

<sub>1</sub> Spatial Capture-Recapture Models

<sub>2</sub> The Four Horsemen (and women)

<sub>3</sub> January 5, 2012



# Contents

5	<b>1 Introduction</b>	<b>5</b>
6	<b>2 Introduction to Bayesian Analysis of GL(M)Ms Using R/WinBUGS</b>	<b>7</b>
7		
8	2.1 Notation . . . . .	8
9	2.2 GLMs and GLMMs . . . . .	9
10	2.3 Bayesian Analysis . . . . .	10
11	2.3.1 Bayes Rule . . . . .	11
12	2.3.2 Bayesian Inference . . . . .	12
13	2.3.3 Prior distributions . . . . .	14
14	2.3.4 Posterior Inference . . . . .	14
15	2.3.5 Small sample inference . . . . .	15
16	2.4 Characterizing posterior distributions by MCMC simulation . . .	16
17	2.5 What Goes on Under the MCMC Hood . . . . .	17
18	2.5.1 Rules for constructing full conditional distributions . . . .	18
19	2.5.2 Metropolis-Hastings algorithm . . . . .	19
20	2.6 Practical Bayesian Analysis and MCMC . . . . .	20
21	2.6.1 Choice of prior distributions . . . . .	20
22	2.6.2 Convergence and so-forth . . . . .	21
23	2.6.3 Bayesian confidence intervals . . . . .	22
24	2.6.4 Estimating functions of parameters . . . . .	22
25	2.7 Bayesian Analysis using WinBUGS . . . . .	23
26	2.7.1 Linear Regression in WinBUGS . . . . .	23
27	2.7.2 Inference about functions of model parameters . . . . .	26
28	2.8 Model Checking and Selection . . . . .	26
29	2.8.1 Goodness-of-fit . . . . .	26
30	2.8.2 Model Selection . . . . .	28
31	2.9 Poisson GLMs . . . . .	29
32	2.9.1 Important properties of the Poisson distribution . . . . .	29
33	2.9.2 Example: Breeding Bird Survey Data . . . . .	30
34	2.9.3 Doing it in WinBUGS . . . . .	31
35	2.9.4 Constructing your own MCMC algorithm . . . . .	33
36	2.10 Poisson GLM with Random Effects . . . . .	38
37	2.11 Binomial GLMs . . . . .	40

38	2.11.1	Binomial regression . . . . .	40
39	2.11.2	Example: Waterfowl Banding Data . . . . .	41
40	2.12	Summary and Outlook . . . . .	43
41	<b>3</b>	<b>Closed population models</b>	<b>45</b>
42	<b>4</b>	<b>Estimating the Size of a Closed Population</b>	<b>47</b>
43	4.1	The Simplest Closed Population Model: Model M0 . . . . .	48
44	4.1.1	The Spatial Context of Capture-Recapture . . . . .	50
45	4.1.2	Conditional likelihood . . . . .	50
46	4.2	Data Augmentation . . . . .	51
47	4.2.1	DA links occupancy models and closed population models	51
48	4.2.2	Model $M_0$ in BUGS . . . . .	53
49	4.2.3	Formal development of data augmentation . . . . .	55
50	4.2.4	Remarks on Data Augmentation . . . . .	56
51	4.2.5	Example: Black Bear Study on Fort Drum . . . . .	57
52	4.3	Temporally varying and behavioral effects . . . . .	61
53	4.4	Models with individual heterogeneity . . . . .	61
54	4.4.1	Analysis of Model Mh . . . . .	64
55	4.4.2	Analysis of the Fort Drum data . . . . .	65
56	4.4.3	Building your own MCMC algorithm . . . . .	68
57	4.4.4	Exercises related to model Mh . . . . .	70
58	4.5	Individual Covariate Models: Toward Spatial Capture-Recapture	71
59	4.5.1	Example: Location of capture as a covariate. . . . .	72
60	4.5.2	Extension of the Model . . . . .	75
61	4.5.3	Invariance of density to maxD . . . . .	78
62	4.5.4	Toward Fully Spatial Capture-recapture Models . . . . .	79
63	4.6	DISTANCE SAMPLING: A primitive Spatial Capture-Recapture	
64		Model . . . . .	79
65	4.6.1	Example: Muntjac deer survey from Nagarahole, India . . . . .	81
66	4.7	Summary and Outlook . . . . .	83
67	<b>5</b>	<b>Fully Spatial Capture-Recapture Models</b>	<b>85</b>
68	<b>6</b>	<b>Fully Spatial Capture-Recapture Models</b>	<b>87</b>
69	6.1	Sampling Design and Data Structure . . . . .	88
70	6.2	The binomial observation model . . . . .	89
71	6.2.1	Distance as a latent variable . . . . .	90
72	6.3	The Binomial Point-process Model . . . . .	91
73	6.3.1	Definition of home range center . . . . .	92
74	6.3.2	The state-space of the point process . . . . .	93
75	6.3.3	Invariance and the State-space as a model assumption . . . . .	94
76	6.3.4	Connection to Model Mh . . . . .	94
77	6.3.5	Connection to Distance Sampling . . . . .	96
78	6.4	Simulating SCR Data . . . . .	96
79	6.4.1	Formatting and manipulating real data sets . . . . .	97

80	6.5	Fitting an SCR Model in BUGS . . . . .	98
81	6.6	Unknown N . . . . .	101
82	6.6.1	Analysis using data augmentation in WinBUGS . . . . .	102
83	6.6.2	Use of other BUGS engines: JAGS . . . . .	104
84	6.7	Case Study: Wolverine Camera Trapping Study . . . . .	105
85	6.7.1	Fitting the model in WinBUGS . . . . .	108
86	6.7.2	Conclusion of Analysis . . . . .	111
87	6.8	Constructing Density Maps . . . . .	112
88	6.8.1	Example: Wolverine density map. . . . .	114
89	6.9	Discrete State-Space . . . . .	114
90	6.9.1	Evaluation of Coarseness of Discrete Approximation . . . . .	116
91	6.9.2	Analysis of the wolverine camera trapping data . . . . .	118
92	6.9.3	SCR models as multi-state models . . . . .	120
93	6.10	Summary and Outlook . . . . .	120
94	<b>7</b>	<b>Other observation models</b>	<b>123</b>
95	<b>8</b>	<b>Maximum likelihood estimation</b>	<b>125</b>
96	<b>9</b>	<b>Likelihood Analysis of SCR Models</b>	<b>127</b>
97	9.1	Likelihood analysis . . . . .	128
98	9.1.1	Implementation (simulated data) . . . . .	129
99	9.2	MLE when N is Unknown . . . . .	133
100	9.2.1	Exercises . . . . .	136
101	9.2.2	Integrated Likelihood using the model under data aug- mentation . . . . .	136
102	9.2.3	Extensions . . . . .	137
103	9.3	Classical model selection and assessment . . . . .	137
104	9.4	Likelihood analysis of the wolverine camera trapping data . . . . .	138
105	9.4.1	Exercises . . . . .	140
106	9.5	Program DENSITY and the R package <b>secr</b> . . . . .	140
107	9.5.1	Analysis using the <b>secr</b> package . . . . .	142
108	9.6	Summary and Outlook . . . . .	144
109			
110	<b>10</b>	<b>MCMC details</b>	<b>147</b>
111	<b>11</b>	<b>MCMC Details</b>	<b>149</b>
112	11.1	Introduction . . . . .	149
113	11.1.1	Why build your own MCMC algorithm? . . . . .	149
114	11.2	MCMC and posterior distributions . . . . .	150
115	11.3	Types of MCMC sampling . . . . .	152
116	11.3.1	Gibbs sampling . . . . .	153
117	11.3.2	Metropolis-Hastings sampling . . . . .	156
118	11.3.3	Metropolis-within-Gibbs . . . . .	159
119	11.4	GLMMs Poisson regression with a random effect . . . . .	160
120	11.4.1	Rejection sampling and slice sampling . . . . .	163

121	11.5 MCMC for closed capture-recapture Model Mh . . . . .	164
122	11.6 MCMC algorithm for the basic spatial capture-recapture model .	164
123	11.6.1 SCR model with binomial encounter process . . . . .	167
124	11.6.2 Looking at model output . . . . .	168
125	11.6.3 Posterior density plots . . . . .	171
126	11.6.4 Serial autocorrelation and effective sample size . . . . .	172
127	11.6.5 Summary results . . . . .	173
128	11.6.6 Other useful commands . . . . .	174
129	11.7 Manipulating the state-space . . . . .	175
130	11.8 MCMC software packages . . . . .	177
131	11.8.1 WinBUGS . . . . .	178
132	11.8.2 OpenBUGS . . . . .	178
133	11.8.3 JAGS Just Another Gibbs Sampler . . . . .	178
134	11.9 Summary and Outlook . . . . .	179
135	<b>12 Goodness of Fit and stuff</b>	<b>181</b>
136	<b>13 Covariate models</b>	<b>183</b>
137	<b>14 State-space Covariates</b>	<b>185</b>
138	14.1 Homogeneous point process revisited . . . . .	186
139	14.2 Inhomogeneous binomial point process . . . . .	188
140	14.3 Examples . . . . .	189
141	14.3.1 Simulation and analysis of inhomogeneous point processes	189
142	14.3.2 Fitting inhomogeneous point process SCR models . . . .	192
143	14.3.3 The jaguar data . . . . .	195
144	14.4 Summary . . . . .	196
145	14.5 Other ideas . . . . .	197
146	<b>15 Inhomogeneous Point Process</b>	<b>199</b>
147	<b>16 Open models</b>	<b>201</b>

148 **Chapter 1**

149 **Introduction**





## Chapter 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  $\mu$  and  $\sigma^2$ . 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  $\psi$ . 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  $\psi$ . 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  $\beta$  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  $y$ :  $\text{Var}(y) = V(\mu)$  (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 parameters<sup>1</sup> 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  $\psi$  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 probability that it is present given that

---

<sup>1</sup>To hear this will be shocking to some readers perhaps.

326 it is not detected in  $T$  samples. In words, the expression we seek is:

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

327 Mathematically, this is

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

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

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

### 337 2.3.2 Bayesian Inference

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

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

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

356 Asserting the general relevance of Bayes rule to all statistical problems, we  
 357 can conclude that the two main features of Bayesian inference are that: (1)  
 358 "parameters"  $\theta$  are regarded as realizations of a random variable and, as a  
 359 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  $\theta$  and therefore we can make direct probability statements to characterize uncertainty about  $\theta$ .

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 like<sup>2</sup> 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  $\theta$  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  $\theta$ , 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 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

---

<sup>2</sup>Characterization from Sims REF XYZ

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).

XXXXXXXX

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

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

XXXXXXXX

### 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  $\theta$  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 worse<sup>3</sup>, “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, have to devise various approximations to what they were trying to characterize. The advent of Markov chain Monte Carlo (MCMC) methods has made it easier to calculate posterior distributions for just about any problem to arbitrary levels of precision.

Broadly speaking, MCMC is a class of methods for drawing random numbers (sampling or simulating) from the target posterior distribution. Thus, even though we might not recognize the posterior as a named distribution or be able to analyze its features analytically, e.g., devise mathematical expressions for the

---

<sup>3</sup>Actual quote from a referee

mean and variance, we can use these MCMC methods to obtain a large sample from the posterior and then use that sample to characterize features of the posterior. What we do with the sample depends on our intentions – typically we obtain the mean or median for use as a point estimate, and take a confidence interval based on Monte Carlo estimates of the quantiles. These are estimates, but not like frequentist estimates. Rather, they are Monte Carlo estimates with an associated Monte Carlo error which is largely determined arbitrarily by the analyst. They are not estimates qualified by a sampling distribution as in classical statistics. If we run our MCMC long enough then our reported value of  $E[\theta|y]$  or any feature of the posterior distribution is precisely what we say it is. There is no “sampling variation” in the frequentist sense of the word. In summary, the MCMC samples provide a Monte Carlo characterization of *the* posterior distribution.

## 2.5 What Goes on Under the MCMC Hood

We will develop and apply MCMC methods in some detail for spatial capture-recapture models in chapter 10. Here we provide a simple illustration of some basic ideas related to the practice of MCMC.

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

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

and

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

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

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

Similarly,

$$[\beta|y, \alpha] is normal(.....)$$

The MCMC algorithm for this model has us simulate in succession, repeatedly, from those two distributions. See Gelman et al. (2004) for more examples of Gibbs sampling for the normal model. A conceptual representation of the MCMC algorithm for this simple model is therefore: XXXX Check out ALGORITHM environment XXXXX

567     **Algorithm**

```

568
569         0. Initialize  $\alpha$  and  $\beta$ 
570
571         Repeat{
572             1. Draw a new value of  $\alpha$  from Eq. \ref{xyz}
573
574             2. Draw a new value of  $\beta$  from Eq. \ref{xyz}
575         }
```

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

577         We will talk about the Metropolis-Hastings algorithm shortly, and we will use it extensively in the analysis of SCR models (e.g., chapter 10).

### 594     **2.5.1 Rules for constructing full conditional distributions**

595         The basic strategy for constructing full-conditional distributions for devising MCMC algorithms can be reduced conceptually to a couple of basic steps summarized as 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 the product of all components;
- 600         (step 3) Remove the ones that don’t have the focal parameter in them.
- 601         (step 4) Do some algebra on the result in order to identify the resulting pdf or pmf.

603         Of the 4 steps, the last of those is the main step that requires quite a bit of statistical experience and intuition because various algebraic tricks can be used to reshape the mess into something noticeable - i.e., a standard, named distribution.

606 But step 4 is not necessary if we decide instead to use the Metropolis-Hastings  
 607 algorithm as described below.

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

## 615 2.5.2 Metropolis-Hastings algorithm

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

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

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

631 which we call the MH acceptance probability. This ratio can sometimes be  $> 1$  in  
 632 which case we set it equal to 1. It is useful to note that  $h()$  can be anything at all.  
 633 Absolutely anything! You can generate candidate values from a *normal*(0,1)  
 634 distribution, from a *uniform*(-3455,3455) distribution, or anything of proper  
 635 support. Note, however, that good choices of  $h()$  are those that approximate  
 636 the posterior distribution. Obviously if  $h() = f(\theta|y)$  (i.e., the posterior) then  
 637 you always accept the draw, and it stands to reason that proposals that are  
 638 more similar to  $f(\theta|y)$  will lead to higher acceptance probabilities. No matter  
 639 the choice of  $h()$ , we can evaluate this ratio numerically because the marginal  
 640  $f(y)$  cancels from both the numerator and denominator, which is the magic of  
 641 the MH algorithm.

642 A special kind of  $h()$  are those that are symmetric, which means that  
 643  $h(a|b) = h(b|a)$  in which case  $h(a|b)$  and  $h(b|a)$  just cancel out from the MH ac-  
 644 ceptance probability and  $r$  is then just the ratio of the target density evaluated  
 645 at the candidate value to that evaluated at the current value. A type of sym-  
 646 metric proposal useful in many situations is the so-called *random-walk* proposal  
 647 distribution where candidate values are drawn from a normal distribution with

mean equal to the current value and some standard deviation, say  $\delta$ , which is prescribed by the user. For parameters that have support on the real line, e.g.,  $\alpha$  in our example above, the random walk proposal generator has us generate  $\alpha^* \sim \text{Normal}(\alpha^t, \delta)$ . If we set  $\delta$  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  $\delta$  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”,  $\tau$ , 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

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

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

## 695 2.6.2 Convergence and so-forth

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

703 The quick diagnostic to whether convergence has been achieved is that your  
 704 Markov chains look “grassy” – see Fig. 2.5 below. Another way to check  
 705 convergence is to update the parameters some more and see if the posterior  
 706 changes. It is good to confirm convergence using the “R-hat” statistic ( $\hat{R}$ ) or  
 707 Brooks-Gelman-Rubin statistic (Gelman et al., 1996) which should be close  
 708 to 1 if the Markov chains have converged and sufficient posterior samples have  
 709 been obtained. In practice,  $\hat{R} = 1.2$  is probably good enough for some problems.  
 710 For some models you can’t actually realize a low  $\hat{R}$ . E.g., if the posterior is a  
 711 discrete mixture of distributions then you can be misled into thinking that your  
 712 Markov chains have not converged when in fact the chains are just jumping back  
 713 and forth in the posterior state-space. So, for example, using model selection  
 714 methods (section XYZ) sometimes suggests non-convergence. Another situation  
 715 is when one of the parameters is on the boundary of the parameter space which  
 716 might appear to be very poor mixing, but all within some extreme region of  
 717 the parameter space.<sup>4</sup> This kind of stuff is normally ok and you need to think  
 718 really hard about the context of the model and the problem before you conclude  
 719 that your MCMC algorithm is ill-behaved.

720 Some models exhibit “poor mixing” of the Markov chains or what people  
 721 might also say “have not covered” (or “slow convergence”) which is a term  
 722 we would disagree with because the samples might well be from the posterior  
 723 (i.e., the Markov chains have converged to the proper stationary distribution)  
 724 but simply mix around the posterior rather slowly. Anyway, poor mixing can  
 725 happen for a huge number of reasons – when parameters are highly correlated  
 726 (even confounded), or barely identified from the data, or the algorithms are

---

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

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  $m^{\text{th}}$  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  $\theta$  is the parameter of interest and let  $\theta^{(i)}$  for  $i = 1, 2, \dots, M$  be the posterior samples of  $\theta$ . Let  $\eta = \exp(\theta)$ , then a posterior sample of  $\eta$  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 5).

## 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 Kéry 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:

```
> x<-rnorm(10)
> mu<- -3.2+ 1.5*x
> 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`:

```

808 > cat("
809 model {
810   for (i in 1:10){
811     y[i]~dnorm(mu[i],tau)           # the "likelihood"
812     mu[i]<- beta0 + beta1*x[i]      # the linear predictor
813   }
814   beta0~dnorm(0,.01)               # prior distributions
815   beta1~dnorm(0,.01)
816   sigma~dunif(0,100)
817   tau<-1/(sigma*sigma)             # tau is a derived parameter
818 }
819 ",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,  $\beta_0$  and  $\beta_1$  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  $\tau$ . **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  $\beta_0$ ,  $\beta_1$ ,  $\sigma$  and  $\tau$ . **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:

```

848 library("R2WinBUGS")    # "attach" the R2WinBUGS library
849 data <- list ( "y","x")
850 inits <- function()
851   list ( beta1=rnorm(1),beta0=rnorm(1),sigma=runif(1,0,2) )
852 parameters <- c("beta0","beta1","sigma","tau")
853 out<-bugs (data, inits, parameters, "normal.txt", n.thin=2, n.chains=2,

```

```
854      n.burnin=2000, n.iter=6000, debug=TRUE,working.dir=getwd())
```

855     **Remarks:** A common question is “how should my data be formatted?”  
 856 That depends on how you describe the model in the **BUGS** language, how  
 857 your data are input into **R** and subsequently formatted. There is no unique  
 858 way to describe any particular model and so you have some flexibility. We  
 859 talk about data format further in the context of capture-recapture models and  
 860 SCR models in chapter 5 and elsewhere. In general, starting values are optional  
 861 but we recommend to always provide reasonable starting values for structural  
 862 parameters, but are not always necessary for random effects. Note that the pre-  
 863 viously created objects defining data, initial values and parameters to monitor  
 864 are passed to the function `bugs()`. In addition, various other things are de-  
 865 clared: The number of Markov chains (`n.chains`), the thinning rate (`n.thin`),  
 866 the number of burn-in iterations (`n.burnin`) and the total number of iterations  
 867 (`n.iter`). To develop a detailed understanding of the various parameters and  
 868 settings used for MCMC, consult a basic reference such as Kéry (2010).

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

```
874 > print(out,digits=2)
875 Inference for Bugs model at "normal.txt", fit using WinBUGS,
876 2 chains, each with 6000 iterations (first 2000 discarded), n.thin = 2
877 n.sims = 4000 iterations saved
878      mean   sd  2.5%  25%   50%   75%  97.5% Rhat n.eff
879 beta0  -2.43 1.84 -6.21 -3.50 -2.42 -1.34  1.27   1  4000
880 beta1   2.62 1.54 -0.42  1.68  2.62  3.57  5.67   1  4000
881 sigma   5.29 1.66  3.11  4.14  4.95  6.05  9.39   1  4000
882 tau     0.05 0.02  0.01  0.03  0.04  0.06  0.10   1  4000
883 deviance 59.85 3.24 56.18 57.47 59.00 61.37 68.32   1   840
884
885 For each parameter, n.eff is a crude measure of effective sample size,
886 and Rhat is the potential scale reduction factor (at convergence, Rhat=1).
887
888 DIC info (using the rule, pD = Dbar-Dhat)
889 pD = 2.6 and DIC = 62.4
```

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

## 896 2.7.2 Inference about functions of model parameters

897 Using the MCMC draws for a given model we can easily obtain the posterior  
 898 distribution of any function of model parameters. We showed this in the above  
 899 example by providing the posterior of  $\tau$  when the model was parameterized in  
 900 terms of standar deviation  $\sigma$ . As another example, suppose that the normal  
 901 regression model above had a quadratic response function of the form

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

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

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

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

## 909 2.8 Model Checking and Selection

910 In general terms model checking - or assessing the adequacy of the model - and  
 911 model selection are quite thorny issues and, despite contrary and, sometimes,  
 912 strongly held belief among practitioners, there are not really definitive, general  
 913 solutions to either problem. We're against dogma on these issues and think  
 914 people need to be open-minded about such things and recognize that models  
 915 can be useful whether or not they pass certain statistical tests. Some models  
 916 are intrinsically better than others because they make more biological sense  
 917 or foster understanding or achieve some objective that some bootstrap or other  
 918 goodness-of-fit test can't decide for you. That said, it gives you some confidence  
 919 if your model seems adequate and we try to provide some fit assessment in most  
 920 real applications of SCR models We provide a very brief overview of concepts  
 921 here, but provide more detailed coverage in chapter 12. See also Kéry (2010,  
 922 ch. xyz) and Link and Barker (2009, ch. xyz) for specific context related to  
 923 Bayesian model checking and selection.

### 924 2.8.1 Goodness-of-fit

925 Goodness-of-fit testing is an important element of any analysis because our  
 926 model represents a general set of hypotheses about the ecological and obser-  
 927 vation processes that generated our data. Thus, if our model "fits" in some  
 928 statistical or scientific sense, then we believe it to be consistent with the hy-  
 929 potheses that went into the model. More formally, we would conclude that  
 930 the data are *not inconsistent* with the hypotheses, or that the model appears  
 931 adequate. If we have enough data, then of course we will reject any set of sta-  
 932 tistical 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 12.

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”,

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

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  $\Pr(\text{Fit} > \text{Fit}_{new})^5$  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 statistic<sup>6</sup>, in which

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

---

<sup>5</sup>Check this definition!

<sup>6</sup>Ref for this?

(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  $\text{Pr}(w = 1)$  indicate stronger evidence to support that " $x$  is in the model" whereas values of  $\text{Pr}(w = 1)$  close to 0 suggest that  $x$  is less important.

This idea seems to be due to Kuo and Mallick (1998)<sup>7</sup> and see Royle and Dorazio (2008, ch. XXXX) for an example in the context of logistic regression. This 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  $\beta$  is sampled from the prior distribution, and the prior distribution is very diffuse, then extreme values of  $\beta$  are likely. Consequently, when the current value of  $\beta$  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 XXXXX<sup>8</sup>) 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 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

<sup>7</sup> Is this also what people call Zellner's G-priors?

<sup>8</sup>see Royle 2008 paper for reference

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 11), 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  $\lambda$  which we write

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

Commonly  $y_i$  is a count of animals or plants at some point in space and  $\lambda$  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  $\lambda_1$  and  $\lambda_2$ , 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 file<sup>9</sup> 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 studies, perhaps with more columns representing covariates. To read in the data and display the first few elements of this matrix, do this:

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

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

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

---

<sup>9</sup>check this data format



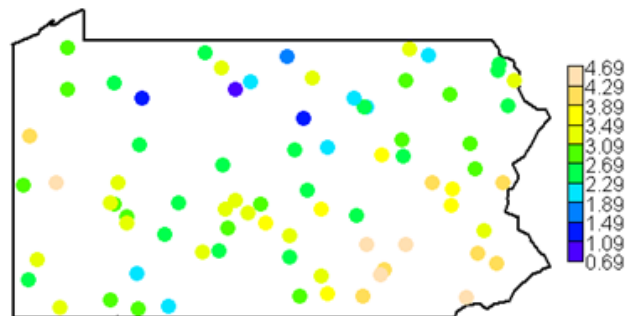


Figure 2.1: Needs a caption

1086 We can ponder the potential effects that might lead to dove counts being  
 1087 high....corn fields, telephone wires, barn roofs along with misidentification of  
 1088 pigeons, these could all correlated reasonably well with the observed count of  
 1089 mourning doves. Unfortunately we don't have any of that information.

1090 We do have a measure of forest cover in the vicinity of each point which is  
 1091 contained in the data set (variable "habitat"). This was derived from a larger  
 1092 GIS coverage of the state (provided in the data file "pahabdata.csv") which  
 1093 can be plotted using the `spatial.plot` function using the following commands

```
1094 > map('state',regions="penn",lwd=2)
1095 > spatial.plot(pahabdata[,2:3],pahabdata[, "dfor"],cx=2)
1096 > map('state',regions="penn",lwd=2,add=TRUE)
```

1097 where the result appears in Fig. 2.2. We see a prominent pattern that  
 1098 indicates high forest coverage in the central part of the state and low forest cover  
 1099 in the SE. Inspecting the previous figure of log-counts suggests a relationship  
 1100 between counts and forest cover which is perhaps not surprising.

### 1101 2.9.3 Doing it in WinBUGS

1102 Here we demonstrate how to fit a Poisson GLM in **WinBUGS** using the co-  
 1103 variate  $x_i$  = forest cover. It is advisable that  $x_i$  be standardized in most cases  
 1104 as this will improve mixing of the Markov chains. Recall that the data we have  
 1105 stored include a standardized covariate (forest cover) and so we don't have to  
 1106 worry about that here. To read the BBS data into **R** and get things set up for  
 1107 **WinBUGS** we issue the following commands:

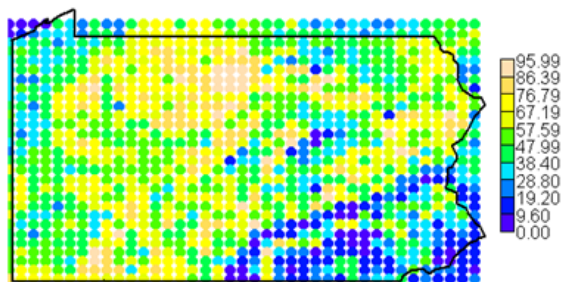


Figure 2.2: Needs a caption

```

1108 data<-read.csv("pa-bbsdovedata-all.csv")
1109 y<-data[,29]                                # pick out 1990
1110 notna<-!is.na(y)
1111 y<-y[notna]                                # discard missing
1112 habitat<-data[notna,4]                     # get habitat data
1113 library("R2WinBUGS")                       # load R2WinBUGS
1114 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**:

```

1118 cat("
1119 model {
1120     for (i in 1:M){
1121         y[i]~dpois(lam[i])
1122         log(lam[i])<- beta0+beta1*habitat[i]
1123     }
1124     beta0~dunif(-5,5)
1125     beta1~dunif(-5,5)
1126 }
1127 ",file="PoissonGLM.txt")
1128
1129 inits <- function() list ( beta0=rnorm(1),beta1=rnorm(1))
1130 parameters <- c("beta0","beta1")
1131 out<-bugs (data, inits, parameters, "PoissonGLM.txt", n.thin=2,n.chains=2,
1132           n.burnin=2000,n.iter=6000,debug=TRUE,working.dir=getwd())

```

1133 **Remarks:** (1) Note the close correspondence in how the model is specified  
 1134 here compared with the normal regression model previously. As an exercise you  
 1135 should discuss the specific differences between the **BUGS** model specifications  
 1136 for the normal and Poisson models.

```
1137 > print(out,digits=3)
1138 Inference for Bugs model at
1139 ‘‘PoissonGLM.txt’’, fit using WinBUGS,
1140 2 chains, each with 4000 iterations (first 1000 discarded), n.thin = 2
1141 n.sims = 3000 iterations saved
```

	mean	sd	2.5%	25%	50%	75%	97.5%	Rhat	n.eff
1142 beta0	3.151	0.025	3.102	3.135	3.151	3.168	3.199	1.001	2300
1143 beta1	-0.498	0.021	-0.539	-0.512	-0.498	-0.484	-0.457	1.001	3000
1144 fit	869.930	19.856	835.500	855.700	868.600	881.900	913.602	1.002	1600
1145 fitnew	76.709	12.519	54.098	68.107	76.215	84.510	102.602	1.001	3000
1146 deviance	1116.605	2.014	1115.000	1115.000	1116.000	1117.000	1122.000	1.001	3000

1148 We might wonder whether this model provides an adequate fit to our data.  
 1149 To evaluate that, we used a Bayesian p-value analysis with fit statistic based  
 1150 on the Freeman-Tukey residual by replacing the model specification above with  
 1151 this:

```
1152 cat("
1153 model {
1154   for (i in 1:M){
1155     y[i]~dpois(lam[i])
1156     log(lam[i])<- beta0+beta1*habitat[i]
1157     d[i]<- pow(pow(y[i],0.5)-pow(lam[i],0.5),2)   #
1158
1159     ynew[i]~dpois(lam[i])
1160     dnew[i]<-pow( pow(ynew[i],0.5)-pow(lam[i],0.5),2)
1161
1162   }
1163   fit<-sum(d[])
1164   fitnew<-sum(dnew[])
1165   beta0~dunif(-5,5)
1166   beta1~dunif(-5,5)
1167 }
1168 ",file="PoissonGLM.txt")
```

1169 The Bayesian p-value is the proportion of times *fitnew* > *fit* which, for this  
 1170 data set, is 0, which was 1.0 in this case (calculation omitted). This suggests  
 1171 that the basic Poisson model does not fit well.

## 1172 2.9.4 Constructing your own MCMC algorithm

1173 It might be helpful to suffer through a couple examples building custom MCMC  
 1174 algorithms. Here, we develop an MCMC algorithm for the Poisson regression  
 1175 model, using a Metropolis-within-Gibbs sampling framework.

1176 We will assume that the two parameters have diffuse normal priors, say  
 1177  $[\alpha] = \text{Norm}(0, 100)$  and  $[\beta] = \text{Norm}(0, 100)$  where each has *standard deviation*

1178 100 (recall that **WinBUGS** parameterizes the normal in terms of  $1/\sigma^2$ ). We  
 1179 need to assemble the relevant elements of the model which are these two prior  
 1180 distributions and the likelihood  $[\mathbf{y}|\alpha, \beta] = \prod_i [y_i|\alpha, \beta]$  which is, mathematically,  
 1181 the product of the Poisson pmf evaluated at each  $y_i$ , given particular values of  
 1182  $\alpha$  and  $\beta$ . Next, we need to identify the full conditionals  $[\alpha|\beta, \mathbf{y}]$  and  $[\beta|\alpha, \mathbf{y}]$ .  
 1183 We use the all-purpose rule for constructing full conditionals (section 2.5.1) to  
 1184 discover that:

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

1185 and

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

1186 Remember, we could replace the “ $\propto$ ” with “ $=$ ” if we put  $[y|\beta]$  or  $[y|\alpha]$  in the  
 1187 denominator. But, in general,  $[y|\alpha]$  or  $[y|\beta]$  will be quite a pain to compute and,  
 1188 more importantly, it is a constant as far as the operative parameters ( $\alpha$  or  $\beta$ ,  
 1189 respectively) are concerned. Therefore, the MH acceptance probability will be  
 1190 the ratio of the full-conditional evaluated at a candidate draw to that evaluated  
 1191 at the current draw, and so the denominator required to change  $\propto$  to  $=$  winds  
 1192 up canceling from the MH acceptance probability. Here we will use the random  
 1193 walk candidate generator so that, for example,  $\alpha^* \sim \text{Normal}(\alpha^t, \delta)$  where  $\delta$   
 1194 is the standard-deviation of the proposal distribution, which is just a tuning  
 1195 parameter<sup>10</sup>. We remark also that calculations are often done on the log-scale  
 1196 to preserve numerical integrity of things when quantities evaluate to small or  
 1197 large numbers, so keep in mind, for example,  $a * b = \exp(\log(a) + \log(b))$ . The  
 1198 “Metropolis within Gibbs” algorithm for a Poisson regression turns out to be  
 1199 remarkably simple:

```
1200 set.seed(2013)
1201
1202 out<-matrix(NA,nrow=1000,ncol=2)    # matrix to store the output
1203 alpha<- -1                          # starting values
1204 beta <- -.8
1205
1206 # begin the MCMC loop ; do 1000 iterations
1207 for(i in 1:1000){
1208
1209   # update the alpha parameter
1210   lambda<- exp(alpha+beta*habitat)
1211   lik.curr<- sum(log(dpois(y,lambda)))
1212   prior.curr<- log(dnorm(alpha,0,100))
1213   alpha.cand<-rnorm(1,alpha,.25)    # generate candidate
1214   lambda.cand<- exp(alpha.cand + beta*habitat)
1215   lik.cand<- sum(log(dpois(y,lambda.cand)))
1216   prior.cand<- log(dnorm(alpha.cand,0,100))
1217   mhratio<- exp(lik.cand +prior.cand - lik.curr-prior.curr)
```

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

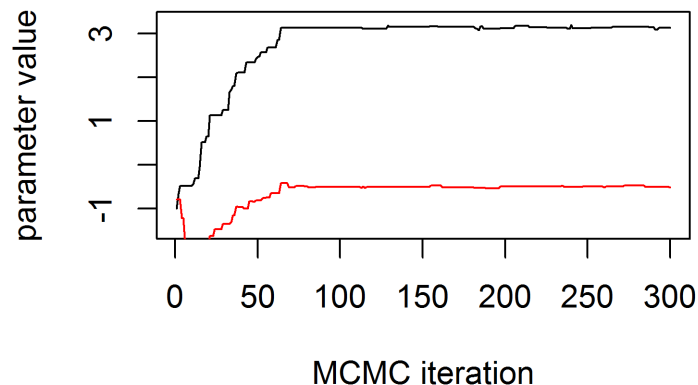


Figure 2.3: MCMC output for Poisson regression parameters (top trace: intercept  $\alpha$ ; bottom trace: slope  $\beta$ ). This is for  $\delta = 0.25$ .

