length(xsp)
6147 X <- cbind(rep(xsp, each=len), rep(xsp, times=len))
6148
6149 # Simulate capture histories, and augment the data
6150 ntraps <- nrow(X)
6151 T <- 5
6152 y <- array(NA, c(N, ntraps, T))
6153

```

```

6154 nz <- 50 # augmentation
6155 M <- nz+nrow(y)
6156 yz <- array(0, c(M, ntraps, T))
6157
6158 sigma <- 0.1 # half-normal scale parameter
6159 lam0 <- 0.5 # basal encounter rate
6160 lam <- matrix(NA, N, ntraps)
6161
6162 set.seed(5588)
6163 for(i in 1:N) {
6164   for(j in 1:ntraps) {
6165     distSq <- (s[i,1]-X[j,1])^2 + (s[i,2] - X[j,2])^2
6166     lam[i,j] <- exp(-distSq/(2*sigma^2)) * lam0
6167     y[i,j,] <- rpois(T, lam[i,j])
6168   }
6169 }
6170 yz[1:nrow(y),,] <- y # Fill

```

Now that we have a simulated capture-recapture dataset y , and we have augmented it to create the new data object yz , we are ready to begin sampling from the posteriors. A commented Gibbs sampler written in **R** is available in the accompanying **R** package **scrbook** (see `?scrIPP`). There are two small parts of the **R** code that distinguish it from previous code we have shown to fit homogeneous point processes. First, we need to update the parameter β conditional on all other parameters in the model. The code to do so is:

```

6178 D1 <- cuhre(2, 1, mu, lower=c(xlims[1], ylims[1]),
6179             upper=c(xlims[2], ylims[2]), beta=beta1)$value
6180 beta1.cand <- rnorm(1, beta1, tune[3])
6181 D1.cand <- cuhre(2, 1, mu, lower=c(xlims[1], ylims[1]),
6182                 upper=c(xlims[2], ylims[2]), beta=beta1.cand)$value
6183 ll.beta1 <- sum( beta1*elev.fn.v(S) - log(D1) )
6184 ll.beta1.cand <- sum( beta1.cand*elev.fn.v(S) - log(D1.cand) )
6185 if(runif(1) < exp(ll.beta1.cand - ll.beta1) ) {
6186   beta1<-beta1.cand
6187 }

```

Next, we need to put the new prior on the activity centers:

```

6189 #ln(prior), denominator is constant
6190 prior.S <- beta1*cov(S[i,1], S[i,2]) # - log(D1)
6191 prior.S.cand <- beta1*(Scand[1] + Scand[2]) # - log(D1)
6192 if(runif(1)< exp((ll.S.cand+prior.S.cand) - (ll.S+prior.S))) {
6193   S[i,] <- Scand
6194   lam <- lam.cand
6195   D[i,] <- dtmp
6196 }

```

We can apply this modified sampler to our data using the code shown in the help file for `scrIPP`. We obtain posterior distributions summarized in Table 10.1. Mixing is good, and as usual, life is very nice when we are working with simulated data.

Fitting continuous space IPP models is somewhat difficult in **BUGS** because our prior $f(x)$ is not one of the available distributions that come with the software³ `seccr` allows users to fit continuous space using polynomials of the x - and y - coordinates, but not for truly continuous covariates. However, these are not really important limitations because discrete space versions are straight-forward, and virtually all spatial covariates are defined as such.

Discrete space

To fit discrete space models, we follow the same steps as outlined in Chapter XXX—we define s_i as pixel ID, and we use the categorical distribution as a prior. A good example of this is in `+citeKery capricaillie`. Here we present an analysis of the simulated data shown in the right panel of Fig. 10.3.1. The spatial covariate, let's call it elevation again, was simulated from a kriging type of model as shown on the help page `ch9simData` in `scrbook`. The points are the number of activity centers in each pixel, generated from a single realization of the IPP $\mu(x) = 2elev$.

The **BUGS** code to fit an IPP model to these data is shown in the following panel.

```

model{
  sigma ~ dunif(0, 1)
  lam0 ~ dunif(0, 5)
  beta ~ dnorm(0,0.1)
  psi ~ dbeta(1,1)
  for(j in 1:nPix) {
    theta[j] <- exp(beta*elevation[j])
    probs[j] <- theta[j]/sum(theta[])
  }
  for(i in 1:M) {

```

³It is possible, if somewhat cumbersome, to add new distributions in **BUGS**.

Table 10.1. Posterior summaries from inhomogeneous point proces model

	Mean	SD	2.5%	50%	97.5%
$\sigma = 0.10$	0.1026	0.0048	0.0935	0.1025	0.1123
$\lambda_0 = 0.50$	0.4419	0.0493	0.3496	0.4400	0.5390
$\psi = 0.66$	0.6826	0.0554	0.5762	0.6820	0.7923
$\beta = 2.00$	2.1601	0.3390	1.5193	2.1583	2.8043
$N = 100$	102.7696	6.2689	92.0000	102.0000	117.0000

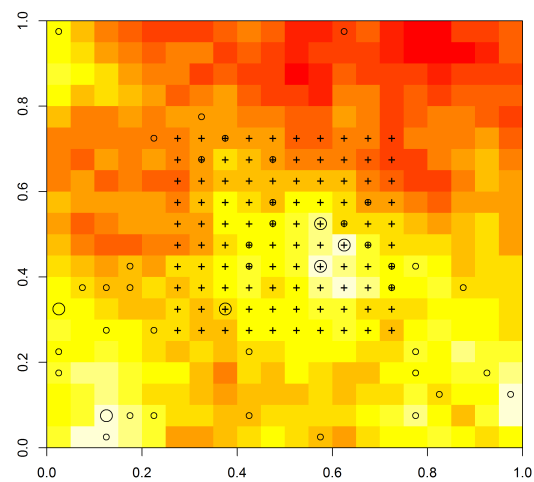


Figure 10.3. Simulated activity centers in discrete space. The spatial covariate, elevation, is highest in the higher areas. Density of activity centers (circles) increases with elevation. Trap locations are shown as crosses.

```

6229   w[i] ~ dbern(psi)
6230   s[i] ~ dcat(probs[])
6231   x0g[i] <- Sgrid[s[i],1]
6232   y0g[i] <- Sgrid[s[i],2]
6233   for(j in 1:ntraps) {
6234     dist[i,j] <- sqrt(pow(x0g[i]-grid[j,1],2) + pow(y0g[i]-grid[j,2],2))
6235     lambda[i,j] <- lam0*exp(-dist[i,j]*dist[i,j]/(2*sigma*sigma)) * w[i]
6236     y[i,j] ~ dpois(lambda[i,j])
6237   }
6238 }
6239
6240 N <- sum(w[])
6241 Density <- N/1 # unit square
6242 }

```

6243 This model can also be fit in **secr**, which refers to the pixel locations as a
 6244 “mask”. **R** code to fit the models using **secr** and **JAGS** is available in **scrbook**,
 6245 see `help(ch9secrYjags)`. Results of the comparison are shown in Table 10.2 and
 6246 are very similar as expected.

6247 Density surface maps can be created for fun, and of course to inform manage-
 6248 ment decisions. [describe how to do this]

6249 10.3.3 The jaguar data

6250 Estimating density of large felines has been a priority for many conservation orga-
 6251 nizations, but no robust methodologies existed before the advent of SCR. Distance
 6252 sampling is not feasible for such rare and cryptic species, and traditional capture-
 6253 recapture methods yield estimates that are highly sensitive to the subjective choice
 6254 of the effective survey area. In this example, we demonstrate how readily density
 6255 can be estimated for a globally imperilled species using SCR. Furthermore, we show
 6256 how inhomogeneous point process models can be used to test important hypotheses
 6257 regarding the factors affecting density.