```

1218 if(runif(1)< mhratio)
1219     alpha<-alpha.cand
1220
1221 # update the beta parameter
1222 lik.curr<- sum(log(dpois(y,exp(alpha+beta*habitat))))
1223 prior.curr<- log(dnorm(beta,0,100))
1224 beta.cand<-rnorm(1,beta,.25)
1225 lambda.cand<- exp(alpha+beta.cand*habitat)
1226 lik.cand<- sum(log(dpois(y,lambda.cand)))
1227 prior.cand<- log(dnorm(beta.cand,0,100))
1228 mhratio<- exp(lik.cand + prior.cand - lik.curr - prior.curr)
1229 if(runif(1)< mhratio)
1230     beta<-beta.cand
1231
1232 out[i,]<-c(alpha,beta)           # save the current values
1233 }
1234
1235
1236 plot(out[,1],ylim=c(-1.5,3.3),type="l",lwd=2,ylab="parameter value",
1237      xlab="MCMC iteration")
1238 lines(out[,2],lwd=2,col="red")

```

1239 The first 300 iterations of the MCMC history of each parameter is shown in  
 1240 Fig. 2.3. The appearance of this is not very appealing but a couple of things

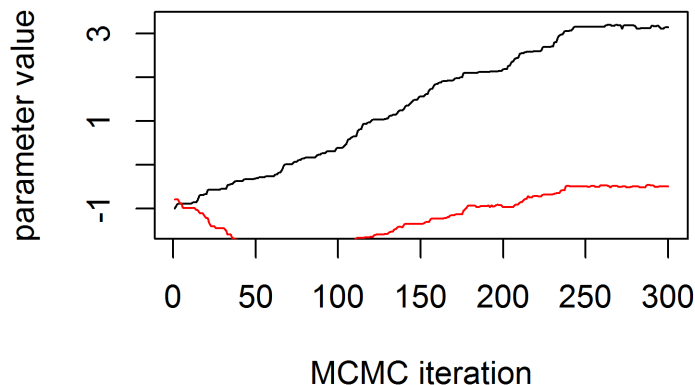


Figure 2.4: Same as previous fig but with  $\delta = 0.05$ .

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

**Remarks:** We used a specific set of starting values for these simulations. It should be clear that starting values closer to the mass of the posterior distribution might cause burn-in to occur faster. As an exercise, evaluate that. (2) Clearly the influence of the proposal standard deviation term is important. Small values lead to much better mixing but it should be noted that values that are too small will slow down burn-in and also lead to high correlation. This suggests there is an optimal value of the Metropolis-Hastings tuning parameter<sup>11</sup>. As an exercise you should contemplate finding that optimal value for this problem<sup>12</sup> (3) For the flat normal prior distributions here we could leave the prior contribution out of the full conditional evaluation since it is locally constant,

<sup>11</sup>Defined previously????

<sup>12</sup>effective sample size definition?

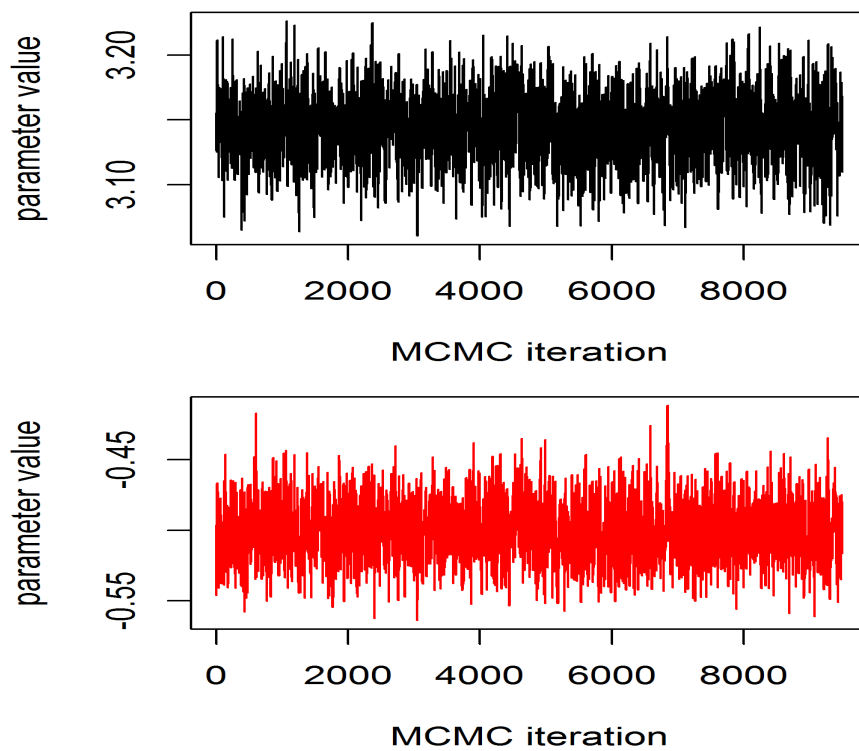


Figure 2.5: nice grassy mcmc output, longer run of previous with  $\delta = 0.05$ .

1262 i.e., constant in the vicinity of the posterior mass, and thus has no practical  
 1263 effect. Removing the prior contribution from the MH acceptance probability is  
 1264 equivalent to saying that the parameters have an improper uniform prior, i.e.,  
 1265  $\alpha \sim \text{const}$ , which is commonly used for mean parameters in practice. Note also  
 1266 that we have used a different prior than in our **WinBUGS** model specifica-  
 1267 tion given previously. As an exercise, evaluate whether this seems to affect the  
 1268 result.

## 1269 2.10 Poisson GLM with Random Effects

1270 What we will be doing in most of this book is dealing with random effects in  
 1271 GLM-like models - similar to what are usually referred to as generalized linear  
 1272 mixed models (GLMMs). We provide a brief introduction by way of example,  
 1273 extending our Poisson regression model to include a random effect.

1274 ANDY STOPPED HERE

1275 **The Log-Normal mixture:** The classical situation involves a GLM with  
 1276 a normally distributed random effect that is additive on the linear predictor.  
 1277 For the Poisson case, we have:

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

1278 where  $\eta_i \sim \text{Normal}(0, \sigma^2)$ . A natural alternative is to have multiplicative  
 1279 gamma-distributed noise,  $\exp(\eta_i) \sim \text{Gamma}(a, b)$  which would correspond to  
 1280 a negative binomial kind of over-dispersion, implying a different mean/variance  
 1281 relationship to the log-normal mixture (the interested reader should work that  
 1282 out). Choosing between such possibilities is not a topic we will get into here be-  
 1283 cause it doesn't seem possible to provide general guidance on it. For this model  
 1284 we carried-out a goodness-of-fit evaluation using the Bayesian p-value based on  
 1285 a Pearson residual statistic. See also (Kéry, 2010, ch. 18) for an example in-  
 1286 volving a binomial mixed model<sup>13</sup>. Anyhow, it is really amazingly simple to  
 1287 express this model in **WinBUGS** and have **WinBUGS** draw samples from  
 1288 the posterior distribution using the following code for the BBS dove counts:

```
1289 data<-read.csv("pa-bbsdovedata-all.csv")
1290 locs<-data[,2:3]
1291 habitat<-data[,4]
1292 y<-data[,29]      # grab year 1990
1293 M<-length(y)
1294
1295 set.seed(2013)
1296
1297 cat("
1298 model {
1299   for (i in 1:M){
1300     y[i]~dpois(lam[i])
1301     log(lam[i])<- alpha+ beta*habitat[i] + eta[i]
```

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



```

1302     frog[i]<-beta*habitat[i] + eta[i]
1303     eta[i] ~ dnorm(0,tau)
1304     d[i]<- pow(pow(y[i],0.5)-pow(lam[i],0.5),2)
1305
1306     ynew[i]~dpois(lam[i])
1307     dnew[i]<- pow(pow(ynew[i],0.5)-pow(lam[i],0.5),2)
1308   }
1309   fit<-sum(d[])
1310   fitnew<-sum(dnew[])
1311
1312   alpha~dunif(-5,5)
1313   beta~dunif(-5,5)
1314   sigma~dunif(0,10)
1315   tau<-1/(sigma*sigma)
1316 }
1317
1318 ",file="model.txt")
1319 data <- list ( "y","M","habitat")
1320 inits <- function()
1321   list ( alpha=rnorm(1),beta=rnorm(1),sigma=runif(1,0,4))
1322 parameters <- c("alpha","beta","sigma","tau","fit","fitnew")
1323 library("R2WinBUGS")
1324
1325 out<-bugs (data, inits, parameters, "model.txt", n.thin=2,n.chains=2,
1326   n.burnin=1000,n.iter=5000,debug=TRUE)

```

1327 This produces the following posterior summary statistics:

```

1328 > print(out,digits=2)
1329 Inference for Bugs model at "model.txt", fit using WinBUGS,
1330 2 chains, each with 5000 iterations (first 1000 discarded), n.thin = 2
1331 n.sims = 4000 iterations saved
1332
1333      mean    sd   2.5%   25%   50%   75%  97.5% Rhat n.eff
1334 alpha    2.98  0.08   2.82   2.93   2.98   3.03   3.12 1.00  1400
1335 beta    -0.53  0.07  -0.68  -0.58  -0.53  -0.49  -0.38 1.01   350
1336 sigma     0.60  0.06   0.49   0.56   0.59   0.64   0.73 1.00  2000
1337 tau      2.88  0.57   1.88   2.47   2.86   3.24   4.12 1.00  2000
1338 fit      26.58  3.72  19.87  23.96  26.37  29.01  34.46 1.00  4000
1339 fitnew    26.83  3.90  19.60  24.12  26.68  29.36  35.04 1.00  4000
1340 deviance 445.94 12.18 424.00 437.40 445.20 453.90 471.50 1.00  4000
1341
1342 [... some output deleted ...]

```

1342 The Bayesian p-value for this model is

```

1343 > mean(out$sims.list$fit>out$sims.list$fitnew)
1344 [1] 0.4815

```

1345 indicating a pretty good fit. Given the site-level random effect, it would be  
 1346 surprising for this model to not fit! One thing we notice is that the posterior  
 1347 standard deviations of the regression parameters are much higher, a result of

the excess variation. Wwe would also notice much less precise predictions of hypothetical new observations.

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, 2004a; 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)<sup>14</sup>. 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 11.

### 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:

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

<sup>14</sup>Some 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”

1385 The inverse-logit (or “expit”) is

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

1386 There are many other possible link functions. However, ecologists seem to adopt  
 1387 the logit link function without question in most applications<sup>15</sup>. We sometimes  
 1388 use the “complementary log-log” (= “cloglog”) link function in ecological appli-  
 1389 cations because it arises naturally in many situations (Royle and Dorazio, 2008,  
 1390 p. 150). For example, consider the “probability of observing a count greater  
 1391 than 0” under a Poisson model:  $\Pr(y > 0) = 1 - \exp(-\lambda)$ . In that case,

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

1392 So that if you have covariates in your linear predictor for  $E(y)$  under a Poisson  
 1393 model then they are linear on the complementary log-log link of  $p$ . In models  
 1394 of species occurrence it seems natural to view occupancy as being derived from  
 1395 local abundance  $N$  (Royle and Nichols, 2003; Royle and Dorazio, 2006; Dorazio,  
 1396 2007). Therefore, models of local abundance in which  $N \sim \text{Poisson}(A\lambda)$  for a  
 1397 habitat patch of area  $A$  implies a model for occupancy  $\psi$  of the form

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

1398 We will use the cloglog link in some analyses of SCR models in chapter 5 and  
 1399 elsewhere.

## 1400 2.11.2 Example: Waterfowl Banding Data

1401 It would be easy to consider a standard “distribution modeling” application  
 1402 where  $K = 1$  and the outcome is occurrence ( $y = 1$ ) or not ( $y = 0$ ) of some  
 1403 species. Such examples abound in books (e.g., Royle and Dorazio (2008, ch. 3);  
 1404 Kéry (2010, ch. 21); Kéry and Schaub (2011, ch. 13)) and in the literature.  
 1405 Instead, we will consider an example involving band returns of waterfowl which  
 1406 were analyzed by Royle and Dubovsky (2001)<sup>16</sup>.

1407 For these data,  $y_i$  is the number of waterfowl bands recovered out of  $B_i$   
 1408 birds banded at some location  $\mathbf{s}_i$ . In this case  $B_i$  is fixed. Thinking about  
 1409 recovery rate as being proportional to harvest rate, we use these data to explore  
 1410 geographic gradients in recovery rate resulting from variability in harvest pres-  
 1411 sure experienced by populations depending on their migration ecology. As such,  
 1412 we fit a basic binomial GLM with a linear response to geographic coordinates  
 1413 (including an interaction term). The data are provided with the **R** package  
 1414 **scrbook**. Here we provide the part of the script for creating the model and  
 1415 fitting the model in **WinBUGS** using the **bugs** function. There are few struc-  
 1416 tural differences between this model and the Poisson GLM fitted previously.  
 1417 The main things are due to the data structure (we have a matrix here instead

<sup>15</sup>a notable exception is distance sampling, which is all about choosing among link functions

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

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:

```

1421 load("mallarddata") # not sure how this will look
1422
1423 sink("model.txt")
1424 cat("
1425 model {
1426   for(t in 1:5){
1427     for(i in 1:nobs){
1428       y[i,t] ~ dbin(p[i,t], B[i,t])
1429       logit(p[i,t]) <- alpha0[t] + alpha1*X[i,1] + alpha2*X[i,2] + alpha3*X[i,1]*X[i,2]
1430     }
1431   }
1432   alpha1~dnorm(0,.001)
1433   alpha2~dnorm(0,.001)
1434   alpha3~dnorm(0,.001)
1435   for(t in 1:5){
1436     alpha0[t] ~ dnorm(0,.001)
1437   }
1438 }
1439 ",fill=TRUE)
1440 sink()
1441
1442 data <- list(B=mallard.banding, y=mallard.recoveries,
1443             nobs=nrow(banding.locs),X=banding.locs)
1444 inits <- function(){
1445   list(alpha0=rnorm(5),alpha1=0,alpha2=0,alpha3=0) }
1446 parms <- list('alpha0','alpha1','alpha2','alpha3')
1447 out <- bugs(data,inits, parms,"model.txt",n.chains=3,
1448            n.iter=2000,n.burnin=1000, n.thin=2,debug=TRUE)

```

Posterior summaries of model parameters are as follows:

```

1450 > print(out,digits=3)
1451 Inference for Bugs model at "model.txt", fit using WinBUGS,
1452 3 chains, each with 2000 iterations (first 1000 discarded), n.thin = 2
1453 n.sims = 1500 iterations saved
1454
1455      mean    sd    2.5%    25%    50%    75%    97.5%  Rhat  n.eff
1456 alpha0[1] -2.346 0.036 -2.417 -2.370 -2.346 -2.323 -2.277 1.001 1500
1457 alpha0[2] -2.356 0.032 -2.420 -2.379 -2.356 -2.335 -2.292 1.001 1500
1458 alpha0[3] -2.220 0.035 -2.291 -2.244 -2.219 -2.197 -2.153 1.001 1500
1459 alpha0[4] -2.144 0.039 -2.225 -2.169 -2.143 -2.116 -2.068 1.000 1500
1460 alpha0[5] -1.925 0.034 -1.990 -1.949 -1.924 -1.901 -1.856 1.004 570
1461 alpha1     -0.023 0.003 -0.028 -0.025 -0.023 -0.022 -0.018 1.001 1500
1462 alpha2      0.020 0.006  0.009  0.016  0.020  0.024  0.031 1.001 1500
1463 alpha3      0.000 0.001 -0.002 -0.001  0.000  0.000  0.002 1.001 1500
1464 deviance 1716.001 4.091 1710.000 1713.000 1715.000 1718.000 1726.000 1.001 1500
1465
1466 [... some output deleted ...]

```

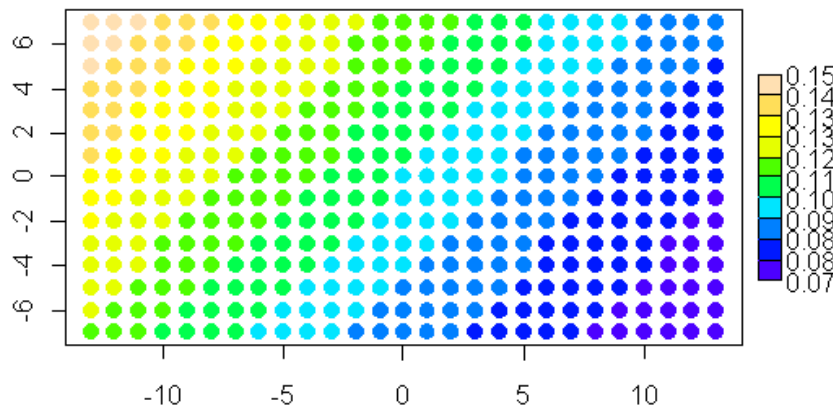


Figure 2.6: Predicted recovery rate of bands.

1466 The basic result suggests a negative east-west gradient and a positive south  
 1467 to north gradient but no interaction. A map of the response surface is shown  
 1468 in Fig. 2.6. We did an additional MCMC run where we saved the binomial  
 1469 parameter  $p$  and computed the Bayesian p-value (double use of “p” here is  
 1470 confusing, but I guess that happens sometimes!) using a fit statistic based on  
 1471 the Freeman-Tukey statistic (see Section XXX above). The result indicates that  
 1472 the linear response surface model does not provide an adequate fit of the data.  
 1473 The reader should contemplate whether this invalidates the basic interpretation  
 1474 of the result.

## 1475 2.12 Summary and Outlook

1476 GLMs and GLMMs are the most useful statistical methods in all of ecology.  
 1477 The principles and procedures underlying these methods are relevant to nearly  
 1478 all modeling and analysis problems in every branch of ecology. Moreover, un-  
 1479 derstanding how to analyze these models is crucial in a huge number of diverse  
 1480 problems. If you understand and can conduct classical likelihood and Bayesian  
 1481 analysis of Poisson and binomial GLM(M)s, then you will be successful analyzing  
 1482 and understanding more complex classes of models that arise. We will  
 1483 see shortly that spatial capture-recapture models are a type of GLMM and  
 1484 thus having a basic understanding of the conceptual origins and formulation of  
 1485 GLM(M)s and their analysis is extremely useful.

1486 We note that GLM(M)s are routinely analyzed by likelihood methods but we  
 1487 have focused on Bayesian analysis here in order to develop the tools that are less

familiar to most ecologists. In particular, Bayesian analysis of models with random effects is relatively straightforward because the models are easy to analyze conditional on the random effect, using methods of MCMC. Thus, we will often analyze SCR models in later chapters by MCMC, explicitly adopting a Bayesian inference framework. In that regard, the various **BUGS** engines (**WinBUGS**, **OpenBUGS**, **JAGS**) are enormously useful because they provide an accessible platform for carrying out analyses by MCMC by just describing the model, and not having to worry about how to actually build MCMC algorithms. That said, the **BUGS** language is more important than just to the extent that it enables one to do MCMC - it is useful as a modeling tool because it fosters understanding, in the sense that it forces you to become intimate with your model. You have to write down all of the probability assumptions, the relationships between observations and latent variables and parameters. This is really a great learning paradigm that you can grow with.

While we have emphasized Bayesian analysis in this chapter, and make primary use of it through the book, we will provide an introduction to likelihood analysis in chapter 9 and use those methods also from time to time. Before getting to that, however, it will be useful to talk about more basic, conventional closed population capture-recapture models and these are the topic of the next chapter.

## 1508 Chapter 3

# 1509 Closed population models





## Chapter 4

# 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 accom-

plish 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.

## 4.1 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 4.1. 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 using standard methods that you are already familiar with (i.e., GLMs - see chapter 2).

Table 4.1: 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

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)):

$$n_1, n_2, \dots, n_K \sim \text{Multin}(N, \pi(1), \pi(2), \dots, \pi(K)) \quad (4.1)$$

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.

### 4.1.1 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 5.

### 4.1.2 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).

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

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

Table 4.2: Modes of analysis of capture-recapture models.

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 4.2 for Model M0).

## 4.2 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.<sup>1</sup> 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  $\psi$  is the “data augmentation parameter.”

### 4.2.1 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

<sup>1</sup> 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  $n$ .

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 4.3. 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  $\psi$ , 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 4.3 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 4.3 - 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 to sampling. We assume that  $z_i \sim \text{Bern}(\psi)$  where  $\psi$  is the probability that an individual in the data set of size  $M$  is a member of the sampled population - in the sense that  $1 - \psi$  is the probability of realizing a “structural zero” in the augmented data set. The zero-inflated binomial model which arises under data augmentation can be formally expressed by the following

1723 set of assumptions:

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

1724 for  $i = 1, \dots, M$ , where  $\delta(0)$  is a point mass at  $y = 0$ .

1725 We note that  $N$  is no longer an explicit parameter of this model. Instead,  
 1726 we estimate  $\psi$  and functions of the latent variables. In particular, under the  
 1727 assumptions of the zero-inflated model,  $z_i \overset{iid}{\sim} \text{Bern}(\psi)$ ; therefore,  $N$  is a function  
 1728 of these latent variables:

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

1729 Further, we note that the latent  $z_i$  parameters can be removed from the model  
 1730 by integration, in which case the joint probability of the data is

$$\Pr(y_1, \dots, y_M | p, \psi) = \prod_{i=1}^M \psi \text{Bin}(y_i | K, p) + I(y_i = 0)(1 - \psi) \quad (4.2)$$

1731 Which can be maximized directly to obtain the MLEs of the structural param-  
 1732 eters  $\psi$  and  $p$  or those of other more complex models (e.g., see Royle, 2006). We  
 1733 could estimate these parameters and then use them to obtain an estimator of  
 1734  $N$  using the so-called “Best unbiased predictor” (see Royle and Dorazio, 2011).

### 1735 4.2.2 Model $M_0$ in BUGS

1736 For model  $M_0$  in which we can aggregate the encounter data to individual-  
 1737 specific encounter frequencies, the augmented data are given by the vector of fre-  
 1738 quencies  $(y_1, \dots, y_n, 0, 0, \dots, 0)$ . The zero-inflated model of the augmented data  
 1739 combines the model of the latent variables,  $z_i \sim \text{Bern}(\psi)$  with the conditional-  
 1740 on- $z$  binomial model:

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

1741 It is convenient to express the conditional-on- $z$  observation model concisely as:

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

1742 Thus, if  $z_i = 0$  then the success probability of the binomial distribution is  
 1743 identically 0 whereas, if  $z_i = 1$ , then the success probability is  $p$ . This is useful  
 1744 in describing the model in the **BUGS** language, as shown below. Note the last  
 1745 line of the model specification here provides the expression for computing  $N$   
 1746 from the data augmentation variables  $z_i$ .

Table 4.3: 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



```

1747 p ~ dunif(0,1)
1748 psi~dunif(0,1)
1749
1750 # nind = number of individuals captured at least once
1751 # nz = number of uncaptured individuals added for PX-DA
1752 for(i in 1:(nind+nz)) {
1753   z[i]~dbern(psi)
1754   mu[i]<-z[i]*p
1755   y[i]~dbin(mu[i],K)
1756 }
1757
1758 N<-sum(z[1:(nind+nz)])

```

1759 Specification of a more general model in terms of the individual encounter  
 1760 observations  $y_{ik}$  is not much more difficult than for the individual encounter  
 1761 frequencies. We define the observation model by a double loop and change the  
 1762 indexing of things accordingly, i.e.,

```

1763 for(i in 1:(nind+nz)) {
1764   z[i]~dbern(psi)
1765   for(k in 1:K){
1766     mu[i,k]<-z[i]*p
1767     y[i,k]~dbin(mu[i,k],1)
1768   }
1769 }

```

1770 In this manner, it is straightforward to incorporate covariates on  $p$  (see discus-  
 1771 sion of this below and also chapt. 8 (REF XYZ) and consider other extensions.

### 1772 4.2.3 Formal development of data augmentation

1773 Use of DA for solving inference problems with unknown  $N$  can be justified as  
 1774 originating from the choice of uniform prior on  $N$ . The  $\text{Unif}(0, M)$  prior for  $N$   
 1775 is innocuous in the sense that the posterior associated with this prior is equal  
 1776 to the likelihood for sufficiently large  $M$ . One way of inducing the  $\text{Unif}(0, M)$   
 1777 prior on  $N$  is by assuming the following hierarchical prior:

$$\begin{aligned}
 N &\sim \text{Bin}(M, \psi) \\
 \psi &\sim \text{Unif}(0, 1)
 \end{aligned}
 \tag{4.3}$$

1778 which includes a new model parameter  $\psi$ . This parameter denotes the prob-  
 1779 ability that an individual in the super-population of size  $M$  is a member of  
 1780 the population of  $N$  individuals exposed to sampling. The model assumptions,  
 1781 specifically the multinomial model (eq. XYZ) and eq. 4.3, may be combined to  
 1782 yield a reparameterization of the conventional model that is appropriate for the  
 1783 augmented data set of known size  $M$ :

$$(n_1, n_2, \dots, n_K) \sim \text{Multin}(M, \psi\pi(1), \psi\pi(2), \dots, \psi\pi(K))
 \tag{4.4}$$

1784 This arises by removing  $N$  from Eq. multinomial XYZ by integrating over the  
 1785 binomial prior distribution for  $N$ . Thus, the models we analyze under data  
 1786 augmentation arise formally by removing the parameter  $N$  from the ordinary  
 1787 model - the model conditional on  $N$  - by integrating over a binomial prior  
 1788 distribution for  $N$ .

1789 Note that the  $M - n$  unobserved individuals in the augmented data set  
 1790 have probability  $\psi\pi(0) + (1 - \psi)$ , indicating that these unobserved individuals  
 1791 are a mixture of individuals that are sampling zeros ( $\psi\pi_0$ , and belong to the  
 1792 population of size  $N$ ) and others that are “structural zeros” (occurring in the  
 1793 augmented data set with probability  $1 - \psi$ ). In Eq. 4.4  $N$  has been eliminated as  
 1794 a formal parameter of the model by marginalization (integration) and replaced  
 1795 with the new parameter  $\psi$ , which we will call the “data augmentation param-  
 1796 eter.” However, the full likelihood containing both  $N$  and  $\psi$  can be analyzed  
 1797 (see Royle et al., 2007).

#### 1798 4.2.4 Remarks on Data Augmentation

1799 Data augmentation may seem like a strange and mysterious black-box, and  
 1800 likely it is unfamiliar to most people even those with extensive experience with  
 1801 capture-recapture models. However, it really is a formal reparameterization of  
 1802 capture-recapture models in which  $N$  is removed from the ordinary (conditional-  
 1803 on- $N$ ) model by integration. In the case of Model M0, data augmentation pro-  
 1804 duces the zero-inflated binomial which is distinct from the original observation  
 1805 model, but only in the sense that it embodies, explicitly, the  $\text{Unif}(0, M)$  prior  
 1806 for  $N$ . Choice of  $M$  might be cause for some concern related to potential sen-  
 1807 sitivity to choice of  $M$ . The guiding principle is that it should be chosen large  
 1808 enough so that the posterior for  $N$  is not truncated, but no larger because large  
 1809 values entail more computational burden. It seems likely that the properties of  
 1810 the Markov chains should be affected by  $M$  and so some optimality might exist  
 1811 (Gopalaswamy, 2012), as in occupancy models (Mackenzie and Royle, 2005).  
 1812 Formal analysis of this is required.

1813 We emphasize the motivation for data augmentation being that it produces a  
 1814 data set of fixed size, so that the parameter dimension in any capture-recapture  
 1815 model is also fixed. As a result, MCMC is a relatively simple proposition us-  
 1816 ing standard Gibbs Sampling. Consider the simplest context - analyzing Model  
 1817 M0 using the occupancy model. In this case, DA converts Model M0 to a ba-  
 1818 sic occupancy model and the parameters  $p$  and  $\psi$  have known full-conditional  
 1819 distributions (in fact, beta distributions) that can be sampled from directly.  
 1820 Furthermore, the data augmentation variables - the latent data augmentation  
 1821 variables  $z$ , can be sampled from Bernoulli full conditionals. MCMC is not  
 1822 too much more difficult for complicated models - sometimes the hyperparam-  
 1823 eters need to be sampled using a Metropolis-Hastings step, but nothing more  
 1824 sophisticated than that is required.

1825 There are other approaches to analyzing models with unknown  $N$ , using re-  
 1826 versible jump MCMC (RJMCMC) or other so-called “trans-dimensional” (TD)

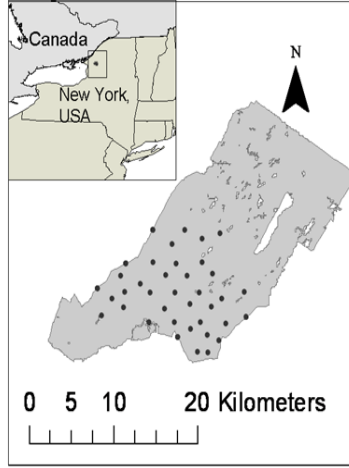


Figure 4.1: Fort Drum study area and hair snare locations.

algorithms<sup>2</sup> (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.

#### 4.2.5 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 (missing). The specific data used here are encounter histories on 47 individuals obtained from an array of 38 baited “hair snares” (Fig. 4.1) 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.

<sup>2</sup>Look these citations up in Royle-Dorazio EURING paper

```

1850 # Consider adding comments to your code.
1851 ## Good idea. This will be done in final draft
1852 trapmat<-read.csv("FDtrapmat.csv")
1853 bearArray<-source("FDbeararray.R")$value
1854 nind<-dim(bearArray)[1]
1855 K<-dim(bearArray)[3]
1856 ntraps<-dim(bearArray)[2]
1857
1858 M=175
1859 nz<-M-nind
1860
1861 Xaug <- array(0, dim=c(M,ntraps,K))
1862 Xaug[1:nind,,]<-bearArray
1863 y<- apply(Xaug,c(1,3),sum)
1864 y[y>1]<-1
1865 ytot<-apply(y,1,sum) # total encounters out of K

```

Note that the raw data,  $y$ , 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 ytot above. The BUGS model file along with commands to fit the model are as follows:

```

1872 set.seed(2013) # to obtain the same results each time
1873 data0<-list(y=y,M=M,K=K)
1874 params0<-list('psi','p','N')
1875 zst=c(rep(1,nind),rbinom(M-nind, 1, .5))
1876 inits = function() {list(z=zst, psi=runif(1), p=runif(1)) }
1877
1878 cat("
1879 model {
1880
1881   psi~dunif(0, 1)
1882   p~dunif(0,1)
1883
1884   for (i in 1:M){
1885     z[i]~dbern(psi)
1886     for(k in 1:K){
1887       tmp[i,k]<-p*z[i]
1888       y[i,k]~dbin(tmp[i,k],1)
1889     }
1890   }
1891   N<-sum(z[1:M])
1892 }
1893 ",file="modelM0.txt")
1894
1895 fit0 = bugs(data0, inits, params0, model.file="modelM0.txt",
1896            n.chains=3, n.iter=2000, n.burnin=1000, n.thin=1,
1897            debug=TRUE,working.directory=getwd())

```

1898 The posterior summary statistics from this analysis are as follows:

```
1899 > print(fit0,digits=2)
1900 Inference for Bugs model at "modelM0.txt", fit using WinBUGS,
1901 3 chains, each with 2000 iterations (first 1000 discarded)
1902 n.sims = 3000 iterations saved
1903      mean      sd   2.5%    25%    50%    75%   97.5% Rhat n.eff
1904 psi      0.29  0.04   0.22   0.26   0.29   0.31   0.36    1  3000
1905 p       0.30  0.03   0.25   0.28   0.30   0.32   0.35    1  3000
1906 N       49.94  1.99  47.00  48.00  50.00  51.00  54.00    1  3000
1907 deviance 489.05 11.28 471.00 480.45 488.80 495.40 513.70    1  3000
1908
1909 [... some output deleted ...]
```

1910 **WinBUGS** did well in choosing an MCMC algorithm for this model – we  
 1911 have  $\hat{R} = 1$  for each parameter, and an effective sample size of 3000, equal to  
 1912 the total number of posterior samples. We see that the posterior mean of  $N$   
 1913 under this model is 49.94 and a 95% posterior interval is (48, 54). We revisit  
 1914 these data later in the context of more complex models.

1915 In order to obtain an estimate of density,  $D$ , we need an area to associate with  
 1916 the estimate of  $N$ , and commonly used procedures to conjure up such an area  
 1917 include buffering the trap array by the home range radius, often estimated by the  
 1918 mean maximum distance moved (MMDM)<sup>3</sup>, 1/2 MMDM (Dice, 1938) or directly  
 1919 from telemetry data (REF XXX NEED REF HERE XXXXX). Typically, the  
 1920 trap array is defined by the convex hull around the trap locations, and this is  
 1921 what we applied a buffer to. We computed the buffer by using an estimate of  
 1922 the mean female home range radius (2.19 km) estimated from telemetry studies  
 1923 (Bales et al., 2005) instead of using an estimate based on our relatively more  
 1924 sparse recapture data<sup>4</sup>. For the Fort Drum study, the convex hull has area  
 1925  $157.135 \text{ km}^2$ , and the buffered convex hull has area  $277.011 \text{ km}^2$ . To create  
 1926 this we used functions contained in the **R** package **rgeos** and created a utility  
 1927 function **bcharea** which is in our **R** package **scrbook**. The commands are as  
 1928 follows:

```
1929 library("rgeos")
1930
1931 bcharea<-function(buff,traplocs){
1932 p1<-Polygon(rbind(traplocs,traplocs[1,]))
1933 p2<-Polygons(list(p1=p1),ID=1)
1934 p3<-SpatialPolygons(list(p2=p2))
1935 p1ch<-gConvexHull(p3)
1936 bp1<-(gBuffer(p1ch, width=buff))
1937 plot(bp1, col='gray')
1938 plot(p1ch, border='black', lwd=2, add=TRUE)
1939 gArea(bp1)
```

---

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

<sup>4</sup>BETH: Why?

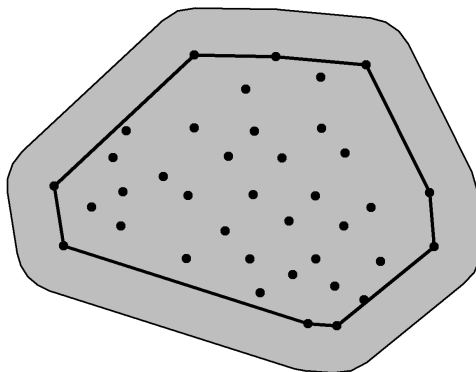


Figure 4.2: buffered convex hull of the bear hair snare array

```

1940 }
1941
1942 bcharea(2.19,traplocs=trapmat)
1943
1943 The resulting buffered convex hull is shown in Fig. 4.2.
1944 To conjure up a density estimate under model  $M_0$ , we compute the appro-
1945 priate posterior summary of  $N$  and the prescribed area ( $277.011 \text{ km}^2$ ):
1946
1946 > summary(fit0$sims.list$N/277.011)
1947   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
1948 0.1697 0.1733 0.1805 0.1803 0.1841 0.2130
1949
1950 > quantile(fit0$sims.list$N/277.011,c(0.025,0.975))
1951    2.5%    97.5%
1952 0.1696684 0.1949381
1953
1953 which yields a density estimate of about  $0.18 \text{ ind/km}^2$ , and a 95% Bayesian
1954 confidence interval of  $(0.170, 0.195)$ .
1955 The obvious limitation of this estimate and, indeed, of the whole process, is
1956 that our choice of “area” is completely subjective - which area should we use?
1957 MMDM? One-half MMDM? Estimated from telemetry data? And, furthermore,
1958 how certain are we of this area? Can we quantify our uncertainty about this
1959 quantity? More important, what exactly is the meaning of this area and, in this
1960 context, how do we gauge bias and/or variance of “estimators” of it? (i.e., what
1961 is it estimating?).

```

### 4.3 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

### 4.4 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 4.5 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 4.3 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  $\phi$ . 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  $\phi$  should depend on *where* an individual lives, i.e., it should be individual-specific  $\phi_i$  (Chandler et al., 2011). This system describes, precisely, that of "random temporary emigration" (Kendall, 1997) where  $\phi_i$  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



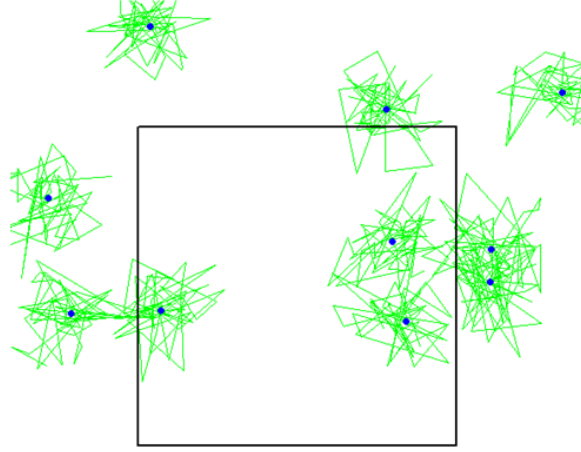


Figure 4.3: A quadrat searched for lizards and the locations of each lizard over some period of time.

2047 and formally with auxiliary spatial information. If individuals are detected with  
 2048 probability  $p_0$ , *conditional* on  $z_i = 1$ , then the marginal probability of detection  
 2049 of individual  $i$  is

$$p_i = p_0 \phi_i$$

2050 so we see clearly that individual heterogeneity in encounter probability is in-  
 2051 duced as a result of the juxtaposition of individuals (i.e., their home ranges)  
 2052 with the sample apparatus and the movement of individuals about their home  
 2053 range.

2054 We will work with a specific type of Model  $M_h$  here, that in which we extend  
 2055 the basic binomial observation model of Model  $M_0$  so that

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

2056 where

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

2057 We could as well write

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

2058 This “logit-normal mixture” was analyzed by Coull and Agresti (1999) and  
 2059 elsewhere. It is a natural extension of the basic model with constant  $p$ , as a  
 2060 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).

#### 4.4.1 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. 9 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
}
```

2103 **4.4.2 Analysis of the Fort Drum data**

2104 The logit-normal heterogeneity model was fitted to the bear data from the Fort  
 2105 Drum study, and we used data augmentation to produce a data set of  $M = 500$   
 2106 individuals. We ran the model using **JAGS** with the instructions given as  
 2107 follows<sup>5</sup>.

```

2108 [... get data as before ....]
2109
2110 set.seed(2013)
2111
2112 cat("
2113 model{
2114   p0 ~ dunif(0,1)          # prior distributions
2115   mup<- log(p0/(1-p0))
2116   taup~dgamma(.1,.1)
2117   sigmap<-sqrt(1/taup)
2118   psi~dunif(0,1)
2119
2120   for(i in 1:(nind+nz)){
2121     z[i]~dbern(psi)        # zero inflation variables
2122     lp[i] ~ dnorm(mup,taup) # individual effect
2123     logit(p[i])<-lp[i]
2124     mu[i]<-z[i]*p[i]
2125     y[i]~dbin(mu[i],K)    # observation model
2126   }
2127
2128   N<-sum(z[1:(nind+nz)])
2129 }
2130 ",file="modelMh.txt")
2131
2132 library("rjags")
2133 jm<- jags.model("modelMh.txt", data=data1, inits=inits, n.chains=4,
2134                n.adapt=1000)
2135 jout<- coda.samples(jm, params1, n.iter=50000, thin=1)

```

2136 CHANGE THIS TO RUN SIGMA DUNIF(0,5) PRIOR INSTEAD OF  
 2137 TAUY

2138 ANDY IS WORKING THIS SECTION RIGHT NOW. KEY IS-  
 2139 SUE IS THAT BEAR DATA HAVE VERY LONG RIGHT TAIL.  
 2140 PSSIBLY NOT EVEN IDENTIFIABLE FOR LOGIT-NORMAL MODEL.  
 2141 NEED TO RUN WINBUGS AND JAGS FOR M=500 AND RUN  
 2142 A 200K RUN AND THEN MAYBE A 500K RUN TO SEE HOW  
 2143 THINGS LOOK. THEN RUN MY R CODE BELOW FOR 5 MIL-  
 2144 LION ITTERS OR SOME bs like that. I'm not sure if this is a teaching  
 2145 moment (Link 2003) or if we need a different example here!

2146 This produces the posterior distribution for  $N$  shown in Fig. 4.4. Posterior  
 2147 summaries of parameters are given as follows:

---

<sup>5</sup>For WinBUGS, should provide starts for lp and sigma or sometimes WinBUGS breaks

```

2148 > summary(jout)
2149 Iterations = 1001:201000
2150 Thinning interval = 1
2151 Number of chains = 4
2152 Sample size per chain = 2e+05
2153
2154 1. Empirical mean and standard deviation for each variable,
2155    plus standard error of the mean:
2156
2157           Mean          SD Naive SE Time-series SE
2158 N      108.63259 52.53176 5.873e-02      2.077726
2159 p0       0.08615  0.05919 6.618e-05      0.001950
2160 psi      0.21841  0.10615 1.187e-04      0.004141
2161 sigmap   1.94096  0.51014 5.703e-04      0.018244
2162
2163 2. Quantiles for each variable:
2164
2165           2.5%       25%       50%       75%       97.5%
2166 N      59.00000 77.00000 93.00000 121.0000 261.0000
2167 p0      0.00418  0.03852  0.07657   0.1240   0.2210
2168 psi     0.11230  0.15373  0.18920   0.2457   0.5241
2169 sigmap  1.11906  1.57643  1.87752   2.2386   3.1166

```

2170 We used  $M = 500$  for this analysis and we note that while the posterior  
2171 mass of  $N$  is concentrated away from this upper bound (Fig. 4.4), the posterior  
2172 has an extremely long right tail, with some posterior values at the upper bound  
2173  $N = 500$ . Maybe or maybe not sufficient data augmentation. The model runs  
2174 effectively in **WinBUGS** but sometimes with apparently inefficient mixing for  
2175 reasons that may be related to bad starting values. In some cases this was  
2176 resolved if we supplied starting values for the  $\text{logit}(p_i)$  parameters and  $\tau$ .

2177 **to do:** insert final results. longer run. more data augmentation. compare  
2178 with winbugs.

2179 The posterior mode compares well with the MLE which we obtained using  
2180 the **R** code contained in Panel 6.1 of Royle and Dorazio (2008). The MLE of  
2181  $\log(n_0)$ , the logarithm of the number of uncaptured individuals, is  $\log(n_0) =$   
2182  $3.86$  and therefore the MLE is  $\hat{N} = \exp(3.86) + 47 = 94.47$  consistent with the  
2183 apparent mode in Fig. 4.4. <sup>6</sup> To convert this to density we use the buffered area  
2184 as computed above  $(255.3 \text{ km}^2)$ <sup>7</sup> and perform the required summary analysis on  
2185 the posterior samples of  $N$ , which results in about  $0.37$  individuals/ $\text{km}^2$ . The  
2186 reader should carry out this analysis to confirm the estimates, and also obtain  
2187 the 95% confidence interval.

---

<sup>6</sup>We note that the result is inconsistent with Gardner et al. (2009) who reported an MLE of  
104.1 ( $\text{density} = 0.437 \text{ inds}/\text{km}^2$ ) although we do not know the reason for this at the present  
time.

<sup>7</sup>WRONG #

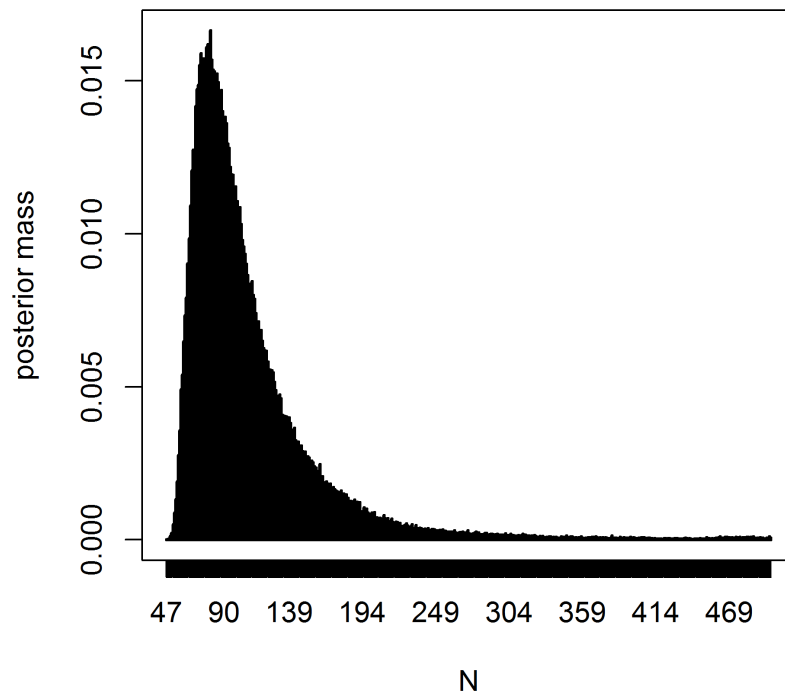


Figure 4.4: Posterior of  $N$  for Fort Drum bear study data under the logit-normal version of model  $M_h$ . From WinBUGS output. 200k samples.

### 2188 4.4.3 Building your own MCMC algorithm

2189 For fun, we construct our own MCMC algorithm using a Metropolized Gibbs  
 2190 sampler for Model  $M_h$ . In chapter 10 we devise MCMC algorithms for spatial  
 2191 capture-recapture models and the basic conceptual and technical considerations  
 2192 are entirely analogous to Model  $M_h$ .

2193 To begin, we first collect all of our model components which are as follows:  
 2194  $[y_i|p_i, z_i]$ ,  $[p_i|\mu_p, \sigma_p]$ , and  $[z_i|\psi]$  for *each*  $i = 1, 2, \dots, M$  and then prior dis-  
 2195 tributions  $[\mu_p]$ ,  $[\sigma_p]$  and  $[\psi]$ . The joint posterior distribution of all unknown  
 2196 quantities in the model is proportional to the joint distribution of all elements  
 2197  $y_i, p_i, z_i$  and also the prior distributions of the prior parameters:

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

2198 For prior distributions, we assume that  $\mu_p, \sigma_p, \psi$  are mutually independent and  
 2199 for  $\mu_p$  and  $\sigma_p$  we use improper uniform priors, and  $\psi \sim \text{Unif}(0, 1)$ . Note that  
 2200 the likelihood contribution for each individual, when conditioned on  $p_i$  and  $z_i$ ,  
 2201 does not depend on  $\psi$ ,  $\mu_p$ , or  $\sigma_p$ . As such, the full-conditionals for the structural  
 2202 parameters  $\psi$  only depends on the collection of data augmentation variables  $z_i$ ,  
 2203 and that for  $\mu_p$  and  $\sigma_p$  will only depends on the collection of latent variables  
 2204  $p_i; i = 1, 2, \dots, M$ . The full conditionals for all the unknowns are as follows:

2205 (1) For  $p_i$ :

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

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

2206 (2) for  $z_i$ :

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

2207 (3) For  $\mu_p$ :

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

2208 (4) For  $\sigma_p$ :

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

2209 (5) For  $\psi$ :

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

2210 What we've done here is identify each of the full conditional distributions  
 2211 in sufficient detail to toss them into our Metropolis-Hastings algorithm. With  
 2212 the exception of  $\psi$  which has a convenient analytic solution – it is a beta dis-  
 2213 tribution which we can easily sample directly. In truth, we could also sample  
 2214  $\mu_p$  and  $\sigma_p^2$  directly with certain choices of prior distributions. For example, if  
 2215  $\mu_p \sim \text{Normal}(0, 1000)$  then the full conditional for  $\mu_p$  is also normal, etc.. We  
 2216 implement an MCMC algorithm for this model in the following block of **R** code.

2217 The basic structure is: initialize the parameters and create any required output  
 2218 or intermediate “holders”, and then begin the main MCMC loop which, in this  
 2219 case, generates 100000 samples.

```

2220
2221 ## obtain the bear data by executing the previous data grabbing
2222 ## function
2223
2224 temp<-getdata()
2225 M<-temp$M
2226 K<-temp$K
2227 ytot<-temp$ytot
2228
2229
2230 ###
2231 ### MCMC algorithm for Model Mh
2232
2233 out<-matrix(NA,nrow=100000,ncol=4)
2234 dimnames(out)<-list(NULL,c("mu","sigma","psi","N"))
2235 lp<- rnorm(M,-1,1)
2236 p<-expit(lp)
2237 mu<- -1
2238 p0<-exp(mu)/(1+exp(mu))
2239 sigma<- 1
2240 psi<- .5
2241 z<-rbinom(M,1,psi)
2242 z[ytot>0]<-1
2243
2244 for(i in 1:100000){
2245
2246   ### update the logit(p) parameters
2247   lpc<- rnorm(M,lp,1) # 0.5 is a tuning parameter
2248   pc<-expit(lpc)
2249   lik.curr<-log(dbinom(ytot,K,z*p)*dnorm(lp,mu,sigma))
2250   lik.cand<-log(dbinom(ytot,K,z*pc)*dnorm(lpc,mu,sigma))
2251   kp<- runif(M) < exp(lik.cand-lik.curr)
2252   p[kp]<-pc[kp]
2253   lp[kp]<-lpc[kp]
2254
2255   p0c<- rnorm(1,p0,.05)
2256   if(p0c>0 & p0c<1){
2257     muc<-log(p0c/(1-p0c))
2258     lik.curr<-sum(dnorm(lp,mu,sigma,log=TRUE))
2259     lik.cand<-sum(dnorm(lp,muc,sigma,log=TRUE))
2260     if(runif(1)<exp(lik.cand-lik.curr)) {
2261       mu<-muc

```

```

2262   p0<-p0c
2263 }
2264 }
2265
2266   sigmac<-rnorm(1,sigma,.5)
2267   if(sigmac>0){
2268     lik.curr<-sum(dnorm(lp,mu,sigma,log=TRUE))
2269     lik.cand<-sum(dnorm(lp,mu,sigmac,log=TRUE))
2270     if(runif(1)<exp(lik.cand-lik.curr))
2271       sigma<-sigmac
2272   }
2273
2274   ### update the z[i] variables
2275   zc<- ifelse(z==1,0,1) # candidate is 0 if current = 1, etc..
2276   lik.curr<- dbinom(ytot,K,z*p)*dbinom(z,1,psi)
2277   lik.cand<- dbinom(ytot,K,zc*p)*dbinom(zc,1,psi)
2278   kp<- runif(M) < (lik.cand/lik.curr)
2279   z[kp]<- zc[kp]
2280
2281   psi<-rbeta(1, sum(z) + 1, M-sum(z) + 1)
2282
2283   out[i,]<- c(mu,sigma,psi,sum(z))
2284 }

```

2285     **Remarks:** (1) for parameters with bounded support, i.e.,  $\sigma_p$  and  $p_0$ , we  
2286 are using a random walk candidate generator but rejecting draws outside of the  
2287 parameter space. (2) We mostly use Metropolis-Hastings except for the data  
2288 augmentation parameter  $\psi$  which we sample directly from its full-conditional  
2289 distribution which is a beta distribution. (3) Even the latent data augmentation  
2290 variables  $z_i$  are updated using Metropolis-Hastings although they too can be  
2291 updated directly from their full-conditional.

#### 2292 4.4.4 Exercises related to model Mh

- 2293     (1) Enclose the MCMC algorithm in an R function and provide arguments for  
2294         some of the parameters of the function that a user might wish to modify.
- 2295     (2) Execute the function and compare the results to those generated from  
2296         WinBUGS in the previous section
- 2297     (3) Note that the prior distribution for the “mean” parameter is given on  
2298          $p_0 = \exp(\mu)/(1 + \exp(\mu))$ . Reformulate the algorithm with a flat prior on  
2299          $\mu$  and see what happens. Contemplate this.
- 2300     (4) Using Bayes rule, figure out the full conditional for  $z_i$  so that you don’t  
2301         have to use MH for that one. It might be more efficient. Is it?



## 4.5 Individual Covariate Models: Toward Spatial Capture-Recapture

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

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

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

$$\text{logit}(p_i) = \alpha_0 + \alpha_1 x_i$$

Classical examples of covariates influencing detection probability are type of animal (juvenile/adult or male/female), a continuous covariate such as body mass (Royle and Dorazio, 2008, chapt. 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; Langtimm, 2010).

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. More on that later.

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

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

where  $\tilde{p}_i$  is the probability that individual  $i$  appeared in the sample. That is,  $\tilde{p}_i = \Pr(y_i > 0)$ . In practice,  $\tilde{p}_i$  is estimated from the conditional-likelihood formed by the encounter histories. Namely,

$$\Pr(y_i | y_i > 0) = \Pr(y_i) / \Pr(y_i > 0)$$

where we substitute

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

with

$$\text{logit}(p_i) = \alpha_0 + \alpha_1 x_i$$

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 in some detail 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 population. For our present purposes, we consider a continuous covariate and we assume that it has a normal distribution:

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

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

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

In addition, we assume that  $p_i$  is functionally related to a covariate  $x_i$ , e.g., by the logit model given above, and we assume a distribution for  $x_i$  appropriate for the context.

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

#### 4.5.1 Example: Location of capture as a covariate.

If we had a regular grid of traps over some closed geographic system then we imagine that the average location of capture would be a decent estimate (heuristically) of an individual’s home range center. Intuitively some measure of typical distance from home range center to traps for an individual should be a decent covariate to explain heterogeneity in encounter probability, i.e., individuals with more exposure to traps should have higher encounter probabilities and vice versa. A version of this idea was put forth by Boulanger and McLellan (2001) (see also Ivan (2012)), but using the Huggins-Alho estimator and with covariate “distance to edge” of the trapping array. A limitation of this basic approach is that it does not provide a solution to the problem that the trap area is fundamentally ill-defined, nor does it readily accommodate the inherent and heterogeneous variation in this measured covariate. Here, we provide an example of this type of heuristically motivated approach using the fully model-based individual covariate model described above analyzed by data augmentation. We take a slightly different approach than that adopted by Boulanger and McLellan (2001). By analyzing the full likelihood and placing a prior distribution on the

individual covariate, we resolve the problem of having an ill-defined area over which the population size is distributed. After you read later chapters of this book, it will be apparent that SCR models represent a formalization of this heuristic procedure.

For our purposes here, we define  $x_i = ||s_i - x_0||$  where  $s_i$  is the average encounter location of individual  $i$  and  $x_0$  is the centroid of the trap array. Conceptually, individuals in the middle of the array should have higher probability of encounter and, as  $x_i$  increases,  $p_i$  should therefore decrease. We note that we have defined  $s_i$  in terms of a sample quantity - the observed mean - which is ad hoc but maybe 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. Suppose further than we imagine no individual could have a home range radius larger than  $D_{max}$ . As such, we think a reasonable distribution for this individual covariate is

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

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.

### 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. To compute the average location of capture for each individual and the distance from the centroid of the trap array, we execute the following R instructions:

```

avg.s<-matrix(NA,nrow=nind,ncol=2)
for(i in 1:nind){
  tmp<-NULL
  for(j in 1:T){
    aa<-bearArray[i,,j]
    if(sum(aa)>0){
      aa<- trapmat[aa>0,]
      tmp<-rbind(tmp,aa)
    }
  }
  avg.s[i,]<-c(mean(tmp[,1]),mean(tmp[,2]))
}
Cx<-mean(trapmat[,1])

```

```

2411 Cy<-mean(trapmat[,2])
2412 avg.s<-rbind(avg.s,matrix(NA,nrow=nz,ncol=2))
2413 xcent<- sqrt( (avg.s[,1]-Cx)^2 + (avg.s[,2]-Cy)^2)

```

2414 To define the maximum distance (maxD) from the centroid, we use that of  
 2415 the farthest trap, and so maxD is computed as follows:

```

2416 minx<- min(trapmat[,1]-Cx)
2417 maxx<-max(trapmat[,1]-Cx)
2418 miny<- min(trapmat[,2]-Cy)
2419 maxy<- max(trapmat[,2]-Cy)
2420 # most extreme point determines maxD
2421 ul<- c(minx,maxy)
2422 maxD<- sqrt( (ul[1]-0)^2 + (ul[2]-0)^2)

```

2423 For the bear data the maxD was about 11.5 km. As such, the model de-  
 2424 scribed above will produce an estimate of the population size of bears within 11.5  
 2425 units of the trap centroid<sup>8</sup>. The BUGS model specification and R commands  
 2426 to package the data and fit the model are as follows:

```

2427 cat("
2428 model{
2429   p0 ~ dunif(0,1)          # prior distributions
2430   mup<- log(p0/(1-p0))
2431   psi~dunif(0,1)
2432   beta~dnorm(0,.01)
2433
2434   for(i in 1:(nind+nz)){
2435     xcent[i]~dunif(0,maxD)
2436     z[i]~dbern(psi)        # DA variables
2437     lp[i] <- mup + beta*xcent[i] # individual effect
2438     logit(p[i])<-lp[i]
2439     mu[i]<-z[i]*p[i]
2440     y[i]~dbin(mu[i],K)    # observation model
2441   }
2442   N<-sum(z[1:(nind+nz)])
2443 }
2444 ",file="modelMcov.txt")
2445 data2<-list(y=ytot,nz=nz,nind=nind,K=T,xcent=xcent,maxD=11.5)
2446 params2<-list('p0','psi','N','beta')
2447 inits = function() {list(z=z, psi=psi, p0=runif(1),beta=rnorm(1) ) }
2448 fit2 = bugs(data2, inits, params2, model.file="modelMcov.txt",working.directory=getwd(),
2449             debug=T, n.chains=3, n.iter=4000, n.burnin=1000, n.thin=4)

```

2450 Posterior summaries are given in Table ?? XYZ, and the posterior distribu-  
 2451 tion of  $N$  is given in Figure XYZ. It might be perplexing that the estimated  $N$

---

<sup>8</sup>To be convincing this might need a little bit of hand-holding

is much lower than obtained by model Mh 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 M0 and Mh 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 Mh. Moreover, importantly, using our new model, *the estimated N applies to an explicit area which is defined by our prescribed value of maxD*. 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 maxD from the centroid of the trap array. As such, the implied “effective trap area”<sup>9</sup> for any maxD is that of a circle with radius maxD.

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

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

## 4.5.2 Extension of the Model

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

<sup>9</sup>This is a bad use of this term. We have never defined ETA or ESA. What is it, exactly?

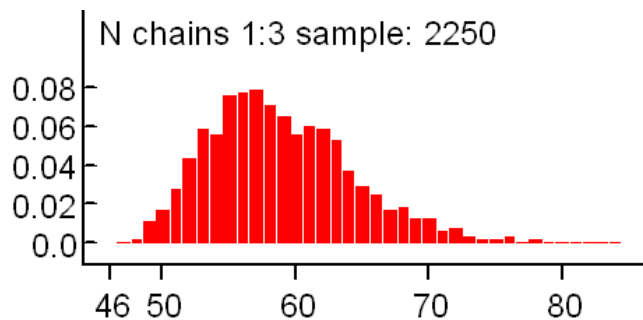


Figure 4.5: Needs a caption

whether we can choose a better distribution for  $\text{maxD}$  - one that doesn't imply a decreasing density as distance from the centroid increases. Conceptually, what we want to do is impose a prior on distance from the centroid,  $x$ , such that density is proportional to the amount of area in each successive distance band as you move farther away from the centroid. In fact, there is theory that exists which tells us what the correct distribution of  $x$  is  $2x/\text{maxD}^2$ . This can be derived by noting that  $F(x) = \Pr(X < x) = \pi i * x * x / \pi i * \text{maxD} * \text{maxD}$ . Then,  $f(x) = dF/dx = 2 * x / (\text{maxD} * \text{maxD})$ . This might be called a triangular distribution, I think, which makes sense because the incremental area in each additional distance band increases linearly with radius (i.e., distance from centroid). It is sometimes comforting to verify things empirically:

```
> u<-runif(10000,-1,1)
> v<-runif(10000,-1,1)
> d<- sqrt(u*u+v*v)
> hist(d[d<1])
> hist(d[d<1],100)
> hist(d[d<1],100,probability=TRUE)
> 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  $D_{\text{max}}$  into  $L$  distance classes of width  $\delta$ , with probabilities proportional to  $2 * x$ . In particular, if the cut-points are  $xg[1] = 0, xg[2], \dots, xg[L + 1] = D_{\text{max}}$  and the interval midpoints are  $xm[i] = xg[i + 1] - \delta$ . Then, the interval probabilities are  $p[i] = 2 * xm[i] * \delta / (D_{\text{max}} * D_{\text{max}})$ , 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

#### 4.5. INDIVIDUAL COVARIATE MODELS: TOWARD SPATIAL CAPTURE-RECAPTURE 79

be used as described previously. The new variable labeled `xround` is actually then the integer category label in units of  $\delta$  from 0. Thus, to convert back to distance in the expression for  $lp[i]$ , `xround[i]` has to be multiplied by  $\delta$ .