6258 In this example, we make use a single year of data from an 8-year camera-

Table 10.2. Comparison of **secr** and **JAGS** results

Software	Par	Est.	SD	lower	upper
secr	N	49.2803	5.7535	41.0087	64.3879
	β	2.1772	0.5628	1.0741	3.2804
	λ_0	0.9203	0.0764	0.7824	1.0825
	σ	0.0990	0.0038	0.0918	0.1068
JAGS	N	48.2072	5.4053	39.0000	60.0000
	β	2.1026	0.5323	1.0889	3.1506
	λ_0	0.9328	0.0766	0.7898	1.0921
	σ	0.1004	0.0041	0.0929	0.1089

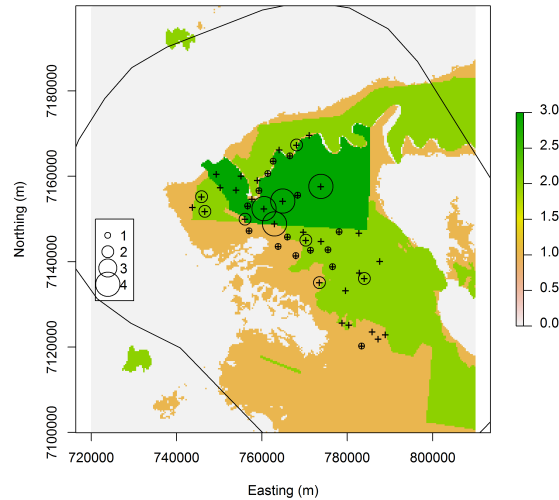


Figure 10.4. Jaguar detections at 46 camera trap stations. The three levels of protection status are no protection (beige), some protection (light green), and national park (dark green). Non-habitat is shown in gray and represents large soybean monocultures.

trapping study of jaguars (*Panthera onca*) in Argentina, along the borders with Brazil and Paraguay. The data consist of 46 camera stations, each consisting of a camera on either side of a road or trail. Forty-five detections of 16 jaguars were made over a 95 day sampling period. The mean number of sampling days at each camera station was 48.2. Eight females and 8 males were detected.

Estimating density is a central objective of this study because ultimately, an estimate of the total population size for the “green corridor” is needed. A second, and related, objective was to assess the influence of poaching on jaguar density. Although jaguars themselves are occasionally killed by poachers, the bigger influence is the effect of poaching on prey species such as peccaries (*Pecari tajacu*, *Tayassu pecari* ASK AGUSTIN WHICH ARE MORE COMMON). To protect jaguars and related species, protected areas have been established (DETAILS) and three levels of protection are recognized as depicted in Fig. 10.3.3.

MENTION SEX-SPECIFIC SIGMAS

MENTION THE RIVER

To assess the influence of protection on jaguar density, we treated protection status as an ordinal variable with 3 levels: no protection, some protection???, national park. Clearly these are ordered, and our hypothesis is that density should

Table 10.3. "Jaguar density estimates and associated parameters"

	Mean	SD	2.5%	50%	97.5%
sigmaF	7340.1547	1998.5926	4736.1803	6933.9760	12440.3079
sigmaM	8154.5222	1581.5224	5801.0509	7916.7572	11890.7957
rho	0.5163	0.1170	0.2874	0.5172	0.7388
lam0	0.0071	0.0024	0.0033	0.0067	0.0126
psi	0.3145	0.0699	0.1908	0.3100	0.4638
beta	4.3238	1.5115	2.4148	4.0295	8.0894
N	20.3828	2.8671	16.0000	20.0000	27.0000
N1	0.2759	0.6360	0.0000	0.0000	2.0000
N2	2.6285	1.8589	0.0000	2.0000	7.0000
N3	17.4784	2.8458	12.0000	17.0000	24.0000

increase with the level of protection. Thus, β in this example is a single “slope” parameter describing the degree to which protection status affects jaguar density. `°` code to fit the model are available in `scrbook`. Parameter estimates are shown in Table XXX. Our results indicate that efforts to protect jaguars by reducing poaching are working. Density was X times higher in the national park than in the unprotected areas. Fig. 10.3.3 shows the estimated density surface.

We note that there is room for improvement in our analysis. The political boundaries use to demarcate protected areas are not as concrete as we might like. In reality poaching pressure is likely to be higher near remote park boundaries than in well-guarded park interiors. One option for addressing this would be to use a continuous measure of poaching pressure such as distance from the nearest town, or some other accessibility metric. It would also be interesting to model density separately for each sex. Many of the detections outside of the park were of males, and thus it is possible that the sexes use habitat differently.

10.4 SUMMARY

When state-space covariates are available, we can model density by replacing the uniform prior on the activity centers with a prior based on a normalized log-linear function of covariates. This distribution has been widely used in ecology to model point processes as well as resource selection probability functions. In our SCR context, use of this new prior results in a model for the inhomogeneous point process describing the location of activity centers, which can be used to test hypotheses about covariates affecting density. In rare cases, these covariates are truly continuous in the sense that they are defined as a function of space. More often, covariates are represented on rasters, which simplifies the analysis. Fitting these models can be accomplished using **BUGS**, **secr**, or the custom **R** code presented in this chapter and found in the package `scrbook`.

Note that density cannot be modeled using traditional CR methods.

All the examples in this section included a single state-space covariate, but this

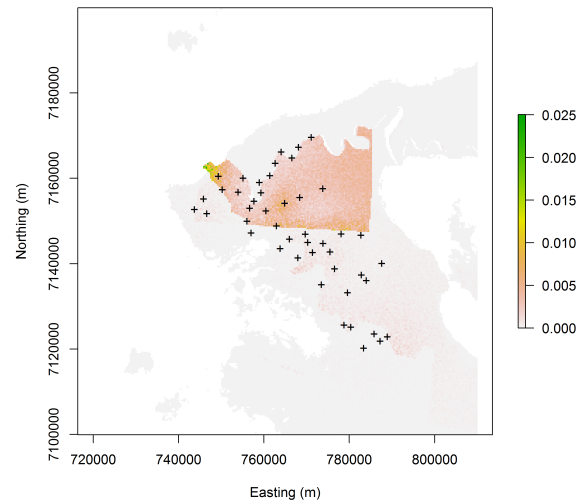


Figure 10.5. Estimated density surface for the jaguar dataset

6304 was for simplicity only. Including multiple covariates poses no additional challenges.
6305 Likewise, additional model structure such sex-specific encounter rate parameters or
6306 behavioral responses can be accommodated.

6307
6308
6309

11

OPEN MODELS

SPATIAL CAPTURE-RECAPTURE FOR UNMARKED POPULATIONS

Traditional capture-recapture models share the fundamental assumption that each individual in a population can be uniquely identified when captured. This can often be accomplished by marking individuals with color bands, ear tags, or some other artificial mark that can be subsequently read in the field. For other species, such as tigers or marbled salamanders, individuals can be easily identified using only their natural markings. Many species, however, do not possess adequate natural markings and are difficult to capture, making it impractical to use standard capture-recapture techniques.

Estimating density when individuals are unmarked can be accomplished using a variety of alternatives to capture-recapture. However, many of these methods have important limitations that warrant the exploration of new approaches. In this chapter we highlight the work of Chandler and Royle (2012) who demonstrated that the “individual recognition” assumption of capture-recapture models is not a requirement of spatial capture-recapture models. They showed that spatial correlation alone is sufficient for making inference about animal distribution and density. That is, if we simply have spatially-correlated count data at a collection of survey points, we can estimate density even if all individuals are unmarked, assuming that the underlying SCR model is valid. This “spatial count model” is virtually identical to other SCR models except that the encounter histories are not directly observed.

The ability to fit SCR models to data from unmarked populations has important consequences. For one, it means that SCR models can be applied to data collected using methods like point counts in which observers record simple counts of animals at an array of survey points. This development also has important implications for traditional SCR studies because many SCR datasets include some individuals that cannot be identified due to poor photo quality or indistinguishable natural

6339 markings.

6340 It's also interesting to note that by disregarding individual identity, we wind up
6341 with a model that closely resembles another large class of spatial models, known as
6342 convolution models (Wolpert and Ickstadt, 1998; Higdon, 1998). These models have
6343 been used for purposes as varied as describing oceanic surface temperatures and
6344 correlation in tree locations within managed forests. The Chandler-Royle model
6345 offers an improvement in some respects over existing convolution models because
6346 it does not require arbitrary decisions about the location and number of "support
6347 points". We will clarify this later in the chapter, and briefly mention how this
6348 model can be used outside of SCR contexts for general purpose spatial modeling of
6349 correlated count data.

12.1 EXISTING MODELS FOR INFERENCE ABOUT DENSITY IN UNMARKED POPULATIONS

6350 When capture-recapture methods are not a viable option, researchers often col-
6351 lect simple count data or even detection/non-detection data to estimate population
6352 parameters. These data are often analyzed using Poisson regression or logistic re-
6353 gression, perhaps with random effects. When detection is imperfect, as it almost
6354 always is, these methods cannot be used to obtain unbiased estimates of population
6355 size or occurrence probability. Even when these data are used an index of abun-
6356 dance or occurrence, standard models may yield unreliable results when covariates
6357 affect both the state variable and detection probability. A classic example is the
6358 finding by Bibby and Buckland (1987) who reported that the probability of de-
6359 tecting songbirds in restocked conifer plantations decreased with vegetation height;
6360 whereas population density was positively related to vegetation height. This in-
6361 tuitive and common phenomenon has led to the development a vast number of
6362 models to estimate population size and detection probability when individuals are
6363 unmarked. A review of these models is beyond the scope of this chapter, but we
6364 mention a few deficiencies of existing methods that warrant the exploration of al-
6365 ternatives for robust inference when standard capture-recapture methods do not
6366 apply.

6367 Distance sampling (Buckland, 2001), which we briefly introduced in chapter
6368 XXXX, is perhaps the most widely used method for estimating population density
6369 when individuals are unmarked and detection probability is less than one. This
6370 class of methods is known to work impecibly when estimating the number of stakes
6371 in a field or the number of duck nests in a wetland. Distance sampling can also work
6372 very well in more interesting situations, and is an extremely powerful method when
6373 the assumptions can be met. However, the assumptions that distance data can be
6374 recorded without error and that animals are distributed randomly with respect to
6375 the transect can be easily violated by common processes such as animal movement
6376 and measurement error may result in substantial bias. Although numerous methods
6377 have been proposed to relax some of these assumptions Royle et al. (2004); ?);

Johnson (2010); Chandler et al. (2011), another issue is that distance sampling is simply not practical in many settings. For example, many species are so rare and elusive that they can only be reliably surveyed using methods such as camera traps.

Other common sampling methods used to estimate density when individuals are unmarked include double-observer sampling, removal sampling, and repeated counts. Models for each of these data structures have been developed (Royle, 2004; ?). To obtain reliable density estimates using these methods, the area surveyed must be well defined and closed with respect to movement and demographic processes. Given a short enough sampling interval, such as a 5-min point-count, the closure assumption may be reasonable. However, short sampling intervals can reduce the number of detections, so observers generally visit each survey point multiple times during a season. But then you can expect animal movement to invalidate the closure assumption, so there is this kind of endless morass of shit to worry about and you might as well jump off a cliff—especially if you start thinking about false positives!

MENTION SIMILARITY BETWEEN THE TEMPORARY EMIGRATION PROBLEM FACED HERE AND THE ONE THAT MOTIVATES SCR

Deep breath. Anyhow, we mention these issues not to suggest that existing models do not have value—indeed we believe that they can be used to obtain reliable density estimates in many situations—rather our aim is to highlight the need for alternative methods when the assumptions of existing methods cannot be met. Additionally, the model we develop in this chapter serves as the foundation for a broad class of SCR models in which all or some of the individuals cannot be uniquely identified.

12.2 SPATIAL CORRELATION AS INFORMATION

All of the previous methods require some sort of auxiliary information to separately model the state and observation processes. That is, we need multiple observers or distance data or repeated visits to ensure that model parameters are identifiable¹. The same is true for Chandler-Royle model, but the auxiliary information comes in the form of spatial correlation (Chandler and Royle, 2012).

It is natural to be suspicious of the claim that spatial correlation is a good thing. Indeed, elaborate methods have been devised to deal with spatial correlation as a nuisance parameter (?), and ecologists have been admonished for failing to obtain “real” replicates uncontaminated by spatial correlation (Hurlbert, 1984). The following heuristic may be helpful.

Imagine a 10×10 grid of camera traps and a single individual exposed to capture whose home range center lies in the center of the trapping grid. If the individual has a small home range size relative to the extent of the trapping grid, we can imagine what the spatial correlation structure of the encounters might look like.

¹Or we can make very strong model assumptions and get away without any auxiliary data (?)

If the animal's movement is symmetric around the activity center then the number of times the individual is detected at each trap (the trap counts) is a function of the distance between the home range center and the trap, *i.e.* traps with the same distance from the activity center will yield counts that are more highly correlated with one another than traps located at different distances from the activity center. Thus, the correlation in counts tells us something about the location of the activity center. That is, correlation carries information about distribution. What about density?

Imagine now that there are two activity centers located in our trapping grid. Using trap counts alone, can our model tell us both where the activity centers are and how many exist in the population exposed to capture? The answer is yes, at least under certain circumstances. EXPLAIN.

Thus, intuition alone suggests that we might be able to estimate the number and location of activity centers using simple count data that are spatially-correlated. The results of Chandler and Royle (2012) provide additional support for this assertion.

12.3 DATA REQUIREMENTS AND SURVEY DESIGNS

One of the important benefits of Chandler-Royle model is that it can be applied to data collected using an enormous variety of survey methods. Whereas traditional SCR models require spatially-referenced encounter histories, this model requires simple count data. Once again, suppose that we have J traps surveyed on K time periods during which no births or deaths occur. Suppose also that an individual can be encountered at multiple traps during a single time period, say one night during a camera-trapping study. If individuals can also be encountered multiple times at a single trap during a night, then the encounter history data could be regarded as

$$z_{ijk} \sim \text{Poisson}(\lambda_{ij}). \quad (12.3.1)$$

if the animals were marked. Here λ_{ij} is the encounter rate for individual i at trap j . A common form of this parameter is

$$\lambda_{ij} = \lambda_0 \exp(\|\mathbf{x}_j - \mathbf{s}_i\|/2\sigma^2)$$

where λ_0 is the baseline encounter rate and σ is the scale parameter describing the distance-related decay in encounter rate.

When individuals cannot be uniquely identified, the z_{ijk} cannot be directly observed. So just when you thought we ran out of things to treat as latent variables, we are now going to regard even the encounter histories as latent. The data are now just a reduced-information summary of the latent encounter histories. That is, they are the sample- and trap-specific totals, aggregated over all individuals:

$$n_{jk} = \sum_{i=1}^N z_{ijk}.$$

This data structure, a matrix of counts made at a collection of sampling locations on one or more occasions is extremely common in ecology. Note also that the reason why we can get by with a single occasion of data ($J \equiv 1$) is that, under the Poisson model,

$$n_{jk} \sim \text{Poisson}(\Lambda_j) \quad (12.3.2)$$

where

$$\Lambda_j = \lambda_0 \sum_i k_{ij},$$

and because Λ_r does not depend on t , we can aggregate the replicated counts, defining $n_{j\cdot} = \sum_k n_{jk}$ and then

$$n_{j\cdot} \sim \text{Poisson}(K\Lambda_j)$$

As such, K and λ_0 serve equivalent roles as affecting baseline encounter rate. This has been noted elsewhere (?).

We imagine that other observation models might be possible (see Discussion) although we focus here on the Poisson encounter model because it has considerable relevance to animal surveys, and has additional methodological context related to point process models which we address in the Discussion.

12.4 ENCOUNTER HISTORIES AS LATENT VARIABLES

State model is the same as other SCR models.