```

2526 delta<-.2
2527 xround<-xcent%%delta + 1
2528 Dgrid<- seq(delta,maxD,delta)
2529 xprobs<- delta*(2*Dgrid/(maxD*maxD))
2530 xprobs<-xprobs/sum(xprobs)
2531
2532 cat("
2533 model{
2534   p0 ~ dunif(0,1)          # prior distributions
2535   mup<- log(p0/(1-p0))
2536   psi~dunif(0,1)
2537   beta~dnorm(0,.01)
2538
2539   for(i in 1:(nind+nz)){
2540     xround[i]~dcat(xprobs[])
2541     z[i]~dbern(psi)          # zero inflation variables
2542     lp[i] <- mup + beta*xround[i]*delta # individual effect
2543     logit(p[i])<-lp[i]
2544     mu[i]<-z[i]*p[i]
2545     y[i]~dbin(mu[i],K)      # observation model
2546   }
2547
2548   N<-sum(z[1:(nind+nz)])
2549 }
2550 ",file="modelMcov.txt")

```

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

```

2553 data2<-list(y=ytot,nz=nz,nind=nind,K=T,xround=xround,xprobs=xprobs,delta=delta)
2554 params2<-list('p0','psi','N','beta')
2555 inits = function() {list(z=z, psi=psi, p0=runif(1),beta=rnorm(1) ) }
2556 fit = bugs(data2, inits, params2, model.file="modelMcov.txt",working.directory=getwd(),
2557           debug=FALSE, n.chains=3, n.iter=11000, n.burnin=1000, n.thin=2)

```

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

Table 4.4: Table: Analysis of Fort Drum bear hair snare data using the individual covariate model, for different values of  $D_{\max}$ , the upper limit of the uniform distribution of ‘distance from centroid of the trap array’

$\max D$	$mn$	$SD$	[1,]	12	0.230	0.038	[2,]	15	0.244	0.041	[3,]	17	0.249	0.044	[4,]	18	0.249	0.043	[5,]	19	0.250
----------	------	------	------	----	-------	-------	------	----	-------	-------	------	----	-------	-------	------	----	-------	-------	------	----	-------

We see that the posterior mean and SD of density (individuals per square km) appear insensitive to choice of  $\max D$  once we get a slight ways away from the maximum observed value of about 11.5. The estimated density of 0.250 per  $\text{km}^2$  is actually quite a bit lower than we reported using model Mh (0.37, see section XYZ above) for which sample area is not an explicit feature of the model. On the other hand it is higher than that reported from Model M0 using the buffered area (0.195). There is no basis really for comparing or contrasting these various estimates and 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 population<sup>10</sup> 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.

### 4.5.3 Invariance of density to $\max D$

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  $\max D$  in this case), we expect that density estimators should be invariant to this area. In the model used above, we note that  $\text{Area}(\max D) = \pi * \max D * \max D$  and  $E[N(\max D)] = \lambda * A(\max D)$  and thus  $E[\text{Density}(\max D)] = \lambda$  which is constant. This should be interpreted as the *prior* density. Absent data, then realizations under the model will have density  $\lambda$  regardless of what  $\max D$  is prescribed to be. As we verified empirically above, the posterior density is also invariant if  $\max D$  as long as the implied area (implied by  $\max D$ ) is large enough so that the data no longer provide information about density (i.e., “far away”), then our estimator of density should

<sup>10</sup>We 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.



become insensitive.

#### 4.5.4 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 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.

## 4.6 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, 2004b; 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 4.1. 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 4.1: 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  $F(x) = \Pr(X \leq x) = L * x / L * B = x/B$  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 data<sup>11</sup>. Thus, we are left with

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

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

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

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

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

#### 4.6.1 Example: Muntjac deer survey from Nagarahole, India

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

<sup>11</sup>Formally we could also say that  $[u] = \int [y|s][s]ds$

```

2698 library("R2WinBUGS")
2699 data<- read.csv("Muntjac.csv")
2700 nind<-nrow(data)
2701 y<-rep(1,nind)
2702 nz<-400
2703 y<-c(y,rep(0,nz))
2704 x<-data[,3]
2705 x<-c(x,rep(NA,nz))
2706 z<-y
2707 data<-list(y=y,x=x,nz=nz,nind=nind,B=150,striparea=708*.15*2)
2708
2709 cat("
2710 model{
2711   b~dunif(0,10)
2712   psi~dunif(0,1)
2713
2714   for(i in 1:(nind+nz)){
2715     z[i]~dbern(psi)      # DA Variables
2716     x[i]~dunif(0,B)      # B=strip width
2717     p[i]<-exp(logp[i])    # DETECTION MODEL
2718     logp[i]<- -((x[i]*x[i])*b)
2719     #logp[i]<- -b*log(x[i]+1)
2720     mu[i]<-z[i]*p[i]
2721     y[i]~dbern(mu[i])    # OBSERVATION MODEL
2722   }
2723   N<-sum(z[1:(nind+nz)])
2724   D<- N/striparea      # area of transects
2725 }
2726 ",file="dsamp.txt")

```

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

```

2729 params<-list('b','N','D','psi')
2730 inits = function() {list(z=z, psi=runif(1), b=runif(1,0,.02) )}
2731 fit = bugs(data, inits, params, model.file="dsamp.txt",
2732 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.<sup>12</sup>

	node	mean	sd	MC error	2.5%	median	97.5%	start	sample
D	1.096	0.1694	0.009122	0.8098	1.078	1.474	501	4500	
N	232.8	35.99	1.938	172.0	229.0	313.0	501	4500	

<sup>12</sup> 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.

```

2738 b 5.678E-4 1.05E-4 4.129E-6 3.867E-4 5.616E-4 7.949E-4 501 4500
2739 deviance 681.2 16.72 0.7536 650.8 680.6 716.6 501 4500
2740 psi 0.5099 0.08238 0.004442 0.3681 0.5033 0.6918 501 4500

```

## 2741 4.7 Summary and Outlook

2742 Traditional closed population capture-recapture models are closely related to  
 2743 binomial generalized linear models. Indeed, the only real distinction is that in  
 2744 capture-recapture models, the population size parameter  $N$  (corresponding also  
 2745 to the size of a hypothetical “complete” data set) is unknown. This requires  
 2746 special consideration in the analysis of capture-recapture models. The classi-  
 2747 cal approach to inference recognizes that the observations don’t have a stan-  
 2748 dard binomial distribution but, rather, a truncated binomial (from which which  
 2749 the so-called “conditional likelihood” derives) since we only have encounter fre-  
 2750 quency data on observed individuals. If instead we analyze the models using  
 2751 data augmentation, the observations can be modeled using a zero-inflated bino-  
 2752 mial distribution. In short, when we deal with the unknown- $N$  problem using  
 2753 data augmentation then we are left with zero-inflated GLM and GLMMs in-  
 2754 stead of ordinary GLM or GLMMs. The analysis of such zero-inflated models is  
 2755 practically convenient, especially using the various Bayesian analysis packages  
 2756 that use the BUGS language.

2757 Spatial capture-recapture models that we will consider in the rest of the  
 2758 chapters of this book are closely related to what have been called individual co-  
 2759 variate models. Heuristically, spatial capture-recapture models arise by defining  
 2760 individual covariates based on observed locations of individuals – we can think of  
 2761 using some function of mean encounter location as an individual covariate. We  
 2762 did this in a novel way, by using distance to the centroid of the trapping array  
 2763 as a covariate. We analyzed the “full likelihood” using data augmentation, and  
 2764 placed a prior distribution on the individual covariate which was derived from  
 2765 an assumption that individual locations are, a priori, uniformly distributed in  
 2766 space. This assumption provides for invariance of the density estimator to the  
 2767 choice of population size area (induced by maximum distance from the centroid  
 2768 of the). The model addressed some important problems in the use of closed pop-  
 2769 ulation models: it allows for heterogeneity in encounter probability due to the  
 2770 spatial context of the problem and it also provides a direct estimate of density  
 2771 because area is a feature of the model (via the prior on the individual covariate).  
 2772 The model is still not completely general because the model does not make use  
 2773 of the fully spatial encounter histories, which provide direct information about  
 2774 the locations and density of individuals. A specific individual covariate model  
 2775 that is in widespread use is classical “distance sampling.” The model underlying  
 2776 distance sampling is precisely a special kind of SCR model - but one without  
 2777 replicate samples. Understanding distance sampling and individual covariate  
 2778 models more broadly provides a solid basis for understanding and analyzing  
 2779 spatial capture-recapture models.



2780 **Chapter 5**

2781 **Fully Spatial**  
2782 **Capture-Recapture Models**





## Chapter 6

# 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.

## 6.1 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  $6 \times 4$  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” henceforth<sup>1</sup> so that we may use that model as a point of reference without having to provide a long-winded enumeration of

---

<sup>1</sup>RC: 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

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.

## 6.2 The binomial observation model

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

$$y_{ij} \sim \text{Bin}(K, p_{ij}) \quad (6.1)$$

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  $n \times m$  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:

$$\text{logit}(p_{ij}) = \alpha_0 + \theta * ||\mathbf{s}_i - \mathbf{x}_j|| \quad (6.2)$$

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

$$\log(p_{ij}) = \log(p_0) - \theta * ||\mathbf{s}_i - \mathbf{x}_j||^2 \quad (6.3)$$

Whatever model encounter probability we choose, we should always keep in mind that the model is described conditional on  $\mathbf{s}_i$ , which is an unobserved random

variable. Thus, to be precise about this, we should write the observation model as

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

Note that we probably expect that the parameter  $\theta$  in Eq. 6.2 or 6.3 should be negative, so that the probability of encounter decreases with distance between the trap and individual home range center. The joint likelihood for the data, conditional on the collection of individual activity centers, can therefore be expressed as

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

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

### 6.2.1 Distance as a latent variable

If we knew precisely every  $\mathbf{s}_i$  in the population (and how many,  $N$ ), then the model specified by eqs. 6.1 and 6.2 or 6.3 is just an ordinary logistic regression type of a model which we learned how to fit using **WinBUGS** previously (chapt. 2), with a covariate  $d_{ij}$ . However, the activity centers are unobservable even in the best possible circumstances. In that case,  $d_{ij}$  is an unobserved variable, analogous to classical “random effects” models. We need to therefore extend the model to accommodate these random variables with an additional model component. A standard, and perhaps not unreasonable, assumption is the so-called “uniformity assumption” which is to say that the  $\mathbf{s}_i$  are uniformly distributed over space (the obvious next question “which space?” is addressed below). This uniformity assumption amounts to a uniform prior distribution on  $\mathbf{s}_i$ , i.e., the pdf of  $\mathbf{s}_i$  is constant, which we may express

$$\Pr(\mathbf{s}_i) \text{propto} \text{const} \tag{6.4}$$

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

- (1) Observation model:  $y_{ij}|\mathbf{s}_i \sim \text{Bin}(K, p_{ij})$
- (2) Encounter probability:  $\text{logit}(p_{ij}) = \alpha_0 + \theta * \|\mathbf{s}_i - \mathbf{x}_j\|$
- (3) Point process model:  $\Pr[\mathbf{s}_i] \propto \text{const}$

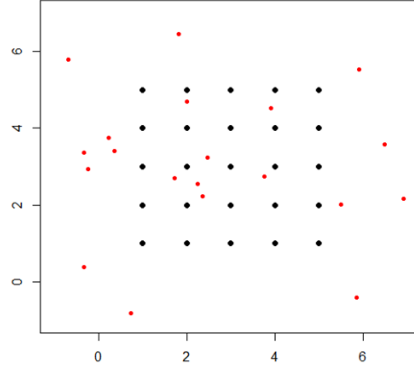


Figure 6.1: Realization of a binomial point process

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

## 6.3 The Binomial Point-process Model

The collection of individual activity centers  $\mathbf{s}_1, \dots, \mathbf{s}_N$  represent a realization of a *binomial point process* (Illian, 2008a, p. xyz). The binomial point process (BPP) is analogous to a Poisson point process in the sense that it represents a “random scatter” of points in space - except that the total number of points is *fixed*, whereas, in a Poisson point process it is random (having a Poisson distribution). As an example, we show in Fig. 6.1 locations of 20 individual activity centers (black dots) in relation to a grid of 25 traps. For a Poisson point process the number of such points in the prescribed state-space would be random whereas often we will simulate fixed numbers of points, e.g., for evaluating the performance of procedures 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. 6.1 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.

### 6.3.1 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.

### 6.3.2 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 11, 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 6.7 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 6.9) 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).

### 6.3.3 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 6.7 below. Note that  $\text{area}(\mathcal{S})$  affects data augmentation. If you increase  $\text{area}(\mathcal{S})$  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 9) we have to integrate the conditional-on-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;

### 6.3.4 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,



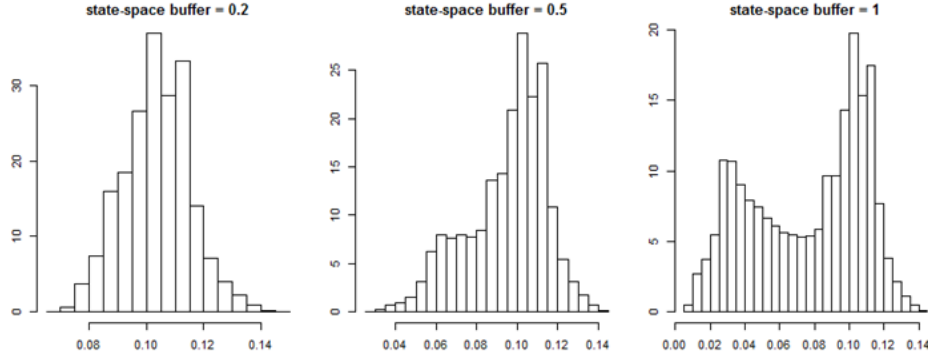


Figure 6.2: Needs a caption

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. 6.2 below which shows a histogram of  $p$  for a hypothetical population of 100000 individuals on a state-space enclosing our  $5 \times 5$  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  $(g, \mathcal{S})$ .

### 3094 6.3.5 Connection to Distance Sampling

3095 It is worth emphasizing that the basic SCR model is a binomial encounter model  
 3096 in which distance is a covariate. As such, it is striking similarity to a classical  
 3097 distance sampling model. Both have distance as a covariate but in classical  
 3098 distance sampling problems the focus is on the distance between the observer  
 3099 and the animal at an instant in time, not the distance between a trap and an  
 3100 animal's home range center. Thus in distance sampling, "distance" is *observed*  
 3101 for those individuals that appear in the sample. Conversely, in SCR problems,  
 3102 it is only imperfectly observed (we have partial information in the form of trap  
 3103 observations). Clearly, it is preferable to observe distance if possible, but as we  
 3104 will discuss in chapter XYZ, distance sampling requires field methods that are  
 3105 often not practical in many situations, e.g. when surveying tigers. Furthermore,  
 3106 SCR models allow us to relax many of the assumption made in classical distance  
 3107 sampling, and SCR models allow for estimates of quantities other than density,  
 3108 such as home range size.

## 3109 6.4 Simulating SCR Data

3110 It is always useful to simulate data because it allows you to understand the  
 3111 system that you're modeling and also calibrate your understanding with the  
 3112 parameter values of the model. That is, you can simulate data using differ-  
 3113 ent parameter values until you obtain data that "looks right" based on your  
 3114 knowledge of the specific situation that you're interested in. Here we provide  
 3115 a simple script to illustrate how to simulate spatial encounter history data. In  
 3116 this exercise we simulate data for 100 individuals and a 25 trap array laid out  
 3117 in a  $5 \times 5$  grid of unit spacing. The specific encounter model is the half-normal  
 3118 model given above and we used this code to simulate data used in subsequent  
 3119 analyses. The 100 activity centers were simulated on a state-space defined by  
 3120 a  $8 \times 8$  square within which the trap array was centered (thus the trap array  
 3121 is buffered by 2 units). Therefore, the density of individuals in this system is  
 3122 fixed at  $100/64$ .

```
3123 set.seed(2013)
3124 # create 5 x 5 grid of trap locations with unit spacing
3125 traplocs<- cbind(sort(rep(1:5,5)),rep(1:5,5))
3126 Dmat<-e2dist(traplocs,traplocs) # in cases where speed doesn't matter, it might be
3127                                # clearer to just show the slow for-loop.
3128                                # Plus, people will want to copy/paste this stuff
3129 ntraps<-nrow(traplocs)
3130
3131 # define state-space of point process. (i.e., where animals live).
3132 # "delta" just adds a fixed buffer to the outer extent of the traps.
3133 delta<-2
3134 Xl<-min(traplocs[,1] - delta)
3135 Xu<-max(traplocs[,1] + delta)
3136 Yl<-min(traplocs[,2] - delta)
```

```

3137 Yu<-max(traplocs[,2] + delta)
3138
3139 N<-100    # population size
3140 K<- 20    # number nights of effort
3141
3142 sx<-runif(N,Xl,Xu)    # simulate activity centers
3143 sy<-runif(N,Yl,Yu)
3144 S<-cbind(sx,sy)
3145 D<- e2dist(S,traplocs) # distance of each individual from each trap
3146
3147 alpha0<- -2.5        # define parameters of encounter probability
3148 sigma<- 0.5          #
3149 theta<- 1/(2*sigma*sigma)
3150 probcap<- expit(-2.5)*exp( - theta*D*D)    # probability of encounter
3151 # now generate the encounters of every individual in every trap
3152 Y<-matrix(NA,nrow=N,ncol=ntraps)
3153 for(i in 1:nrow(Y)){
3154   Y[i,]<-rbinom(ntraps,K,probcap[i,])
3155 }

```

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

```

3164 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.

```

3167 > names(data)
3168 [1] "Y"          "traplocs" "xlim"      "ylim"      "N"          "alpha0"    "beta"
3169 [8] "sigma"      "K"
3170 > Y<-data$Y
3171 > traplocs<-data$traplocs

```

### 6.4.1 Formatting and manipulating real data sets

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

- 3179 (1) The basic 2-dimensional data format, which is an `nind × ntraps` en-  
 3180 counter frequency matrix such as that simulated previously;
- 3181 (2) The maximally informative 3-dimensional array which we establish here  
 3182 the convention that it has dimensions `nind × nperiods × ntraps` and
- 3183 (3) We use a compact format - the “SCR flat format” - which we describe  
 3184 below in section 6.7.

3185 To simulate data in the most informative format - the “3-d array” - we can use  
 3186 the **R** commands given previously but replace the last 4 lines with the following:

```
3187 Y<-array(NA,dim=c(N,K,ntraps))
3188 for(i in 1:nrow(Y)){
3189   for(j in 1:ntraps){
3190     Y[i,1:K,j]<-rbinom(K,1,probcap[i,j])
3191   }
3192 }
```

3193 We see that a collection of  $K$  binary encounter events are generated for  
 3194 *each* individual and for *each* trap. The probabilities have those Bernoulli trials  
 3195 are computed based on the distance from each individuals home range center  
 3196 and the trap (see calculation above), and those are housed in the matrix prob-  
 3197 cap. Our data simulator function `simSRC0.fn` will return the full 3-d array if  
 3198 `array3d=TRUE` is specified in the function call. To recover the 2-d matrix from  
 3199 the 3-d array, and subset the 3-d array to individuals that were captured, we  
 3200 do this:

```
3201 Y2d<- apply(Y,c(1,3),sum) # sum over the ‘‘replicates’’ dimension (2nd margin of the array)
3202 ncaps<-apply(Y2d,1,sum)   # compute how many times each individual was captured
3203 Y<-Y[ncaps>0,,]          # keep those individuals that were captured
```

## 3204 6.5 Fitting an SCR Model in BUGS

3205 Clearly if we somehow knew the value of  $N$  then we could fit this model directly  
 3206 because, in that case, it is a special kind of logistic regression model - one with  
 3207 a random effect, but that enters into the model in a peculiar fashion - and also  
 3208 with a distribution (uniform) which we don’t usually think of as standard for  
 3209 random effects models. So our aim here is to analyze the known- $N$  problem,  
 3210 using our simulated data, as an incremental step in our progress toward fitting  
 3211 more generally useful models.

3212 To begin, we use our simulator to grab a data set and then harvest the  
 3213 elements of the resulting object for further analysis.

```
3214 data<-simSRC0.fn(discard0=FALSE,sd=2013)
3215 y<-data$Y
3216 traplocs<-data$traplocs
3217 nind<-nrow(y)
3218 X<-data$traplocs
```

```

3219 J<-nrow(X)
3220 y<-rbind(y,matrix(0,nrow=(100-nrow(y)),ncol=J ) )
3221 Xl<-data$xlim[1]
3222 Yl<-data$ylim[1]
3223 Xu<-data$xlim[2]
3224 Yu<-data$ylim[2]

```

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

```

3235 cat("
3236 model {
3237   alpha0~dnorm(0,.1)
3238   logit(p0)<- alpha0
3239   theta~dnorm(0,.1)
3240   for(i in 1:N){
3241     s[i,1]~dunif(Xl,Xu)
3242     s[i,2]~dunif(Yl,Yu)
3243     for(j in 1:J){
3244       d[i,j]<- pow(pow(s[i,1]-X[j,1],2) + pow(s[i,2]-X[j,2],2),0.5)
3245       y[i,j] ~ dbin(p[i,j],K)
3246       p[i,j]<- p0*exp(- theta*d[i,j]*d[i,j])
3247     }
3248   }
3249 }
3250 "
3251 ",file = "SCR0a.txt")

```

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

```

3258 sst<-cbind(runif(nind,Xl,Xu),runif(nind,Yl,Yu)) # starting values for s
3259 for(i in 1:nind){
3260   if(sum(y[i,])==0) next
3261   sst[i,1]<- mean( X[y[i,]>0,1] )
3262   sst[i,2]<- mean( X[y[i,]>0,2] )
3263 }
3264
3265 data <- list (y=y,X=X,K=K,N=nind,J=J,Xl=Xl,Yl=Yl,Xu=Xu,Yu=Yu)

```

```

3266 inits <- function(){
3267   list (alpha0=rnorm(1,-4,.4),theta=runif(1,1,2),s=sst)
3268 }
3269
3270 library("R2WinBUGS")
3271 parameters <- c("alpha0","theta")
3272 nthin<-1
3273 nc<-3
3274 nb<-1000
3275 ni<-2000
3276 out <- bugs (data, inits, parameters, "SCR0a.txt", n.thin=nthin,
3277 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:

```

3285 > print(out,digits=3)
3286 Inference for Bugs model at "SCR0a.txt", fit using WinBUGS,
3287 3 chains, each with 2000 iterations (first 1000 discarded)
3288 n.sims = 3000 iterations saved
3289
3290      mean      sd    2.5%    25%    50%    75%   97.5%  Rhat n.eff
3291 alpha0   -2.496  0.224  -2.954  -2.648  -2.48  -2.340  -2.091 1.013   190
3292 theta     2.442  0.419   1.638   2.145   2.44   2.721   3.303 1.005   530
3293 deviance 292.803 21.155 255.597 277.500 291.90 306.000 339.302 1.006   380
3294
3295 For each parameter, n.eff is a crude measure of effective sample size,
3296 and Rhat is the potential scale reduction factor (at convergence, Rhat=1).
3297
3298 DIC info (using the rule, pD = Dbar-Dhat)
3299 pD = -138.8 and DIC = 154.0
3300 DIC is an estimate of expected predictive error (lower deviance is better).

```

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

## 6.6 Unknown $N$

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

We typically have more than 4 traps and, if we’re fortunate, many more individuals in our data set.

For the augmented data, we introduce a set of binary latent variables (the data augmentation variables),  $z_i$ , and the model is extended to describe  $\Pr(z_i = 1)$  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. 10 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  $D = N/\text{area}(S)$ . For our simulator, we're using an  $8 \times 8$  state-space and thus we will compute  $D$  as  $D = N/64$ .

### 6.6.1 Analysis using data augmentation in WinBUGS

As before we begin by obtaining a data set using our `simSCRO.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  $\psi$ .

```

data<-simSCRO.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
  theta~dnorm(0,.1)
  psi~dunif(0,1)

  for(i in 1:M){
    z[i] ~ dbern(psi)
    s[i,1]~dunif(Xl,Xu)
    s[i,2]~dunif(Yl,Yu)
    for(j in 1:J){
      d[i,j]<- pow(pow(s[i,1]-X[j,1],2) + pow(s[i,2]-X[j,2],2),0.5)
      y[i,j] ~ dbin(p[i,j],K)
      p[i,j]<- z[i]*p0*exp(- theta*d[i,j]*d[i,j])
    }
  }
  N<-sum(z[])
  D<-N/64
}
",file = "SCR0a.txt")

```

To prepare our data we have to augment the data matrix `y` with  $M - n$



```

3401 all-zero encounter histories, we have to create starting values for the variables
3402  $z_i$  and also the activity centers  $s_i$  of which, for each, we require  $M$  values.
3403 Otherwise the remainder of the code for bundling the data, creating initial
3404 values and executing WinBUGS looks much the same as before except with
3405 more or differently named arguments.

3406 ## Data augmentation stuff
3407 M<-200
3408 y<-rbind(y,matrix(0,nrow=M-nind,ncol=ncol(y)))
3409 z<-c(rep(1,nind),rep(0,M-nind))
3410
3411 sst<-cbind(runif(M,Xl,Xu),runif(M,Yl,Yu)) # starting values for s
3412 for(i in 1:nind){
3413   if(sum(y[i,])==0) next
3414   sst[i,1]<- mean( X[y[i,]>0,1] )
3415   sst[i,2]<- mean( X[y[i,]>0,2] )
3416 }
3417 data <- list (y=y,X=X,K=K,M=M,J=J,Xl=Xl,Yl=Yl,Xu=Xu,Yu=Yu)
3418 inits <- function(){
3419   list (alpha0=rnorm(1,-4,.4),theta=runif(1,1,2),s=sst,z=z)
3420 }
3421
3422 library("R2WinBUGS")
3423 parameters <- c("alpha0","theta","N")
3424 nthin<-1
3425 nc<-3
3426 nb<-1000
3427 ni<-2000
3428 out <- bugs (data, inits, parameters, "SCR0a.txt", n.thin=ntin,n.chains=nc,
3429   n.burnin=nb,n.iter=ni,debug=TRUE,working.dir=getwd())

```

3430 **Remarks:** (1) Note the differences in this new **WinBUGS** model with  
3431 that appearing in the known- $N$  version. (2) Also the input data has changed  
3432 - the augmented data set has more rows of all-zeros. Previously we knew that  
3433  $N = 100$  but in this analysis we pretend not to know  $N$ , but think that  $N =$   
3434  $200$  is a good upper-bound; (3) Population size  $N(\mathcal{S})$  is a derived parameter,  
3435 being computed by summing up all of the data augmentation variables  $z_i$  (as  
3436 we've done previously); (4) Density,  $D \equiv D(\mathcal{S})$ , is also a derived parameter.  
3437 Summarizing the output from **WinBUGS** produces:

```

3438 > print(out1,digits=2)
3439 Inference for Bugs model at "SCR0a.txt", fit using WinBUGS,
3440 3 chains, each with 2000 iterations (first 1000 discarded)
3441 n.sims = 3000 iterations saved
3442
3443      mean    sd   2.5%   25%   50%   75%  97.5% Rhat n.eff
3444 alpha0  -2.57  0.23  -3.04  -2.72  -2.56  -2.41  -2.15  1.01   320
3445 theta    2.46  0.42   1.63   2.16   2.46   2.73   3.33  1.02   120
3446 N       113.62 15.73  86.00 102.00 113.00 124.00 147.00 1.01   260
3447 D        1.78  0.25   1.34   1.59   1.77   1.94   2.30  1.01   260
3448 deviance 302.60 23.67 261.19 285.47 301.50 317.90 354.91 1.00  1400

```

```

3449 For each parameter, n.eff is a crude measure of effective sample size,
3450 and Rhat is the potential scale reduction factor (at convergence, Rhat=1).
3451
3452 DIC info (using the rule, pD = var(deviance)/2)
3453 pD = 279.9 and DIC = 582.5
3454 DIC is an estimate of expected predictive error (lower deviance is better).

```

The column labeled “MC error” is the Monte Carlo error - the error inherent in the attempt to compute these posterior summaries by MCMC. It is desirable to run the Markov chain algorithm long enough so as to reduce the MC error to a tolerable level. What constitutes tolerable is up to the investigator. Certainly less than 1% is called for. As a general rule, Rhat gets closer to 1 and MC error decreases toward 0 as the number of iterations increases. We see that the estimated parameters ( $\alpha_0$  and  $\theta$ ) are comparable to the previous results obtained for the known-N case, and also not too different from the data-generating values. The posterior of  $N$  overlaps the data-generating value substantially with a mean of 113.62. To obtain these results we fitted the true data-generating model, that based on the half-normal detection model, to a single simulated data set. For fun and excitement we fit the *wrong* model - that with the logistic-linear detection model - to the same data set. This is easily achieved by modifying the **WinBUGS** model specification above, although we provide the **R** script in the **R** package **scrbook**. Those results are given below. We see that the estimate of  $N$ , the main parameter of interest, is very similar to that obtained under the correct model, convergence is worse (as measured by Rhat) 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 12.

```

3475 > print(out2,digits=2)
3476 Inference for Bugs model at "SCR0a.txt", fit using WinBUGS,
3477 3 chains, each with 2000 iterations (first 1000 discarded)
3478 n.sims = 3000 iterations saved
3479      mean    sd  2.5%   25%   50%   75%  97.5% Rhat n.eff
3480 alpha0  -1.59  0.27 -2.16 -1.77 -1.58 -1.42 -1.07 1.05   60
3481 beta    3.77  0.43  2.92  3.48  3.79  4.05  4.66 1.04   70
3482 N       122.57 18.67 90.00 109.00 122.00 135.00 163.00 1.00 3000
3483 D        1.92  0.29  1.41  1.70  1.91  2.11  2.55 1.00 3000
3484 deviance 312.67 22.43 271.00 297.20 311.50 327.00 359.60 1.02 130
3485
3486 For each parameter, n.eff is a crude measure of effective sample size,
3487 and Rhat is the potential scale reduction factor (at convergence, Rhat=1).
3488
3489 DIC info (using the rule, pD = var(deviance)/2)
3490 pD = 247.5 and DIC = 560.1
3491 DIC is an estimate of expected predictive error (lower deviance is better).

```

## 3492 6.6.2 Use of other BUGS engines: JAGS

3493 There are two other popular **BUGS** engines in widespread use: **OpenBUGS**  
 3494 (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`.

## 6.7 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. 6.3) 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. 6.1) 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. 6.6 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. 6.4.1 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
[1,]      1         2     127
[2,]      1         2     128
[3,]      1         2     129
[4,]      1        18     130
[5,]      2         3     106
```

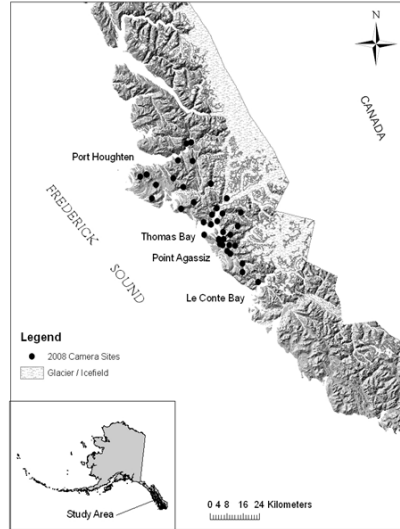


Figure 6.3: Wolverine camera trap locations from Magoun et al. (2011).