This formulation of the model in terms of the aggregate count simplifies computations as the latent variables z_{irt} do not need to be updated in the MCMC estimation scheme (see below). However, retaining z_{irt} in the formulation of the model is important if some individuals are uniquely marked, in which case modifying the MCMC algorithm (see below) to include both types of data is trivial. This is because uniquely identifiable individuals produce observations of some of the z_{irt} variables.

12.5 ESTIMATION BY MCMC

We adopt a Bayesian framework for inference allowing estimation of N while retaining the formulation of the model that is conditional on the latent activity centers \mathbf{s}_i . Specifically, we employ Markov chain Monte Carlo (MCMC) to simulate posterior distributions of the parameters. However, the fact that N is unknown presents a technical challenge because the size of the parameter space can change with each MC iteration. To resolve this, we adopt the formulation of data augmentation in Royle et al. (2007) who used a specific prior construction for N in terms of individual level Bernoulli trials. In particular, we assume $N \sim \text{Unif}(0, M)$ for some large integer M . We construct this prior by assuming $N|M, \phi \sim \text{Bin}(M, \phi)$ and $\phi \sim \text{DUnif}(0, 1)$ which implies, marginally, that N has the requisite $\text{DUnif}(0, M)$

distribution. However the hierarchical formulation of the prior suggests an implementation in which we introduce a set of latent indicator variables $w_i \sim \text{Bern}(\phi)$ and, furthermore, the model implies that z_{irt} are obtained from the specified distribution (Eq. 12.3.1) if $w_i = 1$, or if $w_i = 0$, $z_{irt} = 0$ with probability 1. In effect, extending the model in this way induces a reparameterization for the latent counts that is a zero-inflated version of the original conditional-on- N model. Specifically, the model under data augmentation becomes

$$\begin{aligned} z_{irt} | w_i &\sim \text{Poisson}(\lambda_{ir} w_i) \\ w_i &\sim \text{Bern}(\phi) \end{aligned}$$

Under this formulation $N = \sum_{i=1}^M w_i$, and population density is simply $D = N/A(\mathcal{S})$ where $A(\mathcal{S})$ is the area of the point process state-space \mathcal{S} .

We developed two distinct MCMC implementations for this model. In the first, we devised an algorithm for the model conditional on the latent variables z_{irt} . This formulation is useful for problems in which one or more individual identities are available, in which case the z_{irt} are observable for those individuals. The unobserved z_{irt} are easily updated using their full-conditional distribution which is multinomial with sample size n_{rt} . The remaining parameters are updated using Metropolis-Hastings steps. In the second formulation of the algorithm we applied the Metropolis-Hastings algorithm to the model *unconditional* on the z_{irt} variables. In that case, the marginal distribution for n_{rt} is precisely Eq. 12.3.2. This algorithm is slightly more convenient because it avoids having to update the z_{irt} variables of which there are many.

12.6 NORTHERN PARULA EXAMPLE

To apply our model to data collected in the field, we designed a point count study of the northern parula (*Parula americana*), a Neotropical-Nearctic migratory passerine. This species defends well-defined territories during the breeding season (?), and thus our modeling effort was focused on estimating the number and location of territory centers. Points were located on a 50-m grid to ensure spatial correlation. This small grid spacing contrasts with the conventional practice of spacing points by > 200 m to obtain *i.i.d.* counts. Figure 12.1 depicts the spatially-correlated counts ($n_{r.}$) from the 105 point count locations surveyed three times each during June 2006 at the Patuxent Wildlife Research Center in Laurel Maryland, USA. A total of 226 detections were made with a maximum count of 4 during a single survey. At 38 points, no warblers were detected. All but one of the detections were of singing males, and this one observation was not included in the analysis.

In our analysis of the parula data, we defined the point process state-space by buffering the grid of point count locations by 250 m and used $M = 300$. We simulated posterior distributions using three Markov chains, each consisting of 300000 iterations after discarding the initial 10000 draws. Convergence was satisfactory, as indicated by an \hat{R} statistic of < 1.02 (Gelman and Rubin, 1992).

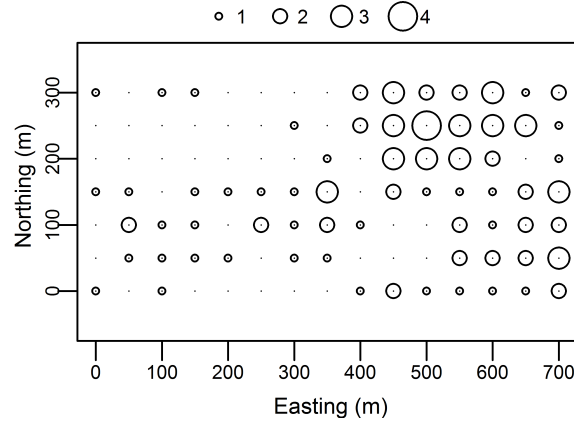


Figure 12.1. Spatially-correlated counts of northern parula on a 50-m grid. The size of the circle represents the total number of detections at each point.

One benefit of a Bayesian analysis is that it can accommodate prior information on the home range size and encounter rate parameters, which are readily available for many species. To illustrate, we analyzed the parula data using two sets of priors. In the first set, all priors were improper, customary non-informative priors (see Table 12.1). Uniform priors were also used in the second set, with the exception of an informative prior for the scale parameter $\sigma \sim \text{Gamma}(13, 10)$. We arrived at this prior using the methods described by Royle et al. (2011b) and published information on the warbler's home range size and detection probability (Simons et al., 2009). More details on this derivation are found in. We briefly note here that this prior includes the biologically-plausible range of values from σ suggested by the published literature.

The posterior distribution for N was highly skewed with a long right tail resulting in a wide 95% credible interval (Table 12.1). Nonetheless, the interval for density, D , includes estimates reported from more intensive field studies (?). This was true when considering both sets of priors, although posterior precision was higher under the informative set of priors. Specifically, the use of prior information reduced posterior density at high, biologically implausible, values of σ , and hence decreased the posterior mass for low values of N (Fig. 12.2).

In addition to estimating density, our model can be used to produce density surface maps, which are often used in applied ecological research to direct management efforts and develop hypotheses regarding the factors influencing abundance. Density surface maps can be produced by discretized the state-space and tallying the number of activity centers occurring in each pixel during each MCMC iteration. Parula density was highest near the northeastern corner of the study plot, which

Table 12.1. Posterior summary statistics for spatial Poisson-count model applied to the northern parula data. Two sets of priors were considered. $M = 300$ was used in both cases. Parulas/ha, D , is a derived parameter.

Par	Prior	Mean	SD	Mode	q0.025	q0.50	q0.975
σ	$U(0, \infty)$	2.154	1.222	1.230	0.896	1.665	5.170
λ_0	$U(0, \infty)$	0.284	0.149	0.212	0.084	0.256	0.665
N	$U(0, M)$	40.953	38.072	4.000	3.000	31.000	143.000
D	–	0.427	0.397	0.0417	0.0313	0.323	1.490
σ	$G(13, 10)$	1.301	0.258	1.230	0.889	1.266	1.908
λ_0	$U(0, \infty)$	0.298	0.132	0.240	0.098	0.279	0.603
N	$U(0, M)$	59.321	36.489	36.000	18.000	50.000	157.000
D	–	0.618	0.380	0.375	0.188	0.521	1.635

may correspond to important habitat features such as suitable nest site locations (Fig. 12.3). We anticipate future model extensions to directly model the point process intensity using habitat covariates.

12.7 ON (IM)PRECISION

12.8 HOW MUCH CORRELATION IS ENOUGH?

12.9 MUTANTS

12.9.1 Other observation models

12.9.2 Linear designs

12.10 SUMMARY

In this paper, we confronted one of the most difficult challenges faced in wildlife sampling — estimation of density in the absence of data to distinguish among individuals. To do so, we developed a novel class of spatially-explicit models that applies to spatially organized counts, where the count locations or devices are located sufficiently close together so that individuals are exposed to encounter at multiple devices. This design yields correlation in the observed counts, and this correlation proves to be informative about encounter probability parameters and hence density. We note that sample locations in count-based studies are typically *not* organized close together in space because conventional wisdom and standard practice dictate that independence of sample units is necessary (Hurlbert, 1984). Our model suggests that in some cases it might be advantageous to deviate from the conventional wisdom if one is interested in direct inference about density. Of course, this is also known in the application of standard spatial capture-recapture models (Borchers and Efford, 2008) where individual identity is preserved across trap encounters, but it is seldom, if ever, considered in the design of more traditional count surveys.

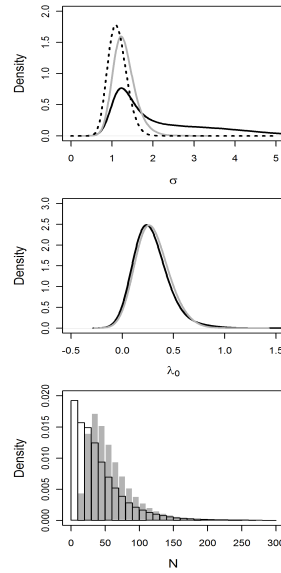


Figure 12.2. Effects of $\sigma \sim \text{Gamma}(13,10)$ prior on the posterior distributions from the northern parula model. Posteriors from model with uniform priors are shown in black, and posteriors from the informative prior model are shown in gray. The prior itself is shown as dotted line in the upper panel.

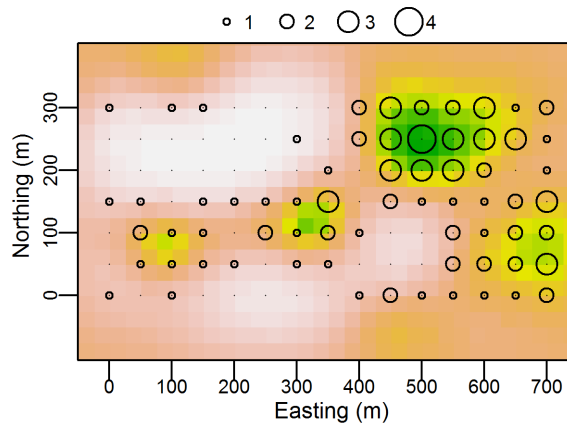


Figure 12.3. Estimated density surface of northern parula activity centers. The grid of point count locations with count totals is superimposed. See Fig. 1 for additional details.

Our model has broad relevance to an incredible number of animal sampling problems. Our motivating problem involved bird point counts where individual identity is typically not available. The model also applies to other standard methods used to sample unmarked populations, such as camera traps or even methods that yield sign (*e.g.* scat, track) counts indexed by space. However, results of our simulation study reveal some important limitations of the basic estimator applied to situations in which none of the individuals can be uniquely identified. In particular, posterior distributions are highly skewed in typical small to moderate sample size situations and posterior precision is low.

Several modifications of the model can lead to improved performance of the estimator. Our simulation results demonstrate that marking a subset of individuals can yield substantial increases in posterior precision. Marking a subset of individuals is commonplace in animal studies such as when a small number of individuals are radio-collared in conjunction with a count-based survey (Bartmann et al., 1987). In many other situations a subset of individuals can be identified by natural marks alone, and thus our model could be applied to data from camera-trapping studies of species such as mountain lions, deer, coyotes for which traditional SCR methods are not effective (Kelly et al., 2008). Thus, the ability to study partially-marked populations adds flexibility to existing SCR methods, and also creates new opportunities for designing efficient SCR studies since the costs of marking all individuals in a population can be prohibitive.

We note the existence of traditional approaches to combining data on marked and unmarked animals based on either the Lincoln-Peterson estimator or so-called “mark-resight” methods. (Bartmann et al., 1987; ?; ?). In their simplest form, mark-resight methods involve fitting standard closed-population mark-recapture models to the data on marked individuals, and the resultant estimate of detection probability (\hat{p}) is used to estimate population size as $\hat{N} = m + u/\hat{p}$ where m and u are the number of marked and unmarked individual, respectively. In this case, the unmarked individuals provide no information about the encounter rate parameters, and thus mark-resight methods cannot be used unless a large sample of marked individuals is available. This contrasts with our approach which can be used even when all individuals are unmarked.

In some cases, such as in point counts of birds, it may not be practical to mark individuals. An alternative to increasing posterior precision is to utilize prior information on home range size. Indeed, extensive information on home range size has been compiled for many species in diverse habitats (*e.g.*, DeGraaf and Yamasaki, 2001). It is easy to embody this information in a prior distribution as we demonstrated for the parula data.

An additional design extension that could increase precision is to use multiple sampling methods, in which one method generates encounter frequencies and the other method generates individuality. For example, camera traps are now commonly used with surveys for sign (scat or tracks), or hair snares for sampling bear populations. These distinct methods would have different basal detection rates

but share an underlying spatial model describing the organization of individuals in space. Our models show promise for using these disparate data types efficiently for estimating density.

12.10.1 Alternative Observation Models

Several aspects of our “spatial N -mixture model” can be modified to accommodate alternative sampling designs or parametric distributions. We considered situations where an individual can be detected more than once at a trap during a single occasion, but under some designs this is not possible. When collecting DNA samples, for instance, an individual can often be detected at most once during an occasion, because multiple samples of biological material cannot be attributed to distinct episodes. Therefore, rather than $z_{irt} \sim \text{Poisson}(\lambda_{ir})$ we have $z_{irt} \sim \text{Bernoulli}(p_{ir})$ where, for example, $p_{ir} = p_0 \exp(-d_{ir}^2/(2\sigma^2))$, and p_0 is the probability of detecting an individual whose home range is centered on trap r . This Bernoulli model is a focus of ongoing investigations.

Both the Poisson and the Bernoulli models produce count observations when aggregated over individuals to form trap-specific totals; however, ecologists often collect so-called “detection/non-detection” data because it can be easier to determine if “at least one” individual was present rather than enumerating all individuals in a location. In this case, the underlying z_{irt} array is the same as the above cases, but we observe $y_{rt} = I(\sum_{i=1}^N z_{irt} > 0)$ where I is the indicator function. This “Poisson-binary model” is a spatially explicit extension of the model of Royle and Nichols (2003) in which the underlying abundance state is inferred from binary data. We have investigated this model to a limited extent but do not report on those results here.

12.10.2 Spatial point process models

Our model has some direct linkages to existing point process models. We note that the observation intensity function (i.e., corresponding to the observation locations) is a compound Gaussian kernel similar to that of the Thomas process (??, pp. 61-62). Also, the Poisson-Gamma Convolution models (Wolpert and Ickstadt, 1998) are structurally similar (see also Higdon (1998) and Best et al. (2000)). In particular, our model is such a model but with a *constant* basal encounter rate λ_0 and *unknown* number and location of “support points”, which in our case are the animal activity centers, \mathbf{s}_i . We can thus regard our model as a model for *estimating* the location and local density of support points in such models, which we believe could be useful in the application of convolution models. Best et al. (2000) devise an MCMC algorithm for the Poisson-Gamma model based on data augmentation, which is similar to the component of our algorithm for updating the z variables in the conditional-on- z formulation of the model. We emphasize that our model

is distinct from these Poisson-Gamma models in that the number *and* location of such support points are estimated.

If individuals were perfectly observable then the resulting point process of locations is clearly a standard Poisson or Binomial (fixed N) cluster process or Neyman-Scott process. If detection is uniform over space but imperfect, then the basic process is unaffected by this random thinning. Our model can therefore be viewed formally as a Poisson (or Binomial) cluster process model but one in which the thinning is non-uniform, governed by the encounter model which dictates that thinning rate increases with distance from the observation points. In addition, our inference objective is, essentially, to estimate the number of parents in the underlying Poisson cluster process, where the observations are biased by an incomplete sampling apparatus (points in space).