```

3537 [6,]      2      18     104
3538 [7,]      5       5      73
3539 [8,]      5       5      89
3540 [9,]      6      18     117
3541 [10,]     6      18     118

```

3542 This “encounter data file” contains 1 row for each unique individual/trap  
 3543 encounter, and 3 variables (columns): `trapid` is an integer that runs from  
 3544 `1:ntraps`, individual runs from `1:nind` and sample runs from `1:nperiods`.  
 3545 Often (as the case here) “sample” will correspond to daily sample intervals.  
 3546 The variable `trapid` will have to correspond to the row of a matrix containing  
 3547 the trap coordinates - a file named `traplocs.csv` available in the **R** package  
 3548 **scrbook**.

3549 Note that these data do not represent a completely informative summary  
 3550 of the data. For example, if no individuals were captured in a certain trap or  
 3551 during a certain period, then this compact data format will have no record.  
 3552 Thus we will need to know `ntraps` and `nperiods` when reformatting this SCR  
 3553 data format into a 2-d encounter frequency matrix or 3-d array. In addition, the  
 3554 encounter data file does not provide information about which periods each trap  
 3555 was operated. This additional information is also necessary as the trap-specific  
 3556 sample sizes must be passed to **BUGS** as data. We provide this information in  
 3557 a 2nd data file - which we call the “trap deployment” file (described below).

3558 The “encounter data file” `wcaps.csv` exists in the **R** package **scrbook** as a  
 3559 .csv file that people can read into **R** and do some basic summary statistics on.

For our purposes we need to convert these data into the “individual x trap” array of binary encounter frequencies, although more general models might require an encounter-history formulation of the model which requires a full 3-d array. To obtain our `nind x ntrap` encounter frequency matrix, we do this the hard way by first converting the encounter data file into a 3-d array and then summarize to trap totals. We have a handy function `SCR23darray.fn` which takes the compact encounter data file with optional arguments `ntraps` and `nperiods`, and converts it to a 3-d array, and then we use the **R** function `apply` to summarize over the “sample” period dimension (by convention here, this is the 2nd dimension):

```
SCR23darray.fn <- function(caps,ntraps=NULL,nperiods=NULL){
  nind<-max(caps[,2])
  if(is.null(ntraps)) ntraps<-max(caps[,1])
  if(is.null(nperiods)) nperiods<- max(caps[,3])
  y<-array(0,c(nind,nperiods,ntraps))
  tmp<-cbind(caps[,2],caps[,3],caps[,1])
  y[tmp]<-1
  y
}
```

# for the wolverine data do this:

```
Y3d <-SCR23darray.fn(wcaps,ntraps=37,nperiods=165)
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:

Trap	Easting	Northing	1	2	3	4	5	6	7	<- column names
1	39040	19216	0	0	0	0	0	0	0	
2	41324	19772	1	1	1	1	1	1	1	
3	44957	12985	0	0	0	0	0	0	0	
4	41151	23220	0	0	0	0	0	0	0	
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:

```

3606 traps<- read.csv("wtraps.csv")
3607 traplocs<- traps[,2:3]
3608 K<- apply(traps[,4:ncol(traps)],1,sum)

```

3609 This results in a matrix `traplocs` which contains the coordinates of each trap and  
 3610 a vector `K` containing the number of days that each trap was operational. We  
 3611 now have all the information required to fit a basic SCR model in **WinBUGS**.

3612 Summarizing these data files for the wolverine study, we see that 21 unique  
 3613 individuals were captured a total of 115 times. Most individuals were captured  
 3614 1-6 times, with 4, 1, 4, 3, 1, and 2 individuals captured 1-6 times, respectively.  
 3615 In addition, 1 individual was captured each 8 and 14 times and 2 individuals  
 3616 each were captured 10 and 13 times. The number of unique traps that captured  
 3617 a particular individual ranged from 1-6, with 5, 10, 3, 1, 1, and 1 individual cap-  
 3618 tured in each of 1-6 traps, respectively, for a total of 50 unique wolverine-trap  
 3619 encounters. These numbers might be hard to get your mind around whereas  
 3620 some tabular summary is often more convenient. For that it seems natural to  
 3621 tabulate individuals by trap and total encounter frequencies. The spatial infor-  
 3622 mation in SCR data is based on multi-trap captures, and so, it is informative to  
 3623 understand how many unique traps each individual is captured in. At the same,  
 3624 it is useful to understand how many total captures we have of each individual  
 3625 because this is, in an intuitive sense, the effective sample size. So, we reproduce  
 3626 Table 1 from Royle et al. (2011c) which shows the trap and total encounter  
 3627 frequencies:

Table 6.1: Individual frequencies of capture for wolverines captured in camera traps in Southeast 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	

### 3628 6.7.1 Fitting the model in WinBUGS

3629 For illustrative purposes here we fit the simplest SCR model with the half-  
 3630 normal distance function although we revisit these data with more complex  
 3631 models in later chapters. The model is summarized by the following 3 compo-  
 3632 nents:

- 3633 (1)  $y_{ij} | \mathbf{s}_i \sim \text{Bin}(K, z_i p_{ij})$   
 3634 (2)  $p_{ij} = p_0 \exp(-\theta \|\mathbf{s}_i - x_j\|^2)$   
 3635 (3)  $\mathbf{s}_i \sim \text{Unif}(\mathcal{S})$   
 3636 (4)  $z_i \sim \text{Bern}(\psi)$

3637 We assume customary flat priors on the structural (hyper-) parameters of the  
 3638 model,  $\alpha_0 = \text{logit}(p_0)$ ,  $\theta$  and  $\psi$ . It remains to define the state-space  $\mathcal{S}$ . For this,  
 3639 we nested the trap array (Fig. 6.3) in a rectangular state-space extending 20  
 3640 km beyond the traps in each cardinal direction. We also considered larger state-  
 3641 spaces up to 50 km to evaluate that choice. The buffer of the state space should  
 3642 be larger enough so that individuals beyond the state-space boundary are not  
 3643 likely to be encountered. Thus some knowledge of typical space usage patterns  
 3644 of the species is useful. The coordinate system was scaled so that a unit distance  
 3645 was equal to 10km, producing a rectangular state-space of dimension  $9.88 \times 10.5$   
 3646 units ( $\text{area} = 10374 \text{ km} \times \text{km}$ ) within which the trap array was nested. As a  
 3647 general rule, we recommend scaling the state-space so that it is defined near the  
 3648 origin  $(x, y) = (0, 0)$ . While the scaling of the coordinate system is theoretically  
 3649 irrelevant, a poorly scaled coordinate system can produce Markov chains that  
 3650 mix poorly. We fitted this model in **WinBUGS** using data augmentation with  
 3651  $M = 300$  potential individuals, using 3 Markov chains each of 12000 total  
 3652 iterations, discarding the first 2000 as burn-in. [R commands for reading in the  
 3653 data and executing the analysis are as follows:

3654 provide those commands here

3655 The output follows (note, we have a parameter “sigma” which we discuss shortly):

```
3656 Buffer = 10 km
3657 > print(out1$out,digits=2)
3658 Inference for Bugs model at "modelfile.txt", fit using WinBUGS,
3659 3 chains, each with 12000 iterations (first 2000 discarded)
3660 n.sims = 30000 iterations saved
3661      mean      sd    2.5%    25%    50%    75%   97.5%  Rhat  n.eff
3662 psi      0.11  0.02   0.07   0.10   0.11   0.13   0.17    1  2400
3663 sigma    1.79  0.29   1.31   1.58   1.75   1.97   2.46    1   600
3664 p0       0.03  0.00   0.02   0.03   0.03   0.03   0.04    1 13000
3665 N        33.02 4.99  25.00  29.00  32.00  36.00  44.00    1  1600
3666 D         4.93 0.75   3.73   4.33   4.78   5.38   6.57    1  1600
3667 beta     0.17 0.05   0.08   0.13   0.16   0.20   0.29    1   600
3668 deviance 441.97 11.49 421.50 434.00 441.20 449.20 466.30    1  6600
3669
3670
3671 Buffer = 20 km
3672 > print(out2$out,digits=2)
3673 Inference for Bugs model at "modelfile.txt", fit using WinBUGS,
3674 3 chains, each with 12000 iterations (first 2000 discarded)
3675 n.sims = 30000 iterations saved
```

```

3676          mean    sd   2.5%   25%   50%   75%  97.5% Rhat n.eff
3677 psi          0.16 0.04   0.10  0.13   0.16  0.18   0.24   1  4200
3678 sigma        1.78 0.32   1.29  1.55   1.73  1.94   2.56   1 20000
3679 p0           0.03 0.00   0.02  0.03   0.03  0.03   0.04   1  3000
3680 N            47.40 9.19  32.00 41.00  46.00 53.00  68.00   1  5900
3681 D            4.57 0.89   3.08  3.95   4.43  5.11   6.55   1  5900
3682 beta         0.17 0.06   0.08  0.13   0.17  0.21   0.30   1 20000
3683 deviance 444.36 11.84 423.60 436.00 443.60 451.80 469.70   1  1800
3684
3685 Buffer = 25 km
3686 > print(out3$out,digits=2)
3687 Inference for Bugs model at "modelfile.txt", fit using WinBUGS,
3688 3 chains, each with 12000 iterations (first 2000 discarded)
3689 n.sims = 30000 iterations saved
3690          mean    sd   2.5%   25%   50%   75%  97.5% Rhat n.eff
3691 psi          0.19 0.04   0.11  0.16   0.19  0.22   0.29 1.00   790
3692 sigma        1.80 0.34   1.30  1.56   1.75  1.98   2.59 1.01   400
3693 p0           0.03 0.00   0.02  0.03   0.03  0.03   0.04 1.00  2800
3694 N            56.66 11.47  37.00 48.00  56.00 64.00  82.00 1.00   570
3695 D            4.53 0.92   2.96  3.84   4.48  5.11   6.55 1.00   570
3696 beta         0.17 0.06   0.07  0.13   0.16  0.20   0.30 1.01   400
3697 deviance 444.75 11.87 423.60 436.40 444.00 452.30 469.80 1.00 24000
3698
3699 Buffer = 30 km
3700 > print(out4$out,digits=2)
3701 Inference for Bugs model at "modelfile.txt", fit using WinBUGS,
3702 3 chains, each with 12000 iterations (first 2000 discarded)
3703 n.sims = 30000 iterations saved
3704          mean    sd   2.5%   25%   50%   75%  97.5% Rhat n.eff
3705 psi          0.23 0.05   0.14  0.19   0.22  0.26   0.34 1.00  1500
3706 sigma        1.79 0.34   1.29  1.55   1.73  1.97   2.58 1.01   560
3707 p0           0.03 0.00   0.02  0.03   0.03  0.03   0.04 1.00 30000
3708 N            67.39 14.12  43.00 57.00  66.00 76.00  98.00 1.00  1200
3709 D            4.54 0.95   2.90  3.84   4.44  5.12   6.60 1.00  1200
3710 beta         0.17 0.06   0.07  0.13   0.17  0.21   0.30 1.01   560
3711 deviance 444.58 11.83 423.60 436.40 443.80 452.20 469.90 1.00  4700
3712
3713 Buffer = 40 km (need to add this)
3714
3715
3716
3717 Buffer = 45 km
3718 > print(out7$out,digits=2)
3719 Inference for Bugs model at "modelfile.txt", fit using WinBUGS,
3720 3 chains, each with 12000 iterations (first 2000 discarded)
3721 n.sims = 30000 iterations saved
3722          mean    sd   2.5%   25%   50%   75%  97.5% Rhat n.eff
3723 psi          0.36 0.08   0.21  0.30   0.35  0.41   0.53   1  5000
3724 sigma        1.78 0.34   1.29  1.55   1.72  1.95   2.60   1  850
3725 p0           0.03 0.00   0.02  0.03   0.03  0.03   0.04   1 3600

```



```

3726 N          106.57 23.34  67.00  90.00 104.00 121.00 157.00    1 3400
3727 D           4.62  1.01  2.90   3.90  4.51  5.25  6.81    1 3400
3728 beta        0.17  0.06  0.07   0.13  0.17  0.21  0.30    1  850
3729 deviance 444.80 11.84 423.60 436.40 444.10 452.30 470.00    1 30000
3730
3731 Buffer = 50 km
3732 > print(out8$out,digits=2)
3733 Inference for Bugs model at "modelfile.txt", fit using WinBUGS,
3734 3 chains, each with 12000 iterations (first 2000 discarded)
3735 n.sims = 30000 iterations saved
3736      mean    sd   2.5%   25%   50%   75%  97.5% Rhat n.eff
3737 psi       0.40  0.09   0.23   0.33   0.39   0.45   0.60 1.01  1300
3738 sigma     1.82  0.48   1.30   1.56   1.74   1.97   2.68 1.05   200
3739 p0        0.03  0.00   0.02   0.03   0.03   0.03   0.04 1.00  5800
3740 N        118.47 26.81  71.00 100.00 117.00 135.00 176.00 1.01  1200
3741 D         4.52  1.02   2.71   3.82   4.46   5.15   6.72 1.01  1200
3742 beta      0.17  0.06   0.07   0.13   0.17   0.21   0.30 1.05   200
3743 deviance 444.84 11.90 423.90 436.50 444.10 452.20 470.30 1.00   500

```

We see that the estimated density is roughly consistent as we increase the state-space buffer from 20 to 50 *km*. We do note that the data augmentation parameter  $\psi$  (and, correspondingly,  $N$ ) increase with the size of the state space in accordance with the deterministic relationship  $N = D * A$ . However, density is constant more or less as we increase the size of the state-space beyond a certain point. For the 10 *km* state-space buffer, we see a noticeable effect on the posterior distribution of  $D$ . This is not a bug but rather a feature. As we noted above, the state-space is part of the model.

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

## 6.7.2 Conclusion of Analysis

Our point estimate of wolverine density from this study of approximately 4.5 individuals/1000 *km*<sup>2</sup> and a 95% posterior interval is around [2.7, 6.3]. Density is estimated imprecisely which might not be surprising given the low sample size ( $n = 21$  individuals!). This seems to be a basic feature of carnivore studies although it should not (in our view) preclude the study of their populations nor attempts to estimate 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. 6.3). 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 XYZXYZ<sup>2</sup> 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.

## 6.8 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:

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

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

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

---

<sup>2</sup>raccoon example or something?

```

3804 xg<-seq(Xl,Xu,,50)
3805 yg<-seq(Yl,Yu,,50)

3806 Step 2: Extract the MCMC histories for the activity centers and the data
3807 augmentation variables. Note that these are each  $N \times \text{niter}$  matrices:

3808 Sxout<-out$sims.list$s[,1]
3809 Syout<-out$sims.list$s[,2]
3810 z<-out$sims.list$z

3811 Step 3: We associate each coordinate with the proper box using the R com-
3812 mand cut(). Note that we keep only the activity centers for which  $z = 1$  (i.e.,
3813 individuals that belong to the population of size  $N$ ):

3814 Sxout<-cut(Sxout[z==1],breaks=xg,include.lowest=TRUE)
3815 Syout<-cut(Syout[z==1],breaks=yg,include.lowest=TRUE)

3816 Step 4: Use the table() command to tally up how many activity centers are
3817 in each  $B(x)$ :

3818 Dn<-table(Sxout,Syout)

3819 Step 5: Use the image() command to display the resulting matrix.

3820 image(xg,yg,Dn/nrow(z),col=terrain.colors(10))

3821 Praise the Lord! This map is somewhat useful or at least it looks pretty and
3822 will facilitate the publication of your papers.

3823 It is worth emphasizing here that density maps will not usually appear uni-
3824 form despite that we have assumed that activity centers are uniformly dis-
3825 tributed. This is because the observed encounters of individuals provide direct
3826 information about the location of the  $i = 1, 2, \dots, n$  activity centers and thus
3827 their “estimated” locations will be affected by the observations. In a limiting
3828 sense, were we to sample space intensely enough, every individual would be
3829 captured a number of times and we would have considerable information about
3830 all  $N$  point locations. Consequently, the uniform prior would have almost no
3831 influence at all on the estimated density surface in this limiting situation. Thus,
3832 in practice, the influence of the uniformity assumption increases as the fraction
3833 of the population encountered decreases.

3834 On the non-intuitiveness of image() – the R function image() might
3835 not be very intuitive to some – it plots  $M[1,1]$  in the lower left corner. If you
3836 want  $M[]$  to be plotted “as you look at it” then  $M[1,1]$  should be in the upper
3837 left corner. We have a function rot() which does that. If you do image(rot(M))
3838 then it puts it on the monitor as if it was a map you were looking at. You can
3839 always specify the  $x$  and  $y$ – labels explicitly as we did above.

3840 Spatial dot plots – Now here is a cruder version based on the “spatial
3841 dot map” function spatial.plot. The useful functions in R are image() and
3842 image.scale() which is a function we grabbed off the web somewhere. Use of
3843 this function requires arguments of point locations and the resulting value to be
3844 displayed. The function is defined and applied as follows:

```

```

3845 spatial.plot<- function(x,y){
3846   nc<-as.numeric(cut(y,20))
3847   plot(x,pch=" ")
3848   points(x,pch=20,col=topo.colors(20)[nc],cex=2)
3849   image.scale(y,col=topo.colors(20))
3850 }
3851 # To execute the function do this:
3852 spatial.plot(cbind(xg,yg), Dn/nrow(z))

```

### 3853 6.8.1 Example: Wolverine density map.

3854 We used the posterior output from the wolverine model fitted previous to com-  
 3855 pute a relatively coarse version of a density map, using a  $10 \times 10$  grid (Fig. 6.4)  
 3856 and using a  $30 \times 30$  grid (Fig. 6.5)<sup>3</sup>. In these figures density is expressed in  
 3857 units of individuals per 1000  $km^2$ , while the area of the pixels is about 1037  
 3858  $km^2$  and 115  $km^2$ , respectively. That calculation is based on<sup>4</sup>:

```

3859 > total.area<- (Yu-Yl)*(Xu-Xl)*1000
3860 > total.area/(10*10)
3861 [1] 1037.427
3862 > total.area/(30*30)
3863 [1] 115.2697

```

3864 A couple of things are worth noting: First is that as we move away from  
 3865 “where the data live” - away from the trap array - we see that the density  
 3866 approaches the mean density. This is a property of the estimator as long as  
 3867 the “detection function” decreases sufficiently rapidly as a function of distance.  
 3868 Relatedly, it is also a property of statistical smoothers such as splines, kernel  
 3869 smoothers, and regression smoothers - predictions tend toward the global mean  
 3870 as the influence of data diminishes. Another way to think of it is that it is a  
 3871 consequence of the prior - which imposes uniformity, and as you get far away  
 3872 from the data, the predictions tend to the prior. The other thing to note about  
 3873 this map is that density is not 0 over water (although the coastline is not shown).  
 3874 This might be perplexing to some who are fairly certain that wolverines do not  
 3875 like water. However, there is nothing about the model that recognizes water  
 3876 from non-water and so the model predicts over water *as if* it were habitat similar  
 3877 to that within which the array is nested. But, all of this is ok as far as estimating  
 3878 density goes and, furthermore, we can compute valid estimates of  $N$  over any  
 3879 well-defined region which presumably wouldn’t include water if we so choose.

## 3880 6.9 Discrete State-Space

3881 The SCR model developed previously in this chapter assumes that individual ac-  
 3882 tivity centers are distributed uniformly over the prescribed state-space. Clearly

<sup>3</sup>Note: Not sure if we should use quantiles for color to make equal area slices. ??? Also should we use the same scale?

<sup>4</sup>This is wrong and needs fixed. Move decimal one place over. i.e., 100 instead of 1000.

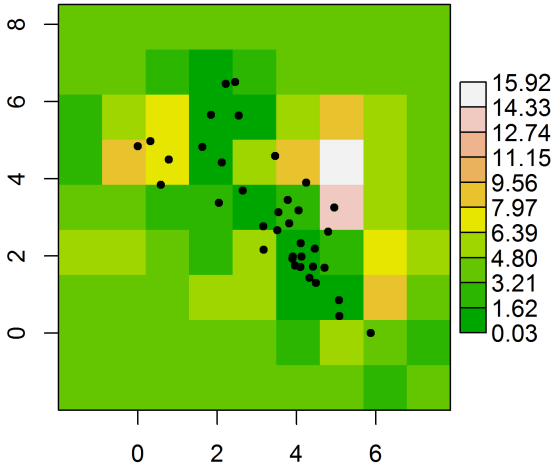


Figure 6.4: Needs a caption

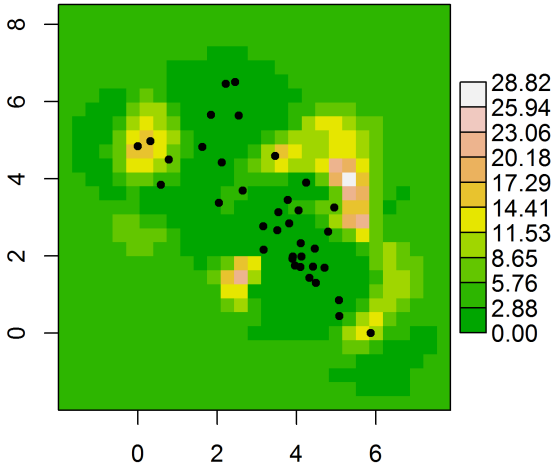


Figure 6.5: Needs a caption

this will not always be a reasonable assumption. In chapter 15 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 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 10 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 10. 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 9, 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).

### 6.9.1 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 6.3.4 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 9) 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  $6 \times 6$ ,  $9 \times 9$ ,  $12 \times 12$ ,  $15 \times 15$ ,  $20 \times 20$ ,  $25 \times 25$  and  $30 \times 30$  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 XYZ<sup>5</sup>.

Table XYZ.

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

Note: WinBUGS based on fewer samples too!

To get SE and time-series SE do this:

<sup>5</sup>Andy to finish later

3971 You can use `as.mcmc.list()` to convert to a coda object. Then use `summary`.

3972 The results in terms of the posterior summaries are, as we expect, very  
 3973 similar using **WinBUGS**. However, it was interesting to note that **WinBUGS**  
 3974 runtime is much worse (note the number of iterations is lower for **WinBUGS**  
 3975 yet the runtime is much longer) and, furthermore, it seems to scale with the size  
 3976 of the discrete state-space grid. While that was expected, it was unexpected  
 3977 that the runtime of **JAGS** would seem relatively consistent as we increase the  
 3978 grid size. We suspect that **WinBUGS** is evaluating the full-conditional for  
 3979 each activity center at all  $G$  possible values whereas it may be that **JAGS**  
 3980 is evaluating the full-conditional only at a subset of values or perhaps using  
 3981 previous calculations more effectively.

3982 While this might suggest that one should always use **JAGS** for this analy-  
 3983 sis, we found in our analysis of the wolverine (next section) that **JAGS** could  
 3984 be extremely sensitive to starting values, producing MCMC algorithms that  
 3985 sometimes simply did not work.

## 3986 6.9.2 Analysis of the wolverine camera trapping data

3987 We reanalyzed the wolverine data using discrete state-space grids with points  
 3988 spaced by 2, 4 and 8 km (depicted in Fig. ??). These were constructed from  
 3989 the 40 km buffered state-space, and deleting the points over water (see Royle  
 3990 et al., 2011c). Our interest in doing this was to evaluate the relative influence  
 3991 of grid resolution on estimated density because the coarser grids will be more  
 3992 efficient from a computational stand-point and so we would prefer to use them,  
 3993 but perhaps not if there is a strong influence on estimated density.

3994 **Note:** Results from WinBUGS are given below based on short runs that  
 3995 took a long long time. I am rerunning those. I will also show a density map for  
 3996 each analysis.

3997 based on 2k burn 3k total and 3 chains = 3k total posterior samples.  
 3998 lots of MC error here.

3999  
 4000 2km

4001 For each parameter, `n.eff` is a crude measure of effective sample size,  
 4002 and `Rhat` is the potential scale reduction factor (at convergence, `Rhat=1`).

4003		mean	sd	2.5%	25%	50%	75%	97.5%	Rhat	n.eff
4004	psi	0.28	0.06	0.17	0.24	0.27	0.32	0.41	1.01	230
4005	sigma	0.64	0.05	0.55	0.60	0.64	0.67	0.73	1.02	88
4006	lam0	-3.00	0.16	-3.33	-3.11	-3.00	-2.90	-2.69	1.04	52
4007	p0	0.05	0.01	0.03	0.04	0.05	0.05	0.06	1.04	52
4008	N	82.95	16.26	55.00	72.00	82.00	93.00	119.02	1.01	240

4009  
 4010 4 km

4011 For each parameter, `n.eff` is a crude measure of effective sample size,  
 4012 and `Rhat` is the potential scale reduction factor (at convergence, `Rhat=1`).

4013		mean	sd	2.5%	25%	50%	75%	97.5%	Rhat	n.eff
4014	psi	0.30	0.06	0.19	0.26	0.29	0.34	0.43	1.01	580



```

4015 sigma 0.62 0.05 0.54 0.59 0.62 0.65 0.72 1.00 2000
4016 lam0 -3.00 0.16 -3.33 -3.10 -2.99 -2.90 -2.67 1.01 390
4017 p0 0.05 0.01 0.03 0.04 0.05 0.05 0.06 1.01 390
4018 N 88.78 16.76 60.00 77.00 87.00 99.00 125.00 1.01 690
4019
4020 8km
4021 For each parameter, n.eff is a crude measure of effective sample size,
4022 and Rhat is the potential scale reduction factor (at convergence, Rhat=1).
4023 mean sd 2.5% 25% 50% 75% 97.5% Rhat n.eff
4024 psi 0.27 0.06 0.17 0.23 0.27 0.31 0.40 1.00 1500
4025 sigma 0.69 0.05 0.60 0.65 0.68 0.72 0.80 1.00 3000
4026 lam0 -3.07 0.17 -3.41 -3.20 -3.07 -2.95 -2.74 1.01 210
4027 p0 0.04 0.01 0.03 0.04 0.04 0.05 0.06 1.01 200
4028 N 82.01 15.98 55.00 71.00 80.00 92.00 118.00 1.00 1300

```

```

4029 We did the analysis in JAGS also. The results are shown below. Note: I
4030 am going to run these again but for longer to finalize the results.

```

```

4031 2km
4032 Iterations = 7001:13000
4033 Thinning interval = 1
4034 Number of chains = 3
4035 Sample size per chain = 6000
4036
4037 Mean SD Naive SE Time-series SE
4038 N 86.28522 16.950626 1.263e-01 0.4878973
4039 lam0 0.04807 0.007512 5.599e-05 0.0002199
4040 p0 0.04581 0.006820 5.083e-05 0.0001996
4041 psi 0.28904 0.062117 4.630e-04 0.0017481
4042 sigma 0.62769 0.043596 3.249e-04 0.0018724
4043
4044 4km
4045 Mean SD Naive SE Time-series SE
4046 N 85.53139 16.998966 1.267e-01 0.5181297
4047 lam0 0.04636 0.007542 5.621e-05 0.0002382
4048 p0 0.04425 0.006867 5.118e-05 0.0002172
4049 psi 0.28650 0.061922 4.615e-04 0.0018276
4050 sigma 0.64281 0.048321 3.602e-04 0.0022911
4051
4052 8km
4053 Mean SD Naive SE Time-series SE
4054 N 83.97039 16.508146 1.230e-01 0.4548782
4055 lam0 0.04519 0.006919 5.157e-05 0.0001738
4056 p0 0.04319 0.006319 4.710e-05 0.0001589
4057 psi 0.28146 0.060653 4.521e-04 0.0016555
4058 sigma 0.66956 0.040989 3.055e-04 0.0015070

```

### 6.9.3 SCR models as multi-state models

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

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

## 6.10 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 11 – 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 9.

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 12). 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 12. 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

4147 main reason is that these conventional methods are predicated on models that  
4148 represent explicit misspecifications of both the observation and ecological pro-  
4149 cess - they are wrong! Not just wrong, because of course all models are wrong,  
4150 but they're not even *plausible* models! Thus while we might be able to show  
4151 adequate fit or whatever, we think as a conceptual and philosophical model one  
4152 should not be using models that are not even plausible data-generating models  
4153 - even if the plausible ones don't fit! Perhaps more charitably, these ordinary  
4154 non-spatial models are models of the wrong system. They do not account for  
4155 trap identity. They don't account for spatial organization or "clustering" of in-  
4156 dividual encounters in space. And, "density" is not a parameter of those models  
4157 because density has no meaning absent an explicit representation of space. If  
4158 we do define space explicitly, e.g., as a buffered minimum convex hull, then the  
4159 normal models ( $M_0$ ,  $M_h$ , etc..) assume that individual capture-probability is  
4160 not related to space, no matter how we define the buffer. Conversely, the SCR  
4161 model is a model for trap-specific encounter data - how individuals are organized  
4162 in space and interact with traps. SCR models provide a coherent framework for  
4163 inference about density or population size and also, because of the formality of  
4164 their derivation, can be extended and generalized to a large variety of different  
4165 situations, as we demonstrate in subsequent chapters.

4166 In the next few chapters we continue to work with this basic SCR design and  
4167 model but consider some important extensions of the basic model. For example,  
4168 we consider extensions to include covariates that vary by individual, trap, or over  
4169 time (chapter 13), spatial covariates on density (chapter 15), open populations  
4170 (chapter 16), model assessment and selection (chapter 12) and other topics. We  
4171 also consider technical details of Bayesian (chapter 10) and maximum likelihood  
4172 (chapter 9) estimation so that the interested reader can develop or extend their  
4173 own methods to suit their needs.

## 4174 Chapter 7

## 4175 Other observation models



## 4176 Chapter 8

# 4177 Maximum likelihood 4178 estimation





## Chapter 9

# Likelihood Analysis of SCR 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 a formulation based on data augmentation. We consider likelihood analysis of SCR models in the context

of the wolverine camera trapping study (Magoun et al., 2011) we analyzed in previous chapters to compare/contrast the results.

## 9.1 Likelihood analysis

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

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

When viewed as a function of  $\theta$  for purposes of estimation, the marginal distribution  $[y|\theta]$  is often referred to as the *integrated likelihood*.

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

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

$$y_{ij}|\mathbf{s}_i \sim \text{Bin}(K, p_{\theta}(\mathbf{x}_j, \mathbf{s}_i)) \quad (9.1)$$

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

$$[\mathbf{y}_i|\mathbf{s}_i, \theta] = \prod_j \text{Bin}(K, p_{\theta}(\mathbf{x}_j, \mathbf{s}_i))$$

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

The so-called “marginal likelihood” is computed by removing  $\mathbf{s}_i$ , by integration, from the conditional-on- $\mathbf{s}$  likelihood. That is, we compute:

$$[y|\theta] = \int_{\mathcal{S}} \mathcal{L}(\theta|\mathbf{y}_i|\mathbf{s}_i)g(\mathbf{s}_i)d\mathbf{s}_i$$

here  $g(s) = 1/|\mathcal{S}|$ .