As a model of a thinned point process, our model has much in common with classical distance sampling models (Buckland, 2001). The main distinction is that our data structure does *not* include observed distances, although the underlying observation model is fundamentally the same as in distance sampling if there is only a single replicate sample and \mathbf{s}_i is defined as an individual's location at an instant in time. For replicate samples, our model preserves (latent) individuality across samples and traps which is not a feature of distance sampling. We note that error in measurement of distance is not a relevant consideration in our model, and we explicitly do not require the standard distance sampling assumption that the probability of detection is 1 if an individual occurs at the survey point. More importantly, distance sampling models cannot be applied to data from many of the sampling designs for which our model is relevant. For example, many rare and endangered species can only be effectively surveyed using methods such as hair snares and camera traps that do not produce distance data (O'Connell et al., 2010).

12.11 CONCLUSION

Concerns about “statistical independence” have prompted ecologists to design count-based studies such that observed random variables can be regarded as *i.i.d.* outcomes (Hurlbert, 1984). Interestingly, this often proves impossible in practice, and elaborate methods have been devised to model spatial dependence as a nuisance parameter. Our paper presents a modeling framework that directly confronts this view by demonstrating that spatial correlation carries information about the locations of individuals, which can be used to estimate density even when individuals are unmarked and distance-related heterogeneity exists in encounter probability.

BIBLIOGRAPHY

- Alho, J. (1990), “Logistic regression in capture-recapture models,” *Biometrics*, 623–635.
- Arnason (1973), “Missing,” *Missing*, Missing, Missing.
- (1974), “Missing,” *Missing*, Missing, Missing.
- Baddeley, A. and Turner, R. (2005), “Spatstat: an R package for analyzing spatial point patterns,” *Journal of Statistical Software*, 12, 1–42, ISSN 1548-7660.
- Bales, S. L., Hellgren, E. C., Leslie Jr., D. M., and Hemphill Jr., J. (2005), “Dynamics of recolonizing populations of black bears in the Ouachita Mountains of Oklahoma,” *Wildlife Society Bulletin*, 1342–1351.
- Bartmann, R. M., White, G. C., Carpenter, L. H., and Garrott, R. A. (1987), “Aerial Mark-Recapture Estimates of Confined Mule Deer in Pinyon-Juniper Woodland,” *The Journal of Wildlife Management*, 51, 41–46.
- Berger, J. O., Liseo, B., and Wolpert, R. L. (1999), “Integrated likelihood methods for eliminating nuisance parameters,” *Statistical Science*, 1–22.
- Best, N. G., Ickstadt, K., and Wolpert, R. L. (2000), “Spatial Poisson Regression for Health and Exposure Data Measured at Disparate Resolutions,” *Journal of the American Statistical Association*, 95, 1076.
- Bibby, C. and Buckland, S. (1987), “Bias of bird census results due to detectability varying with habitat,” *Acta Ecologica*, 8, 103–112.
- Bivand, R. and Rundel, C. (2011), *rgeos: Interface to Geometry Engine - Open Source (GEOS)*, r package version 0.1-8.
- Borchers, D. and Efford, M. (2008), “Spatially explicit maximum likelihood methods for capture-recapture studies,” *Biometrics*, 64, 377–385.
- Borchers, D. L. (missing), “missing,” *Missing*, missing.
- Borchers, D. L., Buckland, S. T., and Zucchini, W. (2002), *Estimating animal abundance: closed populations*, vol. 13, Springer Verlag.
- Boulanger, J. and McLellan, B. (2001), “Closure violation in DNA-based mark-recapture estimation of grizzly bear populations,” *Canadian Journal of Zoology*, 79, 642–651.
- Brooks, S. P., Catchpole, E. A., and Morgan, B. J. T. (2000), “Bayesian Animal Survival Estimation,” *Statistical Science*, 15, 357–376.
- Buckland, S. T. (2001), *Introduction to distance sampling: estimating abundance of biological populations*, Oxford, UK: Oxford University Press.

- 6712 Burnham, K. P. and Anderson, D. R. (2002), *Model selection and multimodel in-*
6713 *ference: a practical information-theoretic approach*, Springer Verlag.
- 6714 Burnham, K. P. and Overton, W. S. (1978), “Estimation of the size of a closed
6715 population when capture probabilities vary among animals,” *Biometrika*, 65,
6716 625.
- 6717 Casella, G. and George, E. I. (1992), “Explaining the Gibbs sampler,” *American*
6718 *Statistician*, 46, 167–174.
- 6719 Chandler, R. and Royle, J. (2012), “Spatially-explicit models for inference about
6720 density in unmarked populations,” *Biometrics (in review)*.
- 6721 Chandler, R., Royle, J., and King, D. (2011), “Inference about density and tempo-
6722 rary emigration in unmarked populations,” *Ecology*, 92, 1429–1435.
- 6723 Converse, S. J. and Royle, J. A. (2010), “Missing,” *Missing*, Missing.
- 6724 Coull, B. A. and Agresti, A. (1999), “The Use of Mixed Logit Models to Reflect
6725 Heterogeneity in Capture-Recapture Studies,” *Biometrics*, 55, 294–301.
- 6726 Dawson, D. and Efford, M. (2009), “Bird population density estimated from acous-
6727 tic signals,” *Journal of Applied Ecology*, 46, 1201–1209.
- 6728 DeGraaf, R. M. and Yamasaki, M. (2001), *New England wildlife: habitat, natural*
6729 *history, and distribution*, University Press of New England.
- 6730 Dice, L. R. (1938), “Some census methods for mammals,” *Journal of Wildlife Man-*
6731 *agement*, 2, 119–130.
- 6732 Dorazio, R. and Royle, J. (2003), “Mixture models for estimating the size of a closed
6733 population when capture rates vary among individuals,” *Biometrics*, 351–364.
- 6734 Dorazio, R. M. (2007), “On the choice of statistical models for estimating occurrence
6735 and extinction from animal surveys,” *Ecology*, 88, 2773–2782.
- 6736 Durbin and Elston (2012), “Missing,” *Missing*, Missing.
- 6737 Efford, M. (2004), “Density estimation in live-trapping studies,” *Oikos*, 106, 598–
6738 610.
- 6739 — (2011), “secre-spatially explicit capture-recapture in R,” .
- 6740 Efford, M., Dawson, D., and Robbins, C. (2004), “DENSITY: software for analysing
6741 capture-recapture data from passive detector arrays,” *Animal Biodiversity and*
6742 *Conservation*, 217–228.
- 6743 Fienberg, S. E., Johnson, M. S., and Junker, B. W. (1999), “Classical multilevel and
6744 Bayesian approaches to population size estimation using multiple lists,” *Journal*
6745 *of the Royal Statistical Society of London A*, 163, 383–405.
- 6746 Gardner, B. (2009), “missing,” *Missing*, missing.
- 6747 Gardner, B., Royle, J., Wegan, M., Rainbolt, R., and Curtis, P. (2010), “Estimating
6748 black bear density using DNA data from hair snares,” *The Journal of Wildlife*
6749 *Management*, 74, 318–325.
- 6750 Gelfand, A. and Smith, A. (1990), “Sampling-based approaches to calculating
6751 marginal densities,” *Journal of the American statistical association*, 85, 398–409.
- 6752 Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004), *Bayesian data*
6753 *analysis, second edition.*, Boca Raton, Florida, USA: CRC/Chapman & Hall.
- 6754 Gelman, A., Meng, X. L., and Stern, H. (1996), “Posterior predictive assessment

- of model fitness via realized discrepancies,” *Statistica Sinica*, 6, 733–759.
- Gelman, A. and Rubin, D. B. (1992), “Inference from iterative simulation using multiple sequences,” *Statistical Science*, 7, 457–511.
- Geman, S. and Geman, D. (1984), “Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6, 721–741.
- Genz, A. S., Meyer, M. R., Lumley, T., and Maechler, M. (2007), “The adapt Package. R package version 1.0-4,” .
- Gilks, W. and Wild, P. (1992), “Adaptive rejection sampling for Gibbs sampling,” *Applied Statistics*, 41, 337–348.
- Gilks, W. R., Thomas, A., and Spiegelhalter, D. J. (1994), “A Language and Program for Complex Bayesian Modelling,” *Journal of the Royal Statistical Society. Series D (The Statistician)*, 43, 169–177, ArticleType: primary_article / Issue Title: Special Issue: Conference on Practical Bayesian Statistics, 1992 (3) / Full publication date: 1994 / Copyright 1994 Royal Statistical Society.