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

$$\prod_i f(y[i, j])$$

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

The key operation for computing the likelihood is solving a 2-dimensional integration problem. There are some general purpose **R** packages that implement a number of multi-dimensional integration routines including **adapt** (Genz et al., 2007) and **R2cuba** (Hahn et al., 2011). In practice, we won’t rely on these extraneous **R** packages but instead will use perhaps less efficient methods in which we replace the integral with a summation over an equal area mesh of points on the state-space  $\mathcal{S}$  and explicitly evaluate the integrand at each point. We invoke the rectangular rule for integration here<sup>1</sup> in which we evaluate the integrand on a regular grid of points of equal area and compute the average of the integrand over that grid of points. Let  $u = 1, 2, \dots, nG$  index a grid of  $nG$  points,  $\mathbf{s}_u$ , where the area of grid cell  $u$  is  $A$ . In this case, the integrand, i.e., the marginal probability of  $y_{ij}$ , is approximated by

$$f(y_{ij}|\theta) = \frac{1}{nG} \sum_{u=1}^{nG} f(\mathbf{y}_i|\mathbf{s}_u) \quad (9.2)$$

This is a specific case of the general expression that could be used for approximating the integral for any arbitrary bivariate distribution  $g(u)$ . The general case is

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

In the present context note that  $[\mathbf{s}] = (1/A)$  and thus the grid-cell area cancels in the above expression to yield eq. 9.2.

Not surprisingly this the same answer we get if  $\mathcal{S}$  were inherently discrete, having  $nG$  unique values with equal probabilities  $1/nG$ , and we apply the Law of Total Probability directly to compute the marginal probability  $[y|\theta]$ .

### 9.1.1 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 **simSCR0.fn** we simulate data for 100 individuals and a 25 trap array layed out in a  $5 \times 5$  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  $8 \times 8$  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:

<sup>1</sup>e.g., [http://en.wikipedia.org/wiki/Rectangle\\_method](http://en.wikipedia.org/wiki/Rectangle_method)

```

4276 data<-simSCRO.fn(discard0=FALSE,sd=2013)
4277 y<-data$Y
4278 traplocs<-data$traplocs
4279 nind<-nrow(y)
4280 X<-data$traplocs
4281 J<-nrow(X)
4282 K<-data$K
4283 Xl<-data$xlim[1]
4284 Yl<-data$ylim[1]
4285 Xu<-data$xlim[2]
4286 Yu<-data$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):

```

4289 delta<- .2
4290 xg<-seq(Xl+delta/2,Xu-delta/2,by=delta)
4291 yg<-seq(Yl+delta/2,Yu-delta/2,by=delta)
4292 npix<-length(xg)           # assumes xg and yg same dimension here
4293 area<- (Xu-Xl)*(Yu-Yl)/((npix)*(npix)) # dont need area for anything
4294 G<-cbind(rep(xg,npix),sort(rep(yg,npix)))
4295 nG<-nrow(G)

```

In this case, the integration grid is set up as a grid with spacing  $\delta = 0.2$  which produces a  $40 \times 40$  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 various quantities associated with the integration grid **G** and *nG*. 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  $\delta$  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:

```

4309
4310 intlik1<-function(parm,y=y,delta=.2,X=traplocs,ssbuffer=2){
4311
4312   Xl<-min(X[,1]) - ssbuffer
4313   Xu<-max(X[,1]) + ssbuffer
4314   Yu<-max(X[,2]) + ssbuffer
4315   Yl<-min(X[,2]) - ssbuffer
4316
4317   xg<-seq(Xl+delta/2,Xu-delta/2,,length=npix)
4318   yg<-seq(Yl+delta/2,Yu-delta/2,,length=npix)
4319   npix<-length(xg)

```

```

4320
4321 G<-cbind(rep(xg,npix),sort(rep(yg,npix)))
4322 nG<-nrow(G)
4323 D<- e2dist(X,G)
4324
4325 alpha0<-parm[1]
4326 alpha1<-parm[2]
4327 probcap<- plogis(alpha0)*exp(-alpha1*D*D)
4328 Pm<-matrix(NA,nrow=nrow(probcap),ncol=ncol(probcap))
4329           # all zero encounter histories
4330 n0<-sum(apply(y,1,sum)==0)
4331           # encounter histories with at least 1 detection
4332 ymat<-y[apply(y,1,sum)>0,]
4333 ymat<-rbind(ymat,rep(0,ncol(ymat)))
4334 lik.marg<-rep(NA,nrow(ymat))
4335 for(i in 1:nrow(ymat)){
4336   Pm[1:length(Pm)]<- (dbinom(rep(ymat[i,],nG),K,probcap[1:length(Pm)],log=TRUE))
4337   lik.cond<- exp(colSums(Pm))
4338   lik.marg[i]<- sum( lik.cond*(1/nG))
4339 }
4340 nv<-c(rep(1,length(lik.marg)-1),n0)
4341 -1*( sum(nv*log(lik.marg)) )
4342 }

```

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 ( $n0$ ). 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. 9.1 above). The marginal likelihood (`lik.marg`) sums up the conditional elements weighted by  $\Pr(s)$  (formula XXX above). Finally, this function assumes that  $K$ , the number of replicates, is constant for each trap. Further, it assumes that the state-space is a square. As an exercise, consider resolving these two issues by generalizing the code.

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:

```

4364 # should take 15-30 seconds
4365

```

```

4366 > starting.values <- c(-2, 2)
4367 > frog<-nlm(intlik1,starting.values,y=y,delta=.1,X=traplocs,ssbuffer=2,hessian=TRUE)
4368 > frog
4369
4370 $minimum
4371 [1] 297.1896
4372
4373 $estimate
4374 [1] -2.504824  2.373343
4375
4376 $gradient
4377 [1] -2.069654e-05  1.968754e-05
4378
4379 $hessian
4380      [,1]      [,2]
4381 [1,]  48.67898 -19.25750
4382 [2,] -19.25750  13.34114
4383
4384 $code
4385 [1] 1
4386
4387 $iterations
4388 [1] 11

```

4389 Details about this output can be found on the help page for `nlm`. We note  
 4390 briefly that `frog$minimum` is the negative log-likelihood value at the MLEs,  
 4391 which are stored in the `frog$estimate` component of the list. The hessian is  
 4392 the observed Fisher information matrix, which can be inverted to obtain the  
 4393 variance-covariance matrix using the commands:

```

4394 > solve(frog$hessian)

```

4395 It is worth drawing attention to the fact that the estimates are different  
 4396 than the Bayesian estimates reported in the previous chapter (section XYZ)!!!  
 4397 How can that be?! There are several reasons for this. First Bayesian inference  
 4398 is based on the posterior distribution and it is not generally the case that the  
 4399 MLE should correspond to any particular value of the posterior distribution. If  
 4400 the prior distributions in a Bayesian analysis are uniform, then the mode of the  
 4401 posterior is the MLE, but note that Bayesians almost always report posterior  
 4402 means and so there will typically be a discrepancy there. Secondly, we have  
 4403 implemented an approximation to the integral here and there might be a slight  
 4404 bit of error induced by that. We will evaluate that shortly. Third, the Bayesian  
 4405 analysis by MCMC is subject to some amount of Monte Carlo error which the  
 4406 analyst should always be aware of in practical situations. All of these different  
 4407 explanations are likely responsible for some of the discrepancy. Accounting  
 4408 for these, as a practical matter, we see general consistency between the two  
 4409 estimates.

4410 To compute the integrated likelihood we used a discrete representation of  
 4411 the state-space so that the integral could be approximated as a summation

over possible values of  $\mathbf{s}$  with each value being weighted by its probability of occurring, which is  $1/nG$  under the assumption that  $\mathbf{s}$  is uniform on the state-space  $\mathcal{S}$ . In chapter 4 we used a discrete state-space in developing a Bayesian analysis of the model in order to be able to modify the state-space in a flexible manner. Bayesian analysis requires simulation of the point process conditional on the observations, and this can be a difficult task when the state-space is continuous but has irregular geometry. Conversely, if the state-space is a regular polygon then Bayesian analysis by MCMC is possibly more efficient with a continuous state-space. We emphasize that the state-space is a part of the model. In some cases there wont be a natural choice of state space beyond “some large rectangle containing the trap grid” and, in such cases, for regular detection functions the estimate of density is invariant to the size of the state-space (i.e., the buffer) as long as it is sufficiently large. However if there are good reasons to restrict the state-space, it will tend to have an influence on the likelihood and hence AIC and so forth. As an illustration, lets do that by changing the state space here.....Use my polygon clipping stuff .....

In summary, we note that, for the basic SCR model, integrated likelihood is a really easy calculation when  $N$  is known. Even for  $N$  unknown it is not too difficult, and we will do that shortly. However, if you can solve the known- $N$  problem then you should be able to do a real analysis, for example by considering different values of  $N$  and computing the results for each value and then making a plot of the log-likelihood or AIC and choosing the value of  $N$  that produces the best log likelihood or AIC. As a homework problem we suggest that the reader take the code given above and try to estimate  $N$  without modifying the code by just repeatedly calling that code for different values of  $N$  and trying to deduce the best value. Nevertheless, we will formalize the unknown- $N$  problem shortly. We note that the software package **DENSITY** (Efford et al., 2004) implements certain types of SCR models using integrated likelihood methods. **DENSITY** has been made into an **R** package called **secr** (Efford, 2011) and we provide an analysis of some data using **secr** shortly along with a discussion of its capabilities.

## 9.2 MLE when N is Unknown

Here we build on the previous introduction to integrated likelihood but we consider 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 dont 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

4455 must be included in the likelihood.

4456 **DETAILS NEEDED HERE**

4457 To summarize, when  $N$  is unknown, the  $n$  observed encounter histories have  
 4458 a multinomial distribution with probabilities  $\pi(i)$  and sample size  $N^2$ . The last  
 4459 cell the “zero cell” is computed by carrying out the integral in expression XYZ  
 4460 above for the all-zero encounter history and we have to account for the fact that  
 4461 there are  $n_0 = N - n$  such encounter histories.

4462 To analyze a specific case, we'll read in our fake data set (simulated using the  
 4463 parameters given above). To set some things up in our workspace we do this:

```
4464 data<-simSCR0.fn(discard0=TRUE,sd=2013)
4465 y<-data$Y
4466 nind<-nrow(y)
4467 X<-data$traplocs
4468 J<-nrow(X)
4469 K<-data$K
4470 Xl<-data$xlim[1]
4471 Yl<-data$ylim[1]
4472 Xu<-data$xlim[2]
4473 Yu<-data$ylim[2]
```

4474 Recall that these data were generated with  $N = 100$ , on an  $8 \times 8$  unit  
 4475 state-space representing the trap locations ( $\mathbf{X}$ ) buffered by 2 units. As before,  
 4476 the likelihood is defined in the **R** workspace as an **R** function which takes an  
 4477 argument being the unknown parameters of the model and additional arguments  
 4478 as prescribed. In particular, as before, we provide the encounter history matrix  
 4479  $\mathbf{y}$ , the trap locations  $\text{traplocs}$ , the spacing of the integration grid ( $\delta$ ) and the  
 4480 state-space buffer. Here is the new likelihood function:

```
4481 intlik2<-function(parm,y=y,delta=.3,X=traplocs,ssbuffer=2){
4482
4483   Xl<-min(X[,1]) -ssbuffer
4484   Xu<-max(X[,1]) + ssbuffer
4485   Yu<-max(X[,2]) + ssbuffer
4486   Yl<-min(X[,2]) - ssbuffer
4487
4488   #delta<- (Xu-Xl)/npix
4489   xg<-seq(Xl+delta/2,Xu-delta/2,delta)
4490   yg<-seq(Yl+delta/2,Yu-delta/2,delta)
4491   npix.x<-length(xg)
4492   npix.y<-length(yg)
4493   area<- (Xu-Xl)*(Yu-Yl)/((npix.x)*(npix.y))
4494   G<-cbind(rep(xg,npix.y),sort(rep(yg,npix.x)))
4495   nG<-nrow(G)
4496   D<- e2dist(X,G)
```

---

<sup>2</sup> Maybe you could show an alternative simulation script to generate data using the `rmulti-`  
`nom` function. This would make it a little more clear for people



```

4497
4498 alpha0<-parm[1]
4499 alpha1<-parm[2]
4500 n0<-exp(parm[3])
4501 probcap<- plogis(alpha0)*exp(-alpha1*D*D)
4502 Pm<-matrix(NA,nrow=nrow(probcap),ncol=ncol(probcap))
4503 ymat<-rbind(y,rep(0,ncol(y)))
4504
4505 lik.marg<-rep(NA,nrow(ymat))
4506 for(i in 1:nrow(ymat)){
4507   Pm[1:length(Pm)]<- (dbinom(rep(ymat[i,],nG),K,probcap[1:length(Pm)],log=TRUE))
4508   lik.cond<- exp(colSums(Pm))
4509   lik.marg[i]<- sum( lik.cond*(1/nG) )
4510 }
4511 nv<-c(rep(1,length(lik.marg)-1),n0)
4512 part1<- lgamma(nrow(y)+n0+1) - lgamma(n0+1)
4513 part2<- sum(nv*log(lik.marg))
4514   -1*(part1+ part2)
4515 }

```

4516 To execute this function for the data that we created with `simSCRO.fn`, we  
 4517 execute the following command (saving the result in our friend `frog`). This  
 4518 results in the usual output, including the parameter estimates, the gradient,  
 4519 and the numerical Hessian which is useful for obtaining asymptotic standard  
 4520 errors (see below):

```

4521 > frog<-nlm(intlik2,c(-2.5,2,log(4)),hessian=TRUE,y=y,X=X,delta=.2,ssbuffer=2)
4522 There were 50 or more warnings (use warnings() to see the first 50)
4523 >
4524 >
4525 > frog
4526 $minimum
4527 [1] 113.5004
4528
4529 $estimate
4530 [1] -2.538334  2.466515  4.232810
4531
4532 [1. Additional output deleted .]

```

4533 While this produces some **R** warnings, these happen to be harmless in this case,  
 4534 and we will see from the `nlm` output that the algorithm performed satisfactory  
 4535 in minimizing the objective function. The estimate of population size for the  
 4536 state-space (using the default state-space buffer) is

```

4537 > nrow(y)+exp(4.2328)
4538 [1] 110.9099

```

4539 Which differs from the data-generating value ( $N = 100$ ) as we might expect. We  
 4540 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:

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

### 9.2.1 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.

### 9.2.2 Integrated Likelihood using the model under data augmentation

Note that this likelihood analysis is based on the standard likelihood in which  $N$  (or  $n_0$ ) is an explicit parameter. This is usually called the “joint likelihood” or “unconditional likelihood”. We could also express the joint likelihood using data augmentation, replacing the parameter  $N$  with  $\psi$  (e.g., Royle and Dorazio, 2008, sec. xyz). We don’t go into detail here, but we do note that the likelihood under data augmentation is a zero-inflated binomial mixture precisely as an occupancy type model (Royle, 2006). The interested reader could adapt the material from Royle and Dorazio (2008) with the **R** code given above for the likelihood and implement the likelihood analysis based on the model under data augmentation. While we can carryout likelihood analysis of models under data augmentation, we primarily advocate data augmentation for Bayesian analysis.

### 9.2.3 Extensions

There are other types of covariates of interest: behavioral response, sex-specificity of parameters and all of these things. Some of these can be added directly to the likelihood if the covariate is fixed and known for all individuals captured or not. This excludes most covariates but it does include behavioral response. Sex-specificity is more difficult since sex is not known for uncaptured individuals. Trap-specific covariates such as trap type or status, or time-specific covariates such as date, are relatively easy to deal with (we leave these as exercises). We apply these various models in Chapter XXXX. To analyze such models, we do Bayesian analysis of the joint likelihood facilitated by the use of data augmentation. For covariates that are not fixed and known for all individuals, it is hard to do MLE for these based on the joint likelihood as we have developed above. Instead what people normally do is use what is colloquially referred to as the “Huggins-Alho” type model which is one of the approaches taken in the software package `secr` (Efford, 2011, see sec. 9.5).

## 9.3 Classical model selection and assessment

In most analyses, one is interested in choosing from among various potential models. A good thing about classical analysis based on likelihood is we can do rote application of AIC without thinking about anything. With distance as a covariate (e.g., distance sampling) this is usually applied to some arbitrary selection of distance functions. We don't recommend this. Given there is hardly ever (if at all) a rational science-based reason for choosing some particular distance function we believe that this standard approach will invariably lead to over-fitting. The fact that AIC is easy to compute does not mean that it should be abused in such fashions. Further discussion is made in chapters XYZ.

**Goodness-of-fit** In many analyses based on likelihood methods it is possible to cook-up fit statistics for which asymptotic distributions are known. In general, however, applied statisticians tend to adopt bootstrapping based on heuristically appealing fit statistics. An omnibus global GoF statistic is not so obvious but we can apply bootstrapping principles to SCR models directly which we discuss in chapter XYZ. Bayesian goodness-of-fit is almost always addressed with Bayesian p-values or some other posterior predictive check (REF XXX). Thus the approach whether Bayesian or classical is the same. We identify a fit statistic, we do a bootstrap (classical) or a Bayesian p-value. Royle et al. (2011) decomposed the fit problem into separate evaluations of the CSR hypothesis and the encounter process model. We discuss all of this in Chapter XYZ.

## 4616 9.4 Likelihood analysis of the wolverine camera 4617 trapping data

4618 Here we compute the MLEs for the wolverine data using an expanded version of  
4619 the function we developed in the previous section. To accommodate that each  
4620 trap might be operational a variable number of nights, we provided an additional  
4621 argument to the likelihood function (allowing for a vector  $K$ ), which requires  
4622 also a modification to the construction of the likelihood. In addition, we had to  
4623 accommodate that the state-space is a general rectangle, and we included a line  
4624 in the code to compute the state-space area which we apply below for computing  
4625 density. The more general function (`intlik3`) is given in the **R** package. It has  
4626 a general purpose wrapper named `scr` which has other capabilities too.

4627 The data were read into our R session and manipulated using the following  
4628 commands. Note that we use the utility **R** function `SCR23darray.fn` which we  
4629 defined in chapt. 4.

```
4630 > wcaps<-source("wcaps.R")$value
4631 > wtraps<-source("wtraps.R")$value
4632 > K.wolv<-apply(wtraps[,4:ncol(wtraps)],1,sum)
4633 >
4634 > xx<-SCR23darray.fn(wcaps,ntraps=37,nperiods=165)
4635 > y.wolv<- apply(xx,c(1,3),sum)
4636 > traplocs.wolv<-wtraps[,2:3]
4637 > traplocs.wolv<-traplocs.wolv/10000
4638 >
4639 > frog<-nlm(intlik3,c(-1.5,1.2,log(4)),hessian=TRUE,y=y.wolv,K=K.wolv,X=traplocs.wolv,
4640 There were 23 warnings (use warnings() to see them)
4641 > frog
4642
4643 $minimum
4644 [1] 220.4355
4645
4646 $estimate
4647 [1] -2.817570  1.255112  3.599040
4648
4649 $gradient
4650 [1] -6.274309e-06  2.146722e-05 -1.045566e-05
4651
4652 $hessian
4653           [,1]      [,2]      [,3]
4654 [1,]  37.687931 -11.852236  4.688911
4655 [2,] -11.852236  30.846144 -9.199113
4656 [3,]  4.688911  -9.199113 13.050428
4657
4658 $code
4659 [1] 1
```

```
4660
4661 $iterations
4662 [1] 12
4663
4664 > exp(3.599)*sqrt(solve(frog$hessian)[3,3])
4665 [1] 11.41059
4666 >
4667
```

4668 We obtained the MLEs for a state-space buffer of 2 (standardized units) and  
4669 for integration grid with spacing  $\delta = .3, .2, .1, .05$ . The MLEs for these 4 cases  
4670 including the relative runtime are given in Table 9.1.

Table 9.1: Run time and MLEs for different integration grid resolutions.

$\delta$	runtime	Estimates		
		$\alpha_0$	$\theta$	$\log(n_0)$
0.30	8.4	-2.819786	1.258468	3.569731
0.20	22.6	-2.817610	1.254757	3.583690
0.10	99.0	-2.817570	1.255112	3.599040
0.05	403.0	-2.817559	1.255281	3.607158

4671 We see the results change only slightly as the fineness of the integration  
4672 grid increases. Conversely, the runtime on the platform of the day for the 4  
4673 cases increases rapidly which, as we have suggested before, could probably be  
4674 regarded in relative terms, across platforms, for gaging the decrease in speed  
4675 as the fineness of the integration grid increases. The effect of this is that we  
4676 anticipate some numerical error in approximating the integral on a mesh of  
4677 points, and that error increases as the coarseness of the mesh increases.

4678 In section 6.9 back in chapt. 4 we used a discrete representation of the state-  
4679 space in order to have control over its extent and shape, for example so that we  
4680 could clip out “non-habitat”. Clearly that formulation of the model is relevant  
4681 to the use of integrated likelihood in the sense that such a representation of  
4682 the state-space underlies the computation of the integral. Thus, for example,  
4683 we could easily compute the MLE of parameters under some model with a  
4684 restricted state-space merely by creating the required state-space at whatever  
4685 grid resolution is desired, and then feed that state-space into the likelihood  
4686 evaluation above. The **R** function `scr` which comes with the **R** package for this  
4687 book accommodates an arbitrary state-space fashioned in this manner, as well  
4688 as state-spaces created by polygons or GIS shapefiles<sup>3</sup>.

4689 Next we studied the effect of the state-space buffer on the MLEs, using a  
4690 fixed  $\delta = .2$  for all analyses. We used state-space buffers of 1 to 4 units stepped  
4691 by .5. This produced the following results, given here are the state-space buffer,  
4692 area of the state-space, the MLE of N for the prescribed state-space and the  
4693 corresponding MLE of density:

---

<sup>3</sup>to be completed!

		ssbuff	Ass	Nhat	Dhat
4694					
4695	[1,]	1.0	66.98212	37.73338	0.5633352
4696	[2,]	1.5	84.36242	46.21008	0.5477567
4697	[3,]	2.0	103.74272	57.00617	0.5494956
4698	[4,]	2.5	125.12302	69.03616	0.5517463
4699	[5,]	3.0	148.50332	82.17550	0.5533580
4700	[6,]	3.5	173.88362	96.44018	0.5546249
4701	[7,]	4.0	201.26392	111.83524	0.5556646

4702 The estimates of  $D$  stabilize rapidly and the incremental difference is within the  
 4703 numerical error associated with approximating the integral. The results suggest  
 4704 that wolverine density is around 0.56 individuals per 100  $km^2$  (recall that a  
 4705 state-space unit is  $10 \times 10 km$ ). This is about 5.6 individuals per thousand  $km^2$   
 4706 which compares with XYZ-lookup-XYZ reported in Royle et al. (2011c) based  
 4707 on a clipped state-space as described in section XYZ (XYZ chapter 4 XYZ).

#### 4708 9.4.1 Exercises

- 4709 1. Compute the 95% confidence interval for wolverine density, somehow.
- 4710 2. Compute the AIC of this model and modify `intlik3` to consider alternative
- 4711 link functions (at least one additional) and compare the AIC of the different
- 4712 models and the estimates. Comment.

### 4713 9.5 Program DENSITY and the R package secr

4714 **DENSITY** is a software program developed by Efford (2004) for fitting spatial  
 4715 capture-recapture models based mostly on classical maximum likelihood esti-  
 4716 mation and related inference methods. Efford (2011) has also released an **R**  
 4717 package named **secr**, that contains many of the functions within **DENSITY**  
 4718 but also incorporates new models and features. Here, we will focus on **secr** as  
 4719 it will continue to be developed, contains more functionality and is based in **R**.  
 4720 To install and run models in **secr**, you must download the package and load  
 4721 it in **R**.

```
4722 > install.packages(secr)
4723 > library(secr)
```

4724 **secr** allows the user to simulate data and fit a suite of models with various  
 4725 detection functions and covariate responses. **secr** uses the standard **R** model  
 4726 specification framework using tildes. E.g., the model command is `secr.fit` and  
 4727 is generally written as

```
4728 > secr.fit(capturedata, model = list(D~1, g0~1, sigma~1), buffer = 20000)
```

4729 where we have `g0~1` indicating the intercept model. To include covariates, this  
 4730 would be written as `g0~b` where  $b$  is a behavioral response covariate. Possible

4731 predictors for detection probability include both pre-defined variables (e.g., **t**  
 4732 and **b** corresponding to “time” and “behavior”), and user-defined covariates of  
 4733 several kinds. The discussion of covariates is developed in chapter XX(8)<sup>4</sup>

4734 Before we can fit the models, the data must first be entered into **secr**. Two  
 4735 input files are required: trap layout (location and identification information for  
 4736 each trap) and capture data (e.g., sampling session, animal identification, trap  
 4737 day, and trap location). SECR requires that you specify the trap type, the  
 4738 two most common for camera trapping/hair snares are proximity detectors and  
 4739 count detectors. The ‘proximity’ detector type allows, at most, one detection  
 4740 of each individual at a particular detector on any occasion. The count detector  
 4741 designation allows repeat encounters of each individual at a particular detector  
 4742 on any occasion. There are other detector types that one can select such as:  
 4743 ‘polygon’ detector type which allows for a trap to be a sampled polygon, e.g.,  
 4744 scat surveys, and ‘signal’ detector which allows for traps that have a strength  
 4745 indicator, e.g., acoustic arrays. The detector types single and multi can be  
 4746 confusing as multi seems like it would appropriate for something like a camera  
 4747 trap, but instead these two designations refer to traps that retain individuals,  
 4748 thus precluding the ability for animals to be captured in other traps during  
 4749 the sampling occasion. The single type indicates trap that can only catch one  
 4750 animal at a time, while multi indicates traps that may catch more than one  
 4751 animal at a time. For a full review of the detector types, one should look at the  
 4752 help manual, which can be accessed in R after installing the SECR package by  
 4753 using the command:

```
4754 > RShowDoc("secr-manual", package = "secr")
```

4755 As with all of the scr models, **secr** fits a detection function relating the probabil-  
 4756 ity of detection to the distance of a detector from an individual activity center.  
 4757 **secr** allows the user to specify one of a variety of detection functions including  
 4758 the commonly used half-normal, hazard rate, and exponential. There are 12 dif-  
 4759 ferent functions, but some are only available for simulating data, and one should  
 4760 take caution when using different detection functions as the interpretation of  
 4761 the parameters, such as sigma, may not be consistent across formulations. The  
 4762 different detection functions are defined in the secr manual and can be found  
 4763 by calling the help function for the detection function:

```
4764 > ?detectfn
```

4765 It is useful to note that **secr** requires the buffer distance to be defined in meters  
 4766 and density will be returned as number of animals per hectare. Thus to make  
 4767 comparisons between **secr** and other models, we will often have to convert the  
 4768 density to the same units. Also, note that sigma is returned in units of meters.

5

---

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

<sup>5</sup>One 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 doesnt explain this, does he? In some sections in chapter 4 or possibly 6 we get into this issue.

### 4770 9.5.1 Analysis using the secr package

4771 To demonstrate the use of the secr package, we will show how to do the same  
 4772 analysis on the wolverine study as shown in section 4.6. To use the secr package,  
 4773 the data need to be formatted in a similar but slightly different manner than we  
 4774 use in WinBUGS. After installing the secr package, we first have to read in the  
 4775 trap locations and other related information, such as if the trap is operational  
 4776 during a sampling occasion. The secr package reads in the trap data through  
 4777 a command called read.traps, which requires the detector type as input. The  
 4778 detector type is important because it will determine the likelihood that secr  
 4779 will use to fit the model. Here, we have selected proximity since individuals are  
 4780 captured at most once in each trap during each sampling occasion.

```
4781 > traps= read.csv(wtraps.csv)
4782 > colnames(traps)[1:3]<- c("trapID","x", "y") #name the first 3 columns
4783 # to match the secr nomenclature
4784
4785 > trapfile <- read.traps(data = traps, detector = "proximity")
```

4786 After reading in the data, we now need to create the encounter matrix or  
 4787 array. The secr package does this through the use of the make.caphist com-  
 4788 mand, where we provide the capture histories in raw data format (each line  
 4789 contains the session, identification number, occasion, and trap id for only 1  
 4790 individual). This is the format that was shown in the data input file wcaps,  
 4791 and we only need a line or two to organize the data into the order that the  
 4792 make.caphist command wants. In creating the capture history, we provide also  
 4793 the trapfile with the trap information, and the format (e.g., here fmt= trapID)  
 4794 so that secr knows how to match the encounters to the trap, and finally, we  
 4795 provide the number of occasions.

```
4796 > wolv.dat <- wcaps[,c(2, 3, 1)]
4797 #NEED TO UPDATE THIS WHEN I GET THE FILES,
4798 ### I JUST GUESSED AT THE CODE, BUT WOULD LIKE TO TRY IT.
4799 > wolv.dat <- cbind(rep(1, dim(wolv.dat)[1]), wolv.dat)
4800 > colnames(wolv.dat) <- c("Session", "ID", "Occasion", "trapID")
4801
4802 > wolvcapt=make.caphist(wolv.dat, trapfile, fmt = "trapID", noccasions = 165)
```

4803 Calling the secr.fit command, will run the model. We are using the basic  
 4804 model (SCR0), so we do not need to make any specifications in the command  
 4805 line except for the providing the buffer size (in m). To specify different models,  
 4806 you can change the default  $D \sim 1$ ,  $g0 \sim 1$ ,  $\sigma \sim 1$ , which the interested reader  
 4807 can do with very little difficulty.

```
4808 > wolv.secr=secr.fit(wolvcapt, model = list(D~1, g0~1, sigma~1), buffer = 20000)
4809
4810 > wolv.secr
```



```

4811 secr.fit( capthist = wolvcapt, buffer = 20000, binomN = 1 )
4812 secr 2.0.0, 18:26:39 05 Jul 2011
4813
4814
4815 Detector type      proximity
4816 Detector number    37
4817 Average spacing    4415.693 m
4818 x-range            593498 652294 m
4819 y-range            6296796 6361803 m
4820 N animals          : 21
4821 N detections        : 115
4822 N occasions        : 165
4823 Mask area          : 1037069 ha
4824
4825 Model              : D~1 g0~1 sigma~1
4826 Fixed (real)       : none
4827 Detection fn       : halfnormal
4828 Distribution        : poisson
4829 N parameters        : 3
4830 Log likelihood      : -746.754
4831 AIC                 : 1499.508
4832 AICc                : 1500.920
4833
4834 Beta parameters (coefficients)
4835      beta      SE.beta      lcl      ucl
4836 D      -9.749576 0.23027860 -10.200913 -9.298238
4837 g0     -4.275736 0.15846104 -4.586313 -3.965158
4838 sigma  8.699202 0.07868944  8.544973  8.853430
4839
4840 Variance-covariance matrix of beta parameters
4841      D      g0      sigma
4842 D      0.053028233 0.000546922 -0.005226926
4843 g0      0.000546922 0.025109900 -0.005885213
4844 sigma -0.005226926 -0.005885213 0.006192027
4845
4846 Fitted (real) parameters evaluated at base levels of covariates
4847      link      estimate SE.estimate      lcl      ucl
4848 D      log 5.831941e-05 1.360973e-05 3.713638e-05 9.158548e-05
4849 g0     logit 1.371121e-02 2.142902e-03 1.008756e-02 1.861207e-02
4850 sigma  log 5.998124e+03 4.727205e+02 5.140849e+03 6.998355e+03

```

4851 Under the fitted (real) parameters, we find  $D$ , the density, given in units  
4852 of individuals/hectare (1 hectare = 100 m<sup>2</sup>). To convert this into individ-  
4853 uals/1000km<sup>2</sup>, we multiply by 100000, thus our density estimate is 5.83 individ-  
4854 uals/1000 km<sup>2</sup>. Sigma is given in units of meters, to convert to kilometers, we  
4855 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. 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.

## 9.6 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 in the equivalent **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 et al. (XYZ) 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, **WinBUGS** is extremely flexible in terms of describing models, although it sometimes can be quite slow. We can devise models in **WinBUGS** 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 **WinBUGS** 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 **WinBUGS** 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 XYZ) and generating predictions of how many individuals have home range centers in any particular polygon.

4899 Basic benefits of Bayesian analysis have been discussed elsewhere (Chapter 2?  
4900 BPA book? Link and Barker?) and we believe these are compelling. On the  
4901 other hand, likelihood analysis makes it easy to do model-selection by AIC.  
4902 Goodness-of-fit is probably no more difficult or easy under either paradigm (see  
4903 next chapter?).

4904 In summary, basic SCR models are easy to implement by either likelihood  
4905 or Bayesian methods but we feel that the typical user will realize much more  
4906 flexibility in model development using existing platforms for Bayesian analysis.  
4907 While these tend to be slow (sometimes excruciatingly slow), this will probably  
4908 not be an impediment in most problems, especially at some near point in the  
4909 future. Since we spent a lot of time here talking about specific technical details  
4910 on how to implement likelihood analysis of SCR models, we provided a corre-  
4911 sponding treatment in the next chapter on how to devise MCMC algorithms  
4912 for SCR models. This is a bit more tedious and requires more coding, but is  
4913 not technically challenging (except perhaps to develop highly efficient algorithms  
4914 which we don't excel at).



## 4915 Chapter 10

## 4916 MCMC details



## Chapter 11

# MCMC Details

### 11.1 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.

#### 11.1.1 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.

## 11.2 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  $\theta$ , 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  $\theta$  is that it can be used to make probability statements about  $\theta$ , such as the probability that  $\theta$  is equal to some value, or the probability that  $\theta$  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  $\theta$  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 most practical applications, it is nearly impossible to directly compute the posterior. Recall Bayes theorem:

$$p(\theta|y) = p(y|\theta) * p(\theta)/p(y), \quad (11.1)$$



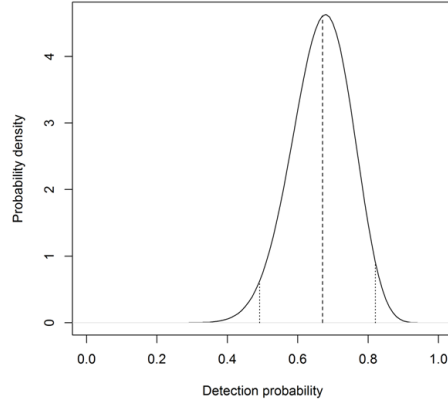


Figure 11.1: Probability density plot of a hypothetical posterior distribution of  $\text{beta}(20,10)$ ; dashed lines indicate mean and upper and lower 95% interval

where  $\theta$  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  $\theta$ ,  $p(\theta)$  the prior probability of  $\theta$ , 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  $\sigma$  is known and our objective is to obtain an estimate of  $\mu$  using Bayesian statistics. To fully specify the model in a Bayesian framework, we first have to define a prior distribution for  $\mu$ . 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  $\mu_0$  and  $\sigma_0^2$  are fixed, the posterior for  $\mu$  has the following form (for the algebraic proof, see XXX):

$$\mu|y \sim \text{Normal}(\mu_n, \sigma_n^2) \quad (11.2)$$

5003 where

$$\mu_n = (sig^2/sig^2 + n * sig0^2) * mu0 + (n * sig0^2/sig^2 + n * sig0^2) * y - bar$$

5004 And

$$sign^2 = sig^2 * sig0^2 / (sig^2 + n * sig0^2)$$

5005 We can directly obtain estimates of interest from this Normal posterior distri-  
 5006 bution, such as the mean mu-hat and its variance; we do not need to apply  
 5007 MCMC, since we can recognize the posterior as a parametric distribution, in-  
 5008 cluding the normalizing constant  $p(y)$ . But generally we will be interested in  
 5009 more complex models with several, say  $n$ , parameters. In this case, computing  
 5010  $p(y)$  from Eq. 11.1 requires  $n$ -dimensional integration, which is can be difficult  
 5011 or impossible. Thus, the posterior distribution in generally only known up to a  
 5012 constant of proportionality:

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

5013 The power of MCMC is that it allows us to approximate the posterior using simu-  
 5014 lation without evaluating the high dimensional integrals and to directly sample  
 5015 from the posterior, even when the posterior distribution is unknown! The price  
 5016 is that MCMC is computationally expensive. Although MCMC first appeared  
 5017 in the scientific literature in 1949 (Metropolis and Ulam, 1949), widespread use  
 5018 did not occur until the 1980s when computational power and speed increased  
 5019 (Gelfand and Smith, 1990). It is safe to say that the advent of practical MCMC  
 5020 methods is the primary reason why Bayesian inference has become so popular  
 5021 during the past three decades. In a nutshell, MCMC lets us generate sequential  
 5022 draws of  $\theta$  (the parameter(s) of interest) from distributions approximating the  
 5023 unknown posterior over  $T$  iterations. The distribution of the draw at  $t$  depends  
 5024 on the value drawn at  $t-1$ ; hence, the draws from a Markov chain.<sup>1</sup> As  $T$  goes  
 5025 to infinity, the Markov chain converges to the desired distribution in our case  
 5026 the posterior distribution for  $\theta|y$ . Thus, once the Markov chain has reached  
 5027 its stationary distribution, the generated samples can be used to characterize  
 5028 the posterior distribution,  $p(\theta|y)$ , and point estimates of  $\theta$ , its standard error  
 5029 and confidence bounds, can be obtained directly from this approximation of the  
 5030 posterior. In practice, although we know that a Markov chain will eventually  
 5031 converge, we can only generate a limited number of samples a process that  
 5032 depending on the model can be quite time consuming. Assessing whether our  
 5033 Markov chain has indeed converged is an important part of MCMC sampling  
 5034 and we will speak about some common diagnostics in Section XX.

## 5035 11.3 Types of MCMC sampling

5036 There are several MCMC algorithms, the most popular being Gibbs sampling  
 5037 and Metropolis-Hastings sampling. We will be dealing with these two classes in

<sup>1</sup>In 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,  $\theta(t-1)$ .

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.

### 11.3.1 Gibbs sampling

Gibbs sampling was named after the physicist J.W. Gibbs by Geman and Geman (1984), who applied the algorithm to a Gibbs distribution<sup>2</sup>. 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 sampling<sup>3</sup>. 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  $\mu$ , the posterior distribution of  $\mu|y$  is also Normal. Conversely, with a fixed (known)  $\mu$ , but unknown variance, the conjugate prior for  $\sigma^2$  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:

$$\text{sig}|\mu, y \sim \text{InvGamma}(an, bn), \quad (11.3)$$

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

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

Or

$$p(\mu, \text{sig}|y) \propto p(y|\mu) * p(\mu) * p(\text{sig})$$

This cannot easily be reduced to a distribution we recognize. However, we can condition  $\mu$  on  $\text{sig}$  (i.e., we treat  $\text{sig}$  as fixed) and remove all terms from the joint posterior distribution that do not involve  $\mu$  to construct the full conditional distribution,

$$p(\mu|\text{sig}, y) \propto p(y|\mu) * p(\mu)$$

The full conditional of  $\mu$  again takes the form of the Normal distribution shown in Eq. ??; similarly,  $p(\text{sig}|\mu, y)$  takes the form of the Inverse Gamma

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

<sup>3</sup>maybe we should think out chapter 2 and concentrate that material here?

distribution shown in Eq. 11.3 both distribution we can easily sample from. And this is precisely what we do when using Gibbs sampling we break down high-dimensional problems into convenient one-dimensional problems by constructing the full conditional distributions for each model parameter separately; and we sample from these full conditionals, which, if we choose conjugate priors, are known parametric distributions. Let's put the concept of Gibbs sampling into the MCMC framework of generating successive samples, using our simple Normal model with unknown mu and sig and conjugate priors as an example. These are the steps you need to build a Gibbs sampler:

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

**Step 1:** Draw  $\theta_1(1)$  from the conditional distribution  $p(\theta_1(1) | \theta_2(0), \dots, \theta_d(0))$ . Here,  $\theta_1$  is mu, which we draw from the Normal distribution in Eq. 11.3 using  $\text{sig}(0)$  as value for sig.

**Step 2:** Draw  $\theta_2(1)$  from the conditional distribution  $p(\theta_2(1) | \theta_1(1), \theta_3(0), \dots, \theta_d(0))$ . Here,  $\theta_2$  is sig, which we draw from the Inverse Gamma distribution of Eq. 11.3, using  $\text{mu}(1)$  as value for mu...

**Step d:** Draw  $\theta_d(1)$  from the conditional distribution  $p(\theta_d(1) | \theta_1(1), \dots, \theta_{d-1}(1))$

In our example we have no additional parameters, so we only need step 0 through to 2. Repeat Steps 1 to d for  $K$  = a large number of samples. In terms of R coding, this means we have to write Gibbs updaters for mu and sig and embed them into a loop over  $K$  iterations. The final code in the form of an R function is shown in Panel 1.

Andy will build the panel environment here soon.

Panel 1: R-code for a Gibbs sampler for a Normal model with unknown mu and sig and conjugate (Normal and Inverse Gamma, respectively) priors for both parameters.

```
Normal.Gibbs<-function(y=y,mu0=mu0, sig0=sig0, a=a,b=b,niter=niter) {
  ybar<-mean(y)
  n<-length(y)
  mu<-runif(1) #mean initial value
  sig<-runif(1) #sd initial value
  an<-n/2 + a
  out<-matrix(nrow=niter, ncol=2)
```

```

5110 colnames(out)<-c('mu', 'sig')
5111
5112 for (i in 1:niter) {
5113
5114   #update mu
5115   mun<- (sig/(sig+n*sig0))*mu0 + (n*sig0/(sig+n* sig0))*ybar
5116   sign <- (sig*sig0)/ (sig+n*sig0)
5117   mu<-rnorm(1,mun, sqrt(sign))
5118
5119   #update sig
5120   bn<- 0.5 * (sum((y-mu)^2)) +b
5121   sig<-1/rgamma(1,shape=an, rate=bn)
5122   out[i,]<-c(mu,sqrt(sig))
5123
5124 }
5125 return(out)
5126 }

```

5127 This is it! You can use the code `NormalGibbs.R` in the **R** package `scrbook`  
 5128 to simulate some data,  $y \sim \text{Normal}(5, 0.5)$  and run your first Gibbs sampler.  
 5129 Your output will be a table with two columns, one per parameter, and  $K$  rows,  
 5130 one per iteration. For this 2-parameter example you can visualize the joint  
 5131 posterior by plotting samples of  $\mu$  against samples of  $\sigma$  (Fig. 2 XXX):

```

5132 plot(out[,1], out[,2])

```

5133 The marginal distribution of each parameter is approximated by just examining  
 5134 the samples of this particular parameter you can visualize it by plotting a  
 5135 histogram of the samples (Fig. 3 a, b XXX):

```

5136 par(mfrow=c(1,2))
5137 hist(out[,1]); hist (out[,2])

```

5138 Finally, recall an important characteristic of Markov chains, namely, that the  
 5139 chain has to have converged (reached its stationary distribution) for samples to  
 5140 come from the posterior distribution. In practice, that means you have to throw  
 5141 out some of the initial samples called the burn-in. We will talk about this in  
 5142 more when we talk about convergence diagnostics. For now, you can use the  
 5143 `plot(out[,1])` or `plot(out[,2])` command to make a time series plot of the  
 5144 samples of each parameter and visually assess how many of the initial samples  
 5145 you should discard. Figure 3 c and d shows plots for the estimates of  $\mu$  and  
 5146 sigma from our simulated data set; you see that in this simple example the  
 5147 Markov chain apparently reaches its stationary distribution very quickly the  
 5148 chains look 'grassy' seemingly from the start. It is hard to discern a burn-in  
 5149 phase visually (but we will see examples further on where the burn-in is clearer)  
 5150 and you may just discard the first 500 draws to be sure you only use samples  
 5151 from the posterior distribution. The mean of the remaining samples are your  
 5152 estimates of  $\mu$  and  $\sigma$ :

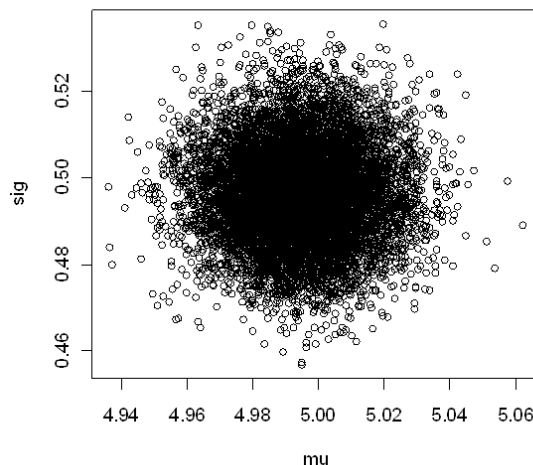


Figure 11.2: Joint posterior distribution of mu and sig from a Normal Model

```

5153 > summary(mod[501:10000,])
5154      mu              sig
5155 Min.   : 4.936      Min.   : 0.4569
5156 1st Qu.: 4.984      1st Qu.: 0.4889
5157 Median : 4.994      Median : 0.4961
5158 Mean   : 4.994      Mean   : 0.4964
5159 3rd Qu.: 5.005      3rd Qu.: 0.5037
5160 Max.   : 5.062      Max.   : 0.5356

```

### 5161 11.3.2 Metropolis-Hastings sampling

5162 Although it is applicable to a wide range of problems, the limitations of Gibbs  
 5163 sampling are immediately obvious what if we do not want to use conjugate priors  
 5164 (or what if we cannot recognize the full conditional distribution as a parametric  
 5165 distribution, or simply do not want to worry about these issues)? The most  
 5166 general solution is to use the Metropolis-Hastings (MH) algorithm, which also  
 5167 goes back to the work by Metropolis et al. (1953). You saw the basics of this  
 5168 algorithm in Chapter 2. In a nutshell, because we do not recognize the posterior  
 5169  $p(\theta|y)$  as a parametric distribution, the MH algorithm generates samples from a  
 5170 known proposal distribution, say  $h(\theta)$ , that depends on  $\theta$  at  $t-1$ . The  $t^{th}$  sample  
 5171 is accepted or rejected based on its joint posterior probability density compared  
 5172 to the density of the sample at  $t-1$ . The original Metropolis algorithm requires  
 5173  $h(\theta)$  to be symmetric so that  $h(\theta^t|\theta^{t-1}) = h(\theta^{t-1}|\theta^t)$ ; but a later development

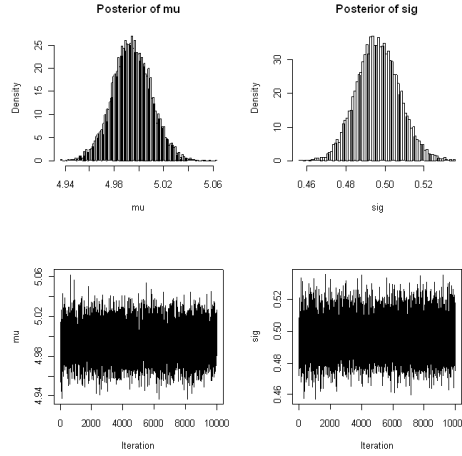


Figure 11.3: Plots of the posterior distributions of  $\mu$ (a) and  $\sigma$  (b) from a Normal model and time series plots of  $\mu$  (c) and  $\sigma$  (d).

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 sampling'. It is worth knowing that there are alternative formulations of the algorithm. For example, in the independent M-H,  $\theta^t$  does not depend on  $\theta^{t-1}$ , 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  $\theta$  at  $t$  from  $h(\theta_t - \theta_{t-1})$ . We often use a Normal proposal distribution, so we draw  $\theta_1$  from  $\text{Normal}(\theta_0, \text{sig}^2)$ , where  $\text{sig}^2$  is the variance of the Normal proposal distribution, a tuning parameter that we have to set.

5195 Step 2: Calculate the ratio of posterior densities for the proposed and the orig-  
 5196 inal value for  $\theta$ :

$$r = p(\theta^t|y)/p(\theta^{t-1}|y)$$

5197 In our example,

$$r = \text{Bern}(y|\theta^t) * \text{Normal}(\theta^t|\mu_0, \sigma_0) / \text{Bernoulli}(y|\theta^{t-1}) * \text{Normal}(\theta^{t-1}|\mu_0, \sigma_0)$$

5198 Step 3: Set

```
5199 \begin{eqnarray*}
5200 \theta^t &= & \theta^{t-1} \text{ with probability } \min(r,1) // \\
5201 &= & \theta^{t-1} \text{ otherwise } \\
5202 \end{eqnarray*}
```

5203 We can do that by drawing a random number  $u$  from a  $\text{Unif}(0,1)$  and accept  
 5204  $\theta^t$  if  $u < r$ . Repeat for  $t = 1, 2, \dots$  a large number of samples. The **R** code for  
 5205 this MH sampler is provided in Panel 2 XXXX.

5206 Panel 2: R code to run a Metropolis sampler on a simple Logit-Normal model.

```
5207
5208 Logreg.MH<-function(y=y, mu0=mu0, sig0=sig0, niter=niter) {
5209
5210   out<-c()
5211
5212   theta<-runif(1, -3,3) #initial value
5213
5214   for (iter in 1:niter){
5215     theta.cand<-rnorm(1, theta, 0.2)
5216
5217     loglike<-sum(dbinom(y, 1, exp(theta)/(1+exp(theta)), log=TRUE))
5218     logprior <- dnorm(theta,mu0 ,sig0, log=TRUE)
5219     loglike.cand<-sum(dbinom(y, 1, exp(theta.cand)/(1+exp(theta.cand)), log=TRUE))
5220     logprior.cand <- dnorm(theta.cand, mu0, sig0, log=TRUE)
5221
5222     if (runif(1)<exp((loglike.cand+logprior.cand)-(loglike+logprior))){
5223       theta<-theta.cand
5224     }
5225     out[iter]<-theta
5226   }
5227
5228   return(out)
5229 }
```

5230 The reason we sum the logs of the likelihood and the prior, rather than  
 5231 multiplying the original values, is simply computational. The product of small  
 5232 probabilities can be numbers very close to 0, which computers do not handle  
 5233 well. Thus we add the logarithms, sum, and exponentiate to achieve the desired  
 5234 result. Similarly, in case you have forgotten some elementary math,  $x/y =$   
 5235  $\exp(\log(x) - \log(y))$ , with the latter being favored for computational reasons.



Comparing MH sampling to Gibbs sampling, where all draws from the conditional distribution are used, in the MH algorithm we discard a portion of the candidate values, which inherently makes it less efficient than Gibbs sampling the price you pay for its increased generality. In Step 1 of the MH sampler we had to choose a variance for the Normal proposal distribution. Choice of the parameters that define our candidate distribution is also referred to as 'tuning', and it is important since adequate tuning will make your algorithm more efficient, i.e. your Markov chain will converge faster. The variance should be chosen so that (a) each step of drawing a new proposal value for  $\theta$  can cover a reasonable distance in the parameter space, as otherwise, the random walk moves too slowly; and (b) proposal values are not rejected too often, as otherwise the random walk will 'get stuck' at specific values for too long. As a rule of thumb, your candidate value should be accepted in about 40% of all cases. Acceptance rates of 20–80% are probably ok, but anything below or above may well render your algorithm inefficient (this does not mean that it will give you wrong results only that you will need more iterations to converge to the posterior distribution). In practice, tuning will require some 'trial-and-error' and some common sense. Or, one can use an adaptive phase, where the tuning parameter is automatically adjusted until it reaches a user-defined acceptance rate, at which point the adaptive phase ends and the actual Markov chain begins. This is computationally a little more advanced. Link and Barker (2009) discuss this in more detail. It is important the samples drawn during the adaptive phase are discarded. You can easily check acceptance rates for the parameters you monitor (that are part of your output) using the `rejectionRate()` function of the package `coda` (we will talk more about this package a little later on). Do not let the term 'rejection rate' confuse you; it is simply  $1 - \text{acceptance rate}$ . There may be parameters for example, individual values of a random effect or latent variables that you do not want to save, though, and in our next example we will show you a way to monitor their acceptance rates with a few extra lines of code.

### 11.3.3 Metropolis-within-Gibbs

One weakness of the MH sampler is that formulating the joint posterior when evaluating whether to accept or reject the candidate values for  $\theta$  becomes increasingly complex or inefficient as the number of parameters in a model increases. It is probably going to sound like MCMC sampling is too good to be true but in these cases you can simply combine MH sampling and Gibbs sampling. You can use Gibbs sampling to break down your high-dimensional parameter space into easy-to-handle one-dimensional conditional distributions and use MH sampling for these conditional distributions. Better yet if you have some conjugacy in your model, you can use the more efficient Gibbs sampling for these parameters and one-dimensional MH for all the others. You have already seen the basics of how to build both types of algorithms, so we can jump straight into an example here and build a Metropolis-within-Gibbs algorithm.

## 11.4 GLMMs Poisson regression with a random effect

Let's assume a model that gets us closer to the problem we ultimately want to deal with a GLMM. Here, we assume we have Poisson counts,  $y$ , from  $i$  plots in  $j$  different study sites, and we believe that the counts are influenced by some plot-specific covariate,  $x$ , but that there is also a random site effect. So our model is:

$$y_{ij} \sim \text{Poisson}(lami_{ij})$$

$$lami_{ij} = \exp(a_j + b * x_i)$$

Let's use Normal priors on  $a$  and  $b$ ,

$$a_j \sim \text{Normal}(\mu_a, \sigma_a)$$

and

$$b \sim \text{Normal}(\mu_b, \sigma_b)$$

<sup>4</sup> Since we want to estimate the random effect in this model, we do not specify  $\mu_a$  and  $\sigma_a$ , but instead, estimate them as well, so we have to specify hyperpriors for these parameters:

$$\mu_a \sim \text{Normal}(\mu_0, \sigma_0)$$

$$\sigma_a \sim \text{InvGamma}(a_0, b_0)$$

With the model fully specified, we can compile the full conditionals, breaking the multi-dimensional parameter space into one-dimensional components:

```
\begin{eqnarray*}
p(a_1|a_2,a_3,a_j,b,y) & \propto & p(y_{i1}|a_1,b) * p(a_1|\mu_a, \sigma_a) \backslash \backslash \\
& \propto & \text{Poisson}(y_{i1} | \exp(a_1 + b*x[j=1])) * \text{Normal}(a_1|\mu_a, \sigma_a) \\
\end{eqnarray*}
\begin{eqnarray*}
p(a_2|a_1,a_3,a_j,b,y) & \propto & p(y_{i2}|a_2,b) * p(a_2|\mu_a, \sigma_a) \backslash \backslash \\
& \propto & \text{Poisson}(y_{i2} | \exp(a_2 + b*x[j=1])) * \text{Normal}(a_2|\mu_a, \sigma_a) \\
\end{eqnarray*}
\text{and so on for all elements of } a.
\begin{eqnarray*}
p(b|a,y) & \propto & p(y|a,b) * p(b) \backslash \backslash \\
& \propto & \text{Poisson}(y | \exp(a + b*x)) * \text{Normal}(b|\mu_b, \sigma_b) \\
\end{eqnarray*}
```

Finally, we need to update the hyperparameters for  $a$ :

$$p(\mu_a|a) \propto p(a|\mu_a, \sigma_a) * p(\mu_a)$$

$$p(\sigma_a|a) \propto p(a|\mu_a, \sigma_a) * p(\sigma_a)$$

---

<sup>4</sup>Why is  $b$  a hyperparameter?

Since we assumed  $a$  to come from a Normal distribution, the choice of priors for  $\mu_a$  Normal and  $\sigma_a$  Inverse Gamma leads to the same conjugacy we observed in our initial Normal model, so that both hyperparameters can be updated using Gibbs sampling.

Now let's build the updating steps for these full conditionals. Again, for the MH steps that update  $a$  and  $b$  we use Normal proposal distributions with standard deviations  $\sigma_a$  and  $\sigma_b$ .

First, we set the initial values  $a(0)$  and  $b(0)$ . Then, starting with  $a(1)$ , we draw  $a(1)$  from Normal( $a(0)$ ,  $\sigma_a$ ), calculate the conditional posterior density of  $a(0)$  and  $a(1)$  and compare their ratios,

$$r = \text{Poisson}(y(j=1)|\exp(a(1)+b*x)) * \text{Normal}(a(1)|\mu_a, \sigma_a) / \text{Poisson}(y(j=1)|\exp(a(0)+b*x)) * \text{Normal}(a(0)|\mu_a, \sigma_a)$$

and accept  $a(1)$  with probability  $\min(r, 1)$ . We repeat this for all  $a$ 's.

For  $b$ , we draw  $b(1)$  from Normal( $b(0)$ ,  $\sigma_b$ ), compare the posterior densities of  $b(0)$  and  $b(1)$ ,

$$r = \text{Poisson}(y|\exp(a+b(1)*x)) * \text{Normal}(b(1)|\mu_b, \sigma_b) / \text{Poisson}(y|\exp(a+b(0)*x)) * \text{Normal}(b(0)|\mu_b, \sigma_b),$$

and accept  $b(1)$  with probability  $\min(r, 1)$ .

For  $\mu_a$  and  $\sigma_a$ , we sample directly from the full conditional distributions (Eq XX and Eq XX):

$$\mu_a(1) \sim \text{Normal}(\mu_n, \sigma_n)$$

where  $\mu_n = (\sigma_a(0)/\sigma_a(0) + n_a * \sigma_a(0)) * \mu_0 + (n_a * \sigma_a(0)/\sigma_a(0) + n_a * \sigma_a(0)) * \bar{a}$  and  $\sigma_n = \sigma_a(0) * \sigma_a(0) / (\sigma_a(0) + n_a * \sigma_a(0))$ . Here,  $\bar{a}$  is the current mean of the vector  $a$ , which we updated before, and  $n_a$  is the length of  $a$ . For  $\sigma_a$  we use  $\sigma_a(1) \sim \text{InvGamma}(a_n, b_n)$ , where  $a_n = n_a/2 + a_0$ , and  $b_n = 1/2 \sum (a(1) - \mu_a(1))^2 + b_0$ .

We repeat these steps over  $K$  iterations of the MCMC algorithm. In this example we may not want to save each value for  $a$ , but are only interested in their mean and standard deviation. Since these two parameters will change as soon as the value for one element in  $a$  changes, their acceptance rates will always be close to 1 and are not representative of how well your algorithm performs. To monitor the acceptance rates of parameters you do not want to save, you simply need to add a few lines of code into your updater to see how often the individual parameters are accepted. The full code for the MCMC algorithm of our Poisson GLMM in Panel 3 shows one way how to monitor acceptance of individual  $a$ 's.

Panel 3: R code for the Metropolis-within-Gibbs sampler for a Poisson regression with random intercepts.

```
Pois.reg<-function(y=y,site=site,mu0=mu0,sig0=sig0,a0=a0,b0=b0,
  mub=mub, sigb=sigb, niter=niter){
  lev<-length(unique(site))      #number of sites
```

```

5347 a<-runif(lev,-5,5) #initial values a
5348 b<-runif(1,0,5) #initial value b
5349 mua<-mean(a)
5350 siga<-sd(a)
5351
5352 out<-matrix(nrow=niter, ncol=3)
5353 colnames(out)<-c('mua','siga','b')
5354
5355 for (iter in 1:niter) {
5356
5357   #update a
5358   aUps<-0 #initiate counter for acceptance rate of a
5359   for (j in 1:lev) { #loop over sites
5360     a.cand<-rnorm(1, a[j], 0.1) #update intercepts a one at a time
5361     loglike<- sum(dpois (y[site==j], exp(a[j] + b*x[site==j]), log=TRUE))
5362     logprior<- dnorm(a[j], mua,siga, log=TRUE)
5363     loglike.cand<- sum(dpois (y[site==j], exp(a.cand + b *x[site==j]), log=TRUE))
5364     logprior.cand<- dnorm(a.cand, mua,siga, log=TRUE)
5365     if (runif(1)< exp((loglike.cand+logprior.cand) (loglike+logprior))) {
5366       a[j]<-a.cand
5367       aUps<-aUps+1
5368     }
5369   }
5370
5371   if(iter %% 100 == 0) { #this lets you check the acceptance rate of a at every 100th iteration
5372     cat(" Acceptance rates\n")
5373     cat(" a =", aUps/lev, "\n")
5374   }
5375
5376   #update b
5377   b.cand<-rnorm(1, b, 0.1)
5378   avec<-rep(a, times=c(rep(10,10)))
5379   loglike<- sum(dpois (y, exp(avec + b*x), log=TRUE))
5380   logprior<- dnorm(b, mub,sigb, log=TRUE)
5381   loglike.cand<- sum(dpois (y, exp(avec + b.cand *x), log=TRUE))
5382   logprior.cand<- dunif(b.cand, mub,sigb, log=TRUE)
5383   if (runif(1)< exp((loglike.cand+logprior.cand) (loglike+logprior) )) {
5384     b<-b.cand
5385   }
5386
5387   #update mua using Gibbs sampling
5388   abar<-mean(a)
5389   mun<- (siga/(siga+lev*sig0))*mu0 + (lev*sig0/(siga+lev* sig0))*abar
5390   sign <- (siga*sig0)/ (siga+lev*sig0)
5391   mua<-rnorm(1,mun, sqrt(sign))
5392
5393   #update siga using Gibbs sampling
5394   a0n<-lev/2 + a0
5395   b0n<- 0.5 * (sum((a-mua)^2)) +b0
5396   siga<-1/rgamma(1,shape=a0n, rate=b0n)

```

```

5397
5398 out[iter,]<-c(mua, sqrt(siga), b)
5399
5400 }
5401
5402 return(out)
5403 }

```

#### 5404 11.4.1 Rejection sampling and slice sampling

5405 While MH and Gibbs sampling are probably the most widely applied algorithms  
 5406 for posterior approximation, there are other options that work under certain  
 5407 circumstances and may be more efficient when applicable. WinBUGS applies  
 5408 these algorithms and we want you to be aware that there is more out there  
 5409 to approximate posterior distributions than Gibbs and MH. One alternative  
 5410 algorithm is rejection sampling. Rejection sampling is not an MCMC method,  
 5411 since each draw is independent of the others. The method can be used when the  
 5412 posterior  $p(\theta|y)$  is not a known parametric distribution but can be expressed  
 5413 in closed form. Then, we can use a so-called envelope function, say,  $g(\theta)$ , that  
 5414 we can easily sample from, with the restriction that  $p(\theta|y) < M * g(\theta)$ . We  
 5415 then sample a candidate value for  $\theta$  from  $g(\theta)$ , calculate  $r = p(\theta|y)/M * g(\theta)$   
 5416 and keep the sample with the probability  $r$ .  $M$  is a constant that has to be  
 5417 picked so that  $r \in [0,1]$ , for example by evaluating both  $p(\theta|y)$  and  $g(\theta)$  at  
 5418  $n$  points and looking at their ratios. Rejection sampling only works well if  
 5419  $g(\theta)$  is similar to  $p(\theta|y)$ , and packages like WinBUGS use adaptive rejection  
 5420 sampling (Gilks and Wild, 1992), where a complex algorithm is used to fit an  
 5421 adequate and efficient  $g(\theta)$  based on the first few draws. Though efficient in  
 5422 some situations, rejection sampling does not work well with high-dimensional  
 5423 problems, since it becomes increasingly hard to define a reasonable envelope  
 5424 function. For an example of rejection sampling in the context of SCR models, see  
 5425 Chapter 9. Another alternative is slice sampling (Neal, 2003). In slice sampling,  
 5426 we sample uniformly from the area under the plot of  $p(\theta|y)$ . Considering a single  
 5427 univariate  $\theta$ . Let's define an auxiliary variable,  $U \sim Uniform(0, p(\theta|y))$ .  
 5428 Then,  $\theta$  can be sampled from the vertical slice of  $p(\theta|y)$  at  $U$  (Figure 4):