- Gopalaswamy (2012), “Missing,” *Missing*, missing.
- Gopalaswamy, A. M., Royle, A. J., Hines, J., Singh, P., Jathanna, D., Kumar, N. S., and Karanth, K. U. (2011), *A Program to Estimate Animal Abundance and Density using Spatially-Explicit Capture-Recapture*, r package version 1.0.4.
- Hahn, T., Bouvier, A., and Kieu, K. (2011), “Package ‘R2Cuba’ R package version 1.0-6,” .
- Hastings, W. (1970), “Monte Carlo sampling methods using Markov chains and their applications,” *Biometrika*, 57, 97–109.
- Hawkins, C. and Racey, P. (2005), “Low population density of a tropical forest carnivore, *Cryptoprocta ferox*: implications for protected area management,” *Oryx*, 39, 35–43.
- Hestbeck (1991), “Missing,” *Missing*, Missing, Missing.
- Higdon, D. (1998), “A process-convolution approach to modelling temperatures in the North Atlantic Ocean,” *Environmental and Ecological Statistics*, 5, 173–190.
- (2002), “Space and Space-Time Modeling using Process Convolutions,” in *Quantitative methods for current environmental issues*, eds. Anderson, C., Barnett, V., Chatwin, P., and El-Shaarawi, A., Springer Verlag, p. 37.
- Huggins, R. M. (1989), “On the statistical analysis of capture experiments,” *Biometrika*, 76, 133.
- Hurlbert, S. (1984), “Pseudoreplication and the design of ecological field experiments,” *Ecological monographs*, 54, 187–211.
- Illian (2008a), “Missing,” *Missing*, Missing.
- Illian, J. (2008b), *Statistical analysis and modelling of spatial point patterns*, Wiley-Interscience.
- Ivan, J. (2012), “Density, demography, and seasonal movements of snowshoe hares in central Colorado,” Ph.D. thesis, COLORADO STATE UNIVERSITY.
- Jackson, R., Roe, J., Wangchuk, R., and Hunter, D. (2006), “Estimating Snow Leopard Population Abundance Using Photography and Capture-Recapture

- Techniques,” *Wildlife Society Bulletin*, 34, 772–781.
- Johnson (2010), “A Model-Based Approach for Making Ecological Inference from Distance Sampling Data,” *Biometrics*, 66, 310318.
- Johnson, D. (1999), “The insignificance of statistical significance testing,” *The journal of wildlife management*, 763–772.
- Karanth, K. U. (1995), “Estimating tiger *Panthera tigris* populations from camera-trap data using capture–recapture models,” *Biological Conservation*, 71, 333–338.
- Kelly, M., Noss, A., Di Bitetti, M., Maffei, L., Arispe, R., Paviolo, A., De Angelo, C., and Di Blanco, Y. (2008), “Estimating puma densities from camera trapping across three study sites: Bolivia, Argentina, and Belize,” *Journal of Mammalogy*, 89, 408418.
- Kendall (1997), “Missing,” *Missing*, missing.
- Kéry, M. (2010), *Introduction to WinBUGS for Ecologists: Bayesian Approach to Regression, ANOVA, Mixed Models and Related Analyses*, Academic Press.
- Kéry, M., Gardner, B., Stoeckle, T., Weber, D., and Royle, J. A. (2010), “Use of Spatial Capture-Recapture Modeling and DNA Data to Estimate Densities of Elusive Animals,” *Conservation Biology*, 25, 356–364.
- Kéry, M., Royle, J., and Schmid, H. (2005), “Modeling avian abundance from replicated counts using binomial mixture models,” *Ecological Applications*, 15, 1450–1461.
- Kery, M. and Schaub, M. (2011), *Bayesian Population Analysis Using WinBugs*, Academic Press.
- King, R. (2009), “Missing,” *missing*, Missing.
- (missing), “Missing,” *missing*, Missing.
- Kumar (missing), “Unpublished data,” .
- Kuo, L. and Mallick, B. (1998), “Variable selection for regression models,” *Sankhyā*: *The Indian Journal of Statistics, Series B*, 65–81.
- Laird, N. M. and Ware, J. H. (1982), “Random-effects models for longitudinal data,” *Biometrics*, 963–974.
- Langtimm (2010), “Missing,” *Missing*, missing.
- Le Cam, L. (1990), “Maximum likelihood: an introduction,” *International Statistical Review/Revue Internationale de Statistique*, 153–171.
- Lewin-Koh, N. J., Bivand, R., contributions by Edzer J. Pebesma, Archer, E., Baddeley, A., Bibiko, H.-J., Dray, S., Forrest, D., Friendly, M., Giraudoux, P., Golicher, D., Rubio, V. G., Hausmann, P., Hufthammer, K. O., Jagger, T., Luque, S. P., MacQueen, D., Niccolai, A., Short, T., Stabler, B., and Turner, R. (2011), *maptools: Tools for reading and handling spatial objects*, r package version 0.8-10.
- Link, W. A. (2003), “Missing,” *missing*, missing.
- Link, W. A. and Barker, R. J. (2009), *Bayesian Inference: With Ecological Applications*, London, UK: Academic Press.
- Liu and Wu (1999), “Parameter expansion for data augmentation,” *J Am Stat Assoc*, 94, 1264–1274.

- MacEachern, S. and Berliner, L. (1994), "Subsampling the Gibbs sampler," *American Statistician*, 188–190.
- MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Royle, J. A., and Langtimm, C. A. (2002), "Estimating site occupancy rates when detection probabilities are less than one," *Ecology*, 83, 2248–2255.
- Mackenzie, D. I. and Royle, J. (2005), "Designing occupancy studies: general advice and allocating survey effort," *Journal of Applied Ecology*, 42, 1105–1114.
- Magoun, A. J., Long, C. D., Schwartz, M. K., Pilgrim, K. L., Lowell, R. E., and Valkenburg, P. (2011), "Integrating motion-detection cameras and hair snags for wolverine identification," *The Journal of Wildlife Management*, 75, 731–739.
- McCarthy, M. A. (2007), *Bayesian Methods for Ecology*, Cambridge: Cambridge University Press.
- McCullagh, P. and Nelder, J. (1989), *Generalized linear models*, Chapman & Hall/CRC.
- Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., Teller, E., et al. (1953), "Equation of state calculations by fast computing machines," *The journal of chemical physics*, 21, 1087–1092.
- Metropolis, N. and Ulam, S. (1949), "The Monte Carlo method," *Journal of the American Statistical Association*, 44, 335–341.
- Millar, R. (2009), "Comparison of hierarchical Bayesian models for overdispersed count data using DIC and Bayes' Factors," *Biometrics*, 65, 962–969.
- Mollet, P., Kery, M., Gardner, B., Pasinelli, G., and A, R. J. (2012), "Population size estimation for capercaillie (*Tetrao urogallus* L.) using DNA-based individual recognition and spatial capture-recapture models," *missing*, missing, missing.
- Neal, R. (2003), "Slice sampling," *Annals of Statistics*, 31, 705–741.
- Nelder, J. and Wedderburn, R. (1972), "Generalized linear models," *Journal of the Royal Statistical Society. Series A (General)*, 370–384.
- Norris III, J. L. and Pollock, K. H. (1996), "Nonparametric MLE under two closed capture-recapture models with heterogeneity," *Biometrics*, 639–649.
- O'Connell, A. F., Nichols, J. D., and Karanth, U. K. (2010), *Camera traps in animal ecology: Methods and analyses*, Springer.
- Overton, W. and Stehman, S. (1995), "The Horvitz-Thompson theorem as a unifying perspective for probability sampling: with examples from natural resource sampling," *American Statistician*, 261–268.
- Pebesma, E. and Bivand, R. (2011), *Package 'sp'*, r package version 0.9-91.
- Pledger, S. (2000), "Unified maximum likelihood estimates for closed capture-recapture models using mixtures," *Biometrics*, 434–442.
- Plummer, M. (2003), "JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling," in *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003)*. March, pp. 20–22.
- (2009), "rjags: Bayesian graphical models using mcmc. R package version 1.0.3-12," .
- (2011), *rjags: Bayesian graphical models using MCMC*, r package version 3-5.

- Plummer, M., Best, N., Cowles, K., and Vines, K. (2006), "CODA: Convergence Diagnosis and Output Analysis for MCMC," *R News*, 6, 7–11.
- R.E., R., Royle, J., Desimone, R., Schwartz, M., Edwards, V., Pilgrim, K., and McKelvey, K. (2012), "Estimating abundance of mountain lions from unstructured spatial samples," .
- Robert, C. P. and Casella, G. (2004), *Monte Carlo statistical methods*, New York, USA: Springer.
- (2010), *Introducing Monte Carlo Methods with R*, New York, USA: Springer.
- Roberts, G. O. and Rosenthal, J. S. (1998), "Optimal scaling of discrete approximations to Langevin diffusions," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60, 255–268.
- Royle, J. (2006), "Site occupancy models with heterogeneous detection probabilities," *Biometrics*, 62, 97–102.
- (2009), "Analysis of capture-recapture models with individual covariates using data augmentation," *Biometrics*, 65, 267–274.
- Royle, J. and Dorazio, R. (2006), "Hierarchical models of animal abundance and occurrence," *Journal of Agricultural, Biological, and Environmental Statistics*, 11, 249–263.
- (2008), *Hierarchical modeling and inference in ecology: the analysis of data from populations, metapopulations and communities*, Academic Press.
- (2010), "Missing," *Missing*, missing, missing.
- (2011), "Missing," *Missing*, missing, missing.
- Royle, J., Dorazio, R., and Link, W. (2007), "Analysis of multinomial models with unknown index using data augmentation," *Journal of Computational and Graphical Statistics*, 16, 67–85.
- Royle, J. and Dubovsky, J. (2001), "Modeling spatial variation in waterfowl band-recovery data," *The Journal of wildlife management*, 726–737.
- Royle, J., Karanth, K., Gopalaswamy, A., and Kumar, N. (2009), "Bayesian inference in camera trapping studies for a class of spatial capture-recapture models," *Ecology*, 90, 3233–3244.
- Royle, J. and Link, W. (2006), "Generalized site occupancy models allowing for false positive and false negative errors," *Ecology*, 87, 835–841.
- Royle, J. and Nichols, J. (2003), "Estimating abundance from repeated presence-absence data or point counts," *Ecology*, 84, 777–790.
- Royle, J. A. (2004), "*N*-mixture models for estimating population size from spatially replicated counts," *Biometrics*, 60, 108–115.
- (2008), "Modeling individual effects in the Cormack–Jolly–Seber model: a state-space formulation," *Biometrics*, 64, 364–370.
- (2010), "missing," *missing*, missing.
- Royle, J. A., Converse, S., and Link, W. (2011a), "Data augmentation for structured populations," *unpublished*.
- Royle, J. A., Dawson, D. K., and Bates, S. (2004), "Modeling abundance effects in distance sampling," *Ecology*, 85, 1591–1597.

- 6927 Royle, J. A., Kéry, M., and Guélat, J. (2011b), “Spatial capture-recapture models
6928 for search-encounter data,” *Methods in Ecology and Evolution*, 1–10.
- 6929 Royle, J. A., Magoun, A. J., Gardner, B., Valkenburg, P., and Lowell, R. E. (2011c),
6930 “Density estimation in a wolverine population using spatial capture–recapture
6931 models,” *The Journal of Wildlife Management*, 75, 604–611.
- 6932 Royle, J. A. and Young, K. V. (2008), “A Hierarchical Model For Spatial Capture-
6933 Recapture Data,” *Ecology*, 89, 2281–2289.
- 6934 Sanathanan, L. (1972), “Estimating the size of a multinomial population,” *The*
6935 *Annals of Mathematical Statistics*, 142–152.
- 6936 Schofield and Barker (missing), “Missing,” *Missing*, missing.
- 6937 Sepúlveda, M., Bartheld, J., Monsalve, R., Gómez, V., and Medina-Vogel, G.
6938 (2007), “Habitat use and spatial behaviour of the endangered Southern river ot-
6939 ter (*Lontra provocax*) in riparian habitats of Chile: conservation implications,”
6940 *Biological Conservation*, 140, 329–338.
- 6941 Sillett (2011), “Missing,” *Missing*, missing.
- 6942 Simons, T. R., Pollock, K. H., Wettroth, J. M., Alldredge, M. W., Pacifici, K., and
6943 Brewster, J. (2009), “Sources of Measurement Error, Misclassification Error, and
6944 Bias in Auditory Avian Point Count Data,” in *Modeling Demographic Processes*
6945 *In Marked Populations*, eds. Thomson, D. L., Cooch, E. G., and Conroy, M. J.,
6946 Boston, MA: Springer US, vol. 3, pp. 237–254.
- 6947 Spiegelhalter, D., Thomas, A., Best, N., and Lunn, D. (2003), *WinBUGS User*
6948 *Manual Version 1.4*.
- 6949 Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002),
6950 “Bayesian measures of model complexity and fit,” *Journal of the Royal Statistical*
6951 *Society. Series B, Statistical Methodology*, 583–639.
- 6952 Stabler, B. (2006), *shapefiles: Read and Write ESRI Shapefiles*, r package version
6953 0.6.
- 6954 Sturtz, S., Ligges, U., and Gelman, A. (2005), “R2WinBUGS: A Package for Run-
6955 ning WinBUGS from R,” *Journal of Statistical Software*, 12, 1–16.
- 6956 Su, Y.-S. and Yajima, M. (2011), *R2jags: A Package for Running jags from R*, r
6957 package version 0.02-17.
- 6958 Tanner, M. A. and Wong, W. H. (1987), “The calculation of posterior distributions
6959 by data augmentation,” *J Am Stat Assoc*, 82, 528–540.
- 6960 Thomas, A., O’Hara, B., Ligges, U., and Sturtz, S. (2006), “Making BUGS Open,”
6961 *R News*, 6, 12–17.
- 6962 Trolle, M. and Kéry, M. (2005), “Camera-trap study of ocelot and other secretive
6963 mammals in the northern Pantanal,” *Mammalia*, 69, 409–416.
- 6964 Tyre, A. J., Tenhumberg, B., Field, S. A., Niejalke, D., Parris, K., and Possing-
6965 ham, H. P. (2003), “Improving precision and reducing bias in biological surveys:
6966 estimating false-negative error rates,” *Ecological Applications*, 13, 1790–1801.
- 6967 Wegan, M. (2008), “Aversive conditioning, population estimation, and habitat pref-
6968 erence of black bears (*Ursus americanus*) on Fort Drum Military Installation in
6969 northern New York,” Ph.D. thesis, Cornell University, Jan.

- 6970 Wilson, K. R. and Anderson, D. R. (1985), “Evaluation of Two Density Estimators
6971 of Small Mammal Population Size,” *Journal of Mammalogy*, 66, 13–21.
- 6972 Wolpert, R. L. and Ickstadt, K. (1998), “Poisson/gamma random field models for
6973 spatial statistics,” *Biometrika*, 85, 251 –267.
- 6974 Yang, H. C. and Chao, A. (2005), “Modeling Animals’ Behavioral Response by
6975 Markov Chain Models for Capture–Recapture Experiments,” *Biometrics*, 61,
6976 1010–1017.
- 6977 Zuur, A., Ieno, E., Walker, N., Saveliev, A., and Smith, G. (2009), *Mixed effects*
6978 *models and extensions in ecology with R*, Springer Verlag.

INDEX

- 6979 bracket notation, 12

- 6980 R package
- 6981 sp, 147
- 6982 R package
- 6983 maptools, 147

- 6984 shapefile, 147

INDEX

- 6985 bracket notation, 12
- 6986 R package
- 6987 sp, 147
- 6988 R package
- 6989 maptools, 147
- 6990 shapefile, 147