```

5429 \theta|U \sim \text{Unif}(B),

```

5430 where  $B = \theta : U < p(\theta|y)$

5431 <sup>5</sup>

5432 Slice sampling can be applied in many situations; however, implementing  
 5433 an efficient slice sampling procedure can be complicated. We refer the inter-  
 5434 ested reader to chapter 7 of Robert and Casella (2010) for a simple example.  
 5435 Both rejection sampling and slice sampling can be applied on one-dimensional  
 5436 conditional distributions within a Gibbs sampling setup.

---

<sup>5</sup>there are supposed to be equations in the caption of figure 4 but it kept causing errors

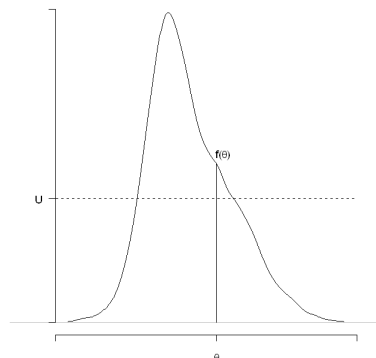


Figure 11.4: Slice sampling. For...

## 11.5 MCMC for closed capture-recapture Model

### Mh

6

## 11.6 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  $S$ ,

$$s_i \sim U(S)$$

---

<sup>6</sup>Andy could move material from chapter 3 to here.

5458 and that the number of times individual  $i$  encounters trap  $j$ ,  $y_{ij}$ , is a random  
5459 Poisson variable with mean  $\text{lam}_{ij}$ ,

$$y_{ij} \sim \text{Poisson}(\text{lam}_{ij})$$

5460 The tie between individual location, movement and trap encounter rates is made  
5461 by the assumption that  $\text{lam}_{ij}$ , is a decreasing function of the distance between  
5462  $s_i$  and  $j$ ,  $D_{ij}$ , of the half-normal form

$$\text{Lam}_{ij} = \text{lam}_0 * \exp(-D_{ij}^2/2 * \text{sig}^2),$$

5463 where  $\text{lam}_0$  is the baseline trap encounter rate at  $D_{ij} = 0$  and  $\text{sig}$  controls the  
5464 shape of the half-normal function.

5465 In order to estimate the number of  $s_i$  in  $S$ ,  $N$ , we use data augmentation  
5466 (sect. 3.XYZ) and create  $M$ - $n$  all-0 encounter histories, where  $n$  is the number  
5467 of individuals we observed and  $M$  is an arbitrary number that is larger than  $N$ .  
5468 We estimate  $N$  by summing over the auxiliary data augmentation variables,  $z_i$ ,  
5469 which is 1 if the individual is part of the population and 0 if not, and assume  
5470 that  $z_i$  is a random Bernoulli variable,

$$z_i \sim \text{Bern}(\psi)$$

5471 To link the two model components, we modify our trap encounter model to

$$\text{Lam}_{ij} = \text{lam}_0 * \exp(-D_{ij}^2/2 * \text{sig}^2) * z_i.$$

5472 The model has the following structural parameters, for which we need to spec-  
5473 ify priors  $\psi$  the Uniform (0,1) is required as part of the data augmentation  
5474 procedure and in general is a natural choice of an uninformative prior for a  
5475 probability; note that this is equivalent to a Beta(1,1) prior, which will come in  
5476 handy later.  $s_i$  since  $s_i$  is a pair of coordinates it is two-dimensional and we use  
5477 a uniform prior limited by the extent of our state-space over both dimensions.  
5478  $\sigma$  we can conceive several priors for sigma but let's assume an improper prior  
5479 one that is Uniform over  $(-\text{Inf}, \text{Inf})$ . We will see why this is convenient when we  
5480 construct the full conditionals for sigma.  $\lambda_0$  analogous, we will use a Uniform  
5481  $(-\text{Inf}, \text{Inf})$  improper prior for sigma. The parameter that is the objective of our  
5482 modeling,  $N$ , is a derived parameter that we can simply obtain by summing all  
5483  $z$ 's:

$$N = \text{sum}(z)$$

5484 **Step 2 - Construct the full conditionals** Having completed step 1,  
5485 let's look at the full conditional distributions for some of these parameters.  
5486 We find that with improper priors, full conditionals are proportional only to  
5487 the likelihood of the observations; for example, take the movement parameter  
5488 sigma:

$$\text{Sig}|s, \text{lam}_0, z, y \propto [y|s, \text{lam}_0, z, \text{sig}] * [\text{sig}]$$

5489 Since the improper prior implies that  $[\text{sig}] \propto 1$ , we can reduce this further  
5490 to

$$\text{Sig}|s, \text{lam}_0, z, y \propto [y|s, \text{lam}_0, z, \text{sig}]$$

5491 The R code to update sigma is shown in Panel 4.<sup>7</sup>

5492 Panel 4: R code to update sigma within an MCMC algorithm for  
5493 an SCR model when using an improper prior

```
5494
5495
5496 sig.cand <- rnorm(1, sigma, 0.1) #draw candidate value
5497 if(sig.cand>0){ #automatically reject sig.cand that are <0
5498   lam.cand <- lam0*exp(-(D*D)/(2*sig.cand*sig.cand))
5499   ll<- sum(dpois(y, lam*z, log=TRUE))
5500   llcand <- sum(dpois(y, lam.cand*z, log=TRUE))
5501   if(runif(1) < exp( llcand - ll) ){
5502     ll<-llcand
5503     lam<-lam.cand
5504     sigma<-sig.cand
5505   }
5506 }
5507
```

5508 These steps are analogous for lam0 and si and we will use MH steps for all of  
5509 these parameters. Similar to the random intercepts in our Poisson GLMM, we  
5510 update each si individually. Note that to be fully correct, the full conditional  
5511 for si contains both the likelihood and prior component, since we did not specify  
5512 an improper, but a Uniform prior on si. However, with a Uniform distribution  
5513 the probability density of any value is 1/(upper limit - lower limit) = constant.  
5514 Thus, the prior components are identical for both the current and the candidate  
5515 value and can be ignored (formally, when you calculate the ratio of posterior  
5516 densities, r, the identical prior component appears both in the numerator and  
5517 denominator, so that they cancel each other out).

5518 We still have to update zi. The full conditional for zi is

$$z_i|y, \sigma, \lambda_0, \text{sprpto}[y|z, \sigma, \lambda_0, s] * [z_i]$$

5519 and since  $z_i \sim \text{Bernoulli}(\psi_i)$ , the term has to be taken into account when  
5520 updating zi. The R code for updating zi is shown in Panel 5.

5521 Panel 5: R code to update z

```
5522
5523 zUps <- 0 #set counter to monitor acceptance rate
5524 for(i in 1:M) {
5525   if(seen[i]) #no need to update seen individuals, since their z =1
5526     next
5527   zcand <- ifelse(z[i]==0, 1, 0)
5528   llz <- sum(dpois(y[i,], lam[i,]*z[i], log=TRUE))
5529   llcand <- sum(dpois(y[i,], lam[i,]*zcand, log=TRUE))

```

---

<sup>7</sup> 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.



```

5530
5531     prior <- dbinom(z[i], 1, psi, log=TRUE)
5532     prior.cand <- dbinom(zcand, 1, psi, log=TRUE)
5533     if(runif(1) < exp( (llcand+prior.cand) - (llz+prior) )) {
5534         z[i] <- zcand
5535         zUps <- zUps+1
5536     }
5537 }

```

5538  $\psi$  itself is a hyperparameter of the model, with an uninformative prior dis-  
5539 tribution of Unif(0,1) or Beta(1,1), so that

$$Psi|z \propto [z|psi] * Beta(1, 1)$$

5540 The Beta distribution is the conjugate prior to the Binomial and Bernoulli  
5541 distributions (remember that  $z \sim Bernoulli(psi)$ ). The general form of a full  
5542 conditional of a Beta-Binomial model with  $y_i \sim Bernoulli(p)$  and  $p \sim Beta(a, b)$   
5543 is

$$p(p|y) \propto Beta(a + \text{sum}(yi), b + n - \text{sum}(yi))$$

5544 In our case, this means we update psi as follows:

```

5545 si<-rbeta(1, 1+sum(z), 1 + M-sum(z))

```

5546 These are all the building blocks you need to write the MCMC algorithm  
5547 for the spatial null model with a Poisson encounter process. You can find the  
5548 full R code (SCR0pois.R) in the online supplementary material.

### 5549 11.6.1 SCR model with binomial encounter process

5550 The equivalent SCR model with a binomial encounter process is very similar.  
5551 Here, each individual  $i$  can only be detected once at any given trap  $j$  during a  
5552 sampling occasion  $k$ . Thus

$$y_{ij} \sim Binomial(p_{ij}, K)$$

5553 Where  $p_{ij}$  is some function of distance between  $\mathbf{s}_i$  and trap location  $\mathbf{x}_j$ . Here  
5554 we use:

$$p_{ij} = 1 - \exp(-\text{lam}_{ij})$$

5555 Recall from Chapter 2 that this is the complementary log-log (cloglog) link func-  
5556 tion, which constrains  $p_{ij}$  to fall between 0 and 1. For our MCMC algorithm that  
5557 means that, instead of using a Poisson likelihood,  $Poisson(y|sigma, \text{lam}_0, s, z)$ ,  
5558 we use a Binomial likelihood,  $Binomial(y, K|sigma, \text{lam}_0, s, z)$ , in all the con-  
5559 ditional distributions. As an example, Panel 6 shows the updating step for  $\text{lam}_0$   
5560 under a binomial encounter model. The full MCMC code for the binomial SCR  
5561 can be found in the online supplements.

```

5562 Panel 6: MCMC updater for lam0 in a SCR model with Binomial encounter
5563 process and cloglog link function on detection. Here, pmat =
5564 1-exp(-lam).
5565
5566     lam0.cand <- rnorm(1, lam0, 0.1)
5567     if(lam0.cand >0){ #automatically reject lam0.cand that are <0
5568         lam.cand <- lam0.cand*exp(-(D*D)/(2*sigma*sigma))
5569         p.cand <- 1-exp(-lam.cand)
5570         ll<- sum(dbinom(y, K, pmat *z, log=TRUE))
5571         llcand <- sum(dbinom(y, K, p.cand *z, log=TRUE))
5572         if(runif(1) < exp( llcand - ll) ){
5573             ll<-llcand
5574             pmat<-p.cand
5575             lam0<- lam0.cand
5576         }
5577     }

```

Another possibility is to model variation in the individual and site specific detection probability,  $p_{ij}$ , directly, without any transformation, such that

```

5580 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  $\text{lam0}$  (equivalent to  $p_0$  in Eq XX) becomes:

```

5586     lam0.cand <- rnorm(1, lam0, 0.1)
5587     if(lam0.cand >0 & lam0.cand < 1 ){ #automatically reject lam0.cand that are
5588         lam.cand <- lam0.cand*exp(-(D*D)/(2*sigma*sigma))
5589         ll<- sum(dbinom(y, K, lam *z, log=TRUE)) #no transformation needed
5590         llcand <- sum(dbinom(y, K, lam.cand *z, log=TRUE))
5591         if(runif(1) < exp( llcand - ll) ){
5592             ll<-llcand
5593             lam<-lam.cand
5594             lam0<- lam0.cand
5595         }
5596     }

```

## 11.6.2 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 ... <sup>8</sup> You can use the same script provided back in Chapter XX to read in the data and build the augmented encounter history

<sup>8</sup>Does this data set come up before Ch6? If not, introduce data here. Or, Andy, would you rather use simulated data?

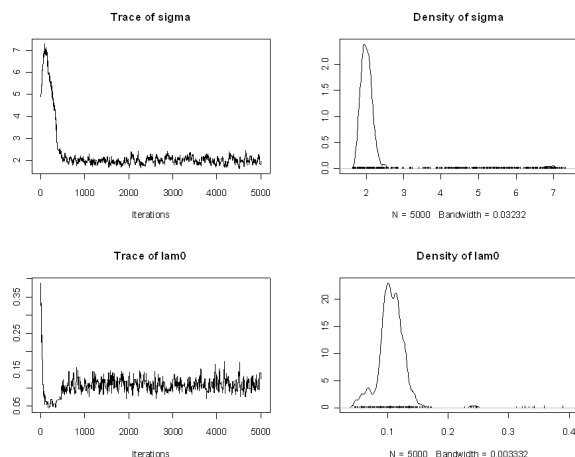


Figure 11.5: Time series and posterior density plots for  $\sigma$  and  $\text{lam0}$ .

5602 array; then source the MCMC code for the binomial encounter model algorithm  
 5603 with the cloglog link and run 5000 iterations. This should take approximately  
 5604 25 minutes.

```
5605 > source('SCR0binom.txt')
5606 > mod0<-SCR.0(y=bigTrap, X=trapmat, M=M, xl=xl, xu=xu, yl=yl, yu=yu, K=8, niter=5000)
```

5607 Before, we used simple R commands to look at model results. However, there  
 5608 is a specific R package to summarize MCMC simulation output and perform  
 5609 some convergence diagnostics package coda (Plummer et al., 2006). Download  
 5610 and install coda, then convert your model output to an mcmc object

```
5611 > chain<-mcmc(mod0)
```

## 5612 Markov chain time series plots

5613 Start by looking at time series plots of your Markov chains using `plot(chain)`.  
 5614 This command produces a time series plot and marginal posterior density plots  
 5615 for each monitored parameter, similar to what we did before using the `hist()`  
 5616 and `plot()` commands (Fig. 5). Time series plots will tell you several things:  
 5617 First, the way the chains move through the parameter space gives you an idea  
 5618 of whether your MH steps are well tuned. If chains were constant over many  
 5619 iterations you would probably need to decrease the tuning parameter of the  
 5620 (Normal) proposal distribution. If a chain moves along some gradient to a  
 5621 stationary state very slowly, you may want to increase the tuning parameter so  
 5622 that the parameter space is explored more efficiently.

5623 Second, you will be able to see if your chains converged and how many initial  
 5624 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 areas of the parameter space simultaneously; if after a while all chains oscillate around the same average value, chances are good that they indeed converged to the posterior distribution. Gelman and Rubin came up with a diagnostic statistic that essentially compares within-chain and between-chain variance to check for convergence of multiple chains (Gelman et al., 2004). Of course, running several parallel chains is computationally expensive. Extra computational demands are not the only and by no means the major concern some people voice when it comes to running several parallel MCMC chains to assess convergence. Again, consider the fact that we do not know anything about the true form of the posterior distribution we are trying to approximate. How do we, then, know how to pick overdispersed initial values? We dont all we can do is pick overdispersed values relative to our expectations of what the posterior should look like. To use a quote from the home page of Charlie Geyer, a Bayesian statistician from the University of Minnesota, “If you don’t know any good starting points [...], then restarting the sampler at many bad starting points is [...] part of the problem, not part of the solution.” (<http://users.stat.umn.edu/~charlie/mcmc/diag.html>). His suggestion is that your only chance to discover a potential problem with your MCMC sampler is to run it for a very long time. But again, there is no way of knowing how long is long enough. It is up to you to decide, which school of thoughts appeals more to you one long versus several parallel Markov chains. Irrespectively, part of developing an MCMC sampler should be to make sure (within reasonable limits) that you are not missing regions of high posterior density because of the way you specify your starting values. 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

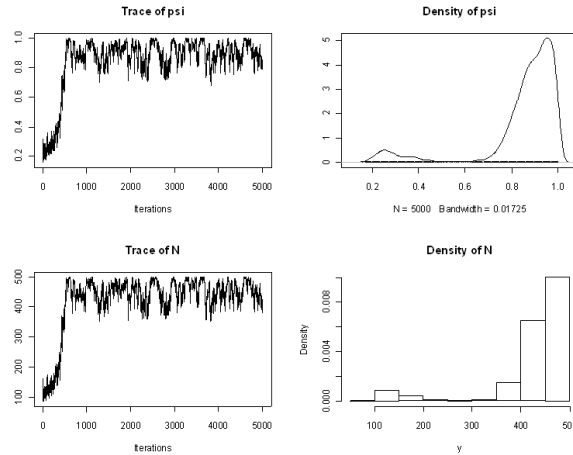


Figure 11.6: Time series and posterior density plots of  $\psi$  and  $N$  for the bear data set truncated by the upper limit of  $M$  (500).

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.

### 11.6.3 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  $\psi$  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.

#### 11.6.4 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 can use to do so. `effectiveSize()` will directly give you an estimate of the effective sample size for you parameters:

```
> effectiveSize(chain)
      sigma      lam0      psi      N
3.930303 78.259159 30.436348 32.047392
```

Alternatively, you can use the `autocorr.diag()` function, which will show you the degree of autocorrelation for different lag values (which you can specify within the function call, we use the defaults below):

```
> autocorr.diag(mcmc(mod))
      sigma      lam0      psi      N
Lag 0  1.0000000 1.0000000 1.0000000 1.0000000
Lag 1  0.9979948 0.9494134 0.9847503 0.9774201
Lag 5  0.9915567 0.8038168 0.9111951 0.9113525
Lag 10 0.9836016 0.6714021 0.8462108 0.8509803
Lag 50 0.8985337 0.1983780 0.6138516 0.6233994
```

Whichever function you use, if you find that your supposedly long Markov chain has not generated enough pseudo-iid samples, you should consider a longer run. In the present case we see that autocorrelation is especially high for the parameter `sigma` and our effective sample size for this parameter is 4! <sup>9</sup> This means we would 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

---

<sup>9</sup>Anyone have any idea how the autocorrelation in `sigma` could be reduced?

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.

### 11.6.5 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
Number of chains = 1
Sample size per chain = 4000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

      Mean      SD Naive SE Time-series SE
sigma  1.9986  0.13805 0.0021827      0.016091

```

```

5775 lam0      0.1096  0.01523 0.0002407      0.001401
5776 psi       0.6113  0.09148 0.0014465      0.010734
5777 N        489.8535 71.79695 1.1352094      8.431119
5778
5779 2. Quantiles for each variable:
5780
5781           2.5%      25%      50%      75%      97.5%
5782 sigma    1.75780    1.89847    1.9900    2.0944    2.2772
5783 lam0     0.08357    0.09824    0.1087    0.1192    0.1427
5784 psi      0.45110    0.54838    0.6052    0.6639    0.8192
5785 N        366.00000 440.00000 485.0000 530.0000 654.0000

```

Looking at the MC errors, we see that in spite of the high autocorrelation, the MC error for sigma is below the 1Our algorithm gives us a posterior distribution of N, but we are usually interested in the density, D. Density itself is not a parameter of our model, but we can derive a posterior distribution for D by dividing each value of N (N at each iteration) by the area of the state-space (here 3032.719 km<sup>2</sup>) and we can use summary statistics of this distribution to characterize D:

```

5793 > summary(window(chain[,4]/ 3032.719, start=1001))
5794 Iterations = 1001:5000
5795 Thinning interval = 1
5796 Number of chains = 1
5797 Sample size per chain = 4000
5798
5799 1. Empirical mean and standard deviation for each variable,
5800    plus standard error of the mean:
5801
5802           Mean           SD      Naive SE Time-series SE
5803    0.1615229    0.0236741    0.0003743    0.0027801
5804
5805 2. Quantiles for each variable:
5806
5807    2.5%    25%    50%    75%    97.5%
5808 0.1207 0.1451 0.1599 0.1748 0.2156

```

If we compare our mean density of 0.16/km<sup>2</sup> (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).

### 11.6.6 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:



```

5816 > rejectionRate(chain)
5817      sigma      lam0      psi      N
5818 0.44108822 0.77675535 0.00000000 0.01940388

```

5819 Recall that rejection rates should lie between 0.2 and 0.8, so our tuning seems  
 5820 to have been appropriate here. Psi is never rejected since we update it with  
 5821 Gibbs sampling, where all candidate values are kept. And since N is the sum  
 5822 of all z, all it takes for N to change from one iteration to the next are small  
 5823 changes in the z-vector, so the rejection rate of N is always low. If you have run  
 5824 several parallel chains, you can combine them into a single mcmc object using  
 5825 the mcmc.list() command on the individual chains (note that each chain has to  
 5826 be converted to an mcmc object before combining them with mcmc.list()). You  
 5827 can then easily obtain the Gelman-Rubin diagnostic (Gelman et al., 2004), in  
 5828 WinBUGS called R-hat, using gelman.diag(), which will indicate if all chains  
 5829 have converged to the same stationary distribution. For details on these and  
 5830 other functions, see the coda manual, which can be found together with the  
 5831 package on the CRAN mirror.

## 5832 11.7 Manipulating the state-space

5833 So far, we have constrained the location of the activity centers to fall within the  
 5834 outermost coordinates of our rectangular state space by posing upper and lower  
 5835 bounds for x and y. But what if S has an irregular shape maybe there is a large  
 5836 water body we would like to remove from S, because we know our terrestrial  
 5837 study species does not occur there. Or the study takes place in a clearly defined  
 5838 area such as an island. As mentioned before, this situation is difficult to handle  
 5839 in WinBUGS. In some simple cases we can adjust the state space by setting  
 5840 SXi to be some function of SYi or vice versa. In this manner, we can cut off  
 5841 corners of the rectangle to approximate the actual state space. In R, we are  
 5842 much more flexible, as we can use the actual state-space polygon to constrain  
 5843 out si.<sup>10</sup> To illustrate that, let's look at a camera trapping study of Florida  
 5844 panthers (*Puma concolor coryi*) conducted in the Picayune Strand Restoration  
 5845 Project (PSRP) area, southwest Florida (Fig. 7), by XXX, and financed by  
 5846 XXX. In the 1960ies the PSRP area was slated for housing development, but  
 5847 then bought back by the State of Florida and is currently being restored to its  
 5848 original hydrology and vegetation. In an effort to estimate the density of the  
 5849 local Florida panther population, 98 camera traps were operated in the area for  
 5850 21 months between 2005 and 2007. Florida panthers are wide-ranging animals  
 5851 and in order to account for their wide movements, the state-space was defined as  
 5852 the trapping grid buffered by 15 km around its outermost coordinates. However,  
 5853 the resulting rectangle contained some ocean in its southwestern corner (Fig. 7).  
 5854 In order to precisely describe the state-space, the ocean has to be removed. You  
 5855 can create a precise state-space polygon in ArcGIS and read it into R, or create  
 5856 the polygon directly within R. In the present case we intersected two shape files

---

<sup>10</sup> Have to check if we can use panther stuff for the book; otherwise, use raccoon example.

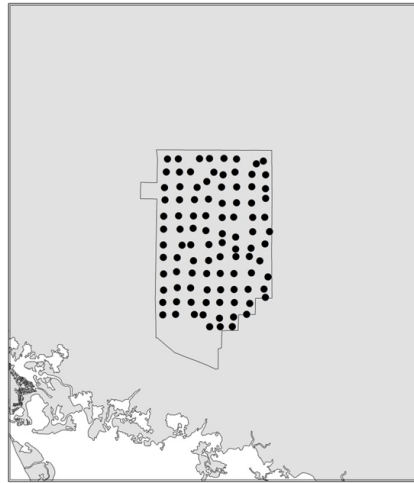


Figure 11.7: Rectangular state-space for a Florida panther camera trapping study in the PSRP area (grey outline, red block inset map of Florida) contain some ocean (white) that needs to be removed from the state-space.

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:

```

5865 xl= xmin-b
5866 xu= xmax+b
5867 yl= ymin-b
5868 yu= ymax+b
5869
5870 dd <- data.frame(Id=c(1,1,1,1,1),X=c(xl,xu,xu,xl,xl),Y=c(yl,yl,yu,yu,yl)) #create data
5871 ddTable <- data.frame(Id=c(1),Name=c("Item1"))
5872 ddShapefile <- convert.to.shapefile(dd, ddTable, "Id", 5) #convert #to shapefile, type
5873 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

5880 directly from the state-space object SSp:

```
5881 > area <- SSp@polygons[[1]]@Polygons[[1]]@area /1000000
```

5882 Note that dividing by 1000000 will return the area in km<sup>2</sup> if your coordi-  
 5883 nates describing the polygon are in UTM. If your state-space consists of several  
 5884 disjunct polygons, you will have to sum the areas of all polygons to obtain the  
 5885 size of the state-space. To include this polygon into our MCMC sampler we  
 5886 need one last spatial R package `sp` (Pebesma and Bivand, 2011), which has a  
 5887 function, `over()`, which allows us to check if a pair of coordinates falls within a  
 5888 polygon or not. All we have to do is embed this new check into the updating  
 5889 steps for `s`:

```
5890         Scand <- as.matrix(cbind(rnorm(M, S[,1], 2),
5891                                 rnorm(M, S[,2], 2)))          #draw candidate value
5892
5893 Scoord<-SpatialPoints(Scand*1000)      #convert to spatial points on UTM (m) scale
5894 SinPoly<-over(Scoord,SSp) # check if scand is within the polygon
5895
5896         for(i in 1:M) {
5897 if(is.na(SinPoly[i])==FALSE) { #if scand falls within polygon, continue update
5898   [rest of the updating step remains the same]
```

5899 Note that it is much more time-efficient to draw all  $M$  candidate values for  $s$   
 5900 and check once if they fall within the state-space, rather than running the `over()`  
 5901 command for every individual pair of coordinates. To make sure that our initial  
 5902 values for  $s$  also fall within the polygon of  $S$ , we use the function `runifpoint()`  
 5903 from the package `spatstat` (Baddeley and Turner, 2005), which generates random  
 5904 uniform points within a specified polygon. You'll find this modified MCMC al-  
 5905 gorithm in the online supplementary material (SCR0poisSSp). Finally, observe  
 5906 that we are converting candidate coordinates of  $S$  back to meters to match the  
 5907 UTM polygon. In all previous examples, for both the trap locations and the  
 5908 activity centers we have used UTM coordinates divided by 1000 to estimate  
 5909 sigma on a km scale. This is adequate for wide ranging individuals like bears.  
 5910 In other cases you may center all coordinates on 0. No matter what kind of  
 5911 transformation you use on your coordinates, make sure to always convert can-  
 5912 didate values for  $S$  back to the original scale (UTM) before running the `over()`  
 5913 command.

## 5914 11.8 MCMC software packages

5915 Throughout most of this book we will use WinBUGS and, occasionally, JAGS  
 5916 to run MCMC analyses. Here, we will briefly discuss the main pros and cons of  
 5917 these two programs as well as WinBUGS successor OpenBUGS. You can find  
 5918 scripts to simulate data and run the basic SCR model in all three programs in  
 5919 the online supplementary material (simSCR0poisBUGS).

### 5920 11.8.1 WinBUGS

5921 In a nutshell, WinBUGS (and the other programs) do everything that we just  
5922 went through in this chapter (and quite a bit more). Looking through your  
5923 model, WinBUGS determines which parameters it can use standard Gibbs sam-  
5924 pling for (i.e. for conjugate full conditional distributions). Then, it determines,  
5925 in the following hierarchy, whether to use adaptive rejection sampling, slice  
5926 sampling or in the 'worst' case Metropolis-Hastings sampling for the other full  
5927 conditionals (Spiegelhalter et al., 2003). If it uses MH sampling, it will auto-  
5928 matically tune the updater so that it works efficiently. While WinBUGS is a  
5929 convenient piece of software that is still widely used, its major drawback is that  
5930 it is no longer being developed, i.e. no new functions or distributions are added  
5931 and no bugs are fixed.

### 5932 11.8.2 OpenBUGS

5933 OpenBUGS is essentially the successor of WinBUGS. While the latter is no  
5934 longer worked on, OpenBUGS is constantly developed further. The name  
5935 'OpenBUGS' refers to the software being open source, so users do not need  
5936 to download a license key, like they have to for WinBUGS (although the license  
5937 key for WinBUGS is free and valid for life).

5938 Compared to WinBUGS, OpenBUGS has a lot more built-in functions. The  
5939 method of how to determine the right updater for each model parameter has  
5940 changed and the user can manually control the MCMC algorithm used to update  
5941 model parameters. Several other changes have been implemented in OpenBUGS  
5942 and a detailed list of differences between the two BUGS versions, can be found  
5943 at <http://www.openbugs.info/w/OpenVsWin>

5944 While OpenBUGS is a useful program for a lot of MCMC sampling appli-  
5945 cations, for reasons we do not understand, simple SCR models do not converge  
5946 in OpenBUGS. It is therefore advisable that you check any OpenBUGS SCR  
5947 model results against result from WinBUGS. Also, currently, the R package  
5948 BRugs (Thomas et al., 2006) necessary for running OpenBUGS through R  
5949 has problems with 64-bit machines, so you may have to use the 32-bit version  
5950 of R and OpenBUGS in order to make it work. The BUGS project site at  
5951 <http://www.openbugs.info> provides a lot of information on and download links  
5952 for OpenBUGS.

5953 There is an extensive help archive for both WinBUGS and OpenBUGS and  
5954 you can subscribe to a mailing list, where people pose and answer questions of  
5955 how to use these programs at <http://www.mrc-bsu.cam.ac.uk/bugs/overview/list.shtml>

### 5956 11.8.3 JAGS Just Another Gibbs Sampler

5957 JAGS, currently at Version 3.1.0, is another free program for analysis of Bayesian  
5958 hierarchical models using MCMC simulation. Originally, JAGS was the only  
5959 program using the BUGS language that would run on operating systems other  
5960 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>.<sup>11</sup>

## 11.9 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 these 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 ad-

<sup>11</sup>As we make progress on the book, let's be sure to add linkages to places where we use JAGS in examples.

6002 equate description of the state-space? In this chapter we saw that in R it only  
 6003 takes a few lines of code to use any arbitrary polygon shapefile as the state-  
 6004 space, which is especially useful when you are dealing with coastlines or large  
 6005 bodies of water that need removing from the state-space. Another example is  
 6006 the SCR R package SPACECAP (Gopalaswamy et al., 2011) that was developed  
 6007 because implementation of an SCR model with a discrete state-space was ineff-  
 6008 icient in WinBUGS. Another situations in which using BUGS/JAGS becomes  
 6009 increasingly complicated or inefficient is when using point processes other than  
 6010 the Uniform Poisson point process which underlies the basic SCR model (see  
 6011 Chapter X). In the Chapters 9 and XX you will see examples of different point  
 6012 processes, implemented using custom-made MCMC algorithms.<sup>12</sup> Finally,  
 6013 the Chapters XX and XX deal with unmarked or partially marked populations  
 6014 using hand-made MCMC algorithms to handle the (partially) latent individual  
 6015 encounter histories. While some of these models can be written in BUGS/JAGS,  
 6016 <sup>13</sup>, they are painstakingly slow; others cannot be implemented in BUGS/JAGS  
 6017 at all. In conclusion, while you can certainly get by using BUGS/JAGS for  
 6018 standard SCR models, knowing how to write your own MCMC sampler allows  
 6019 you to tailor these models to your specific needs.

---

<sup>12</sup>Richard, Beth expand on that?

<sup>13</sup>the 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 dont remember. We may have to try writing the others before saying that they dont work in BUGS/JAGS; they are certainly much faster in R, though.

## 6020 Chapter 12

## 6021 Goodness of Fit and stuff





## 6022 Chapter 13

## 6023 Covariate models



## Chapter 14

# 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”<sup>1</sup> 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 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.

---

<sup>1</sup>The phrase “complete spatial-randomness” is reserved for the homogeneous Poisson point process

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.

## 14.1 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,  $\mu$  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.

```
mu <- 4                                # intensity
Np <- rpois(1, mu)                     # Np is random
PPP <- cbind(runif(Np), runif(Np))     # Poisson point process

Nb <- 4                                # Nb is fixed
BPP <- cbind(runif(Nb), runif(Nb))     # Binomial point process
```

Note that in both models, the  $N$  points are independent of one another and distributed uniformly throughout  $\mathcal{S}$ . Thus, the intensity at any point  $x \in \mathcal{S}$  is  $\mu = 1/A(\mathcal{S})$  where  $A(\mathcal{S})$  denotes the area of the state-space. In the **R** code above, the area of the state-space is 1 unit, and thus the intensity is  $\mu = 1/1$ .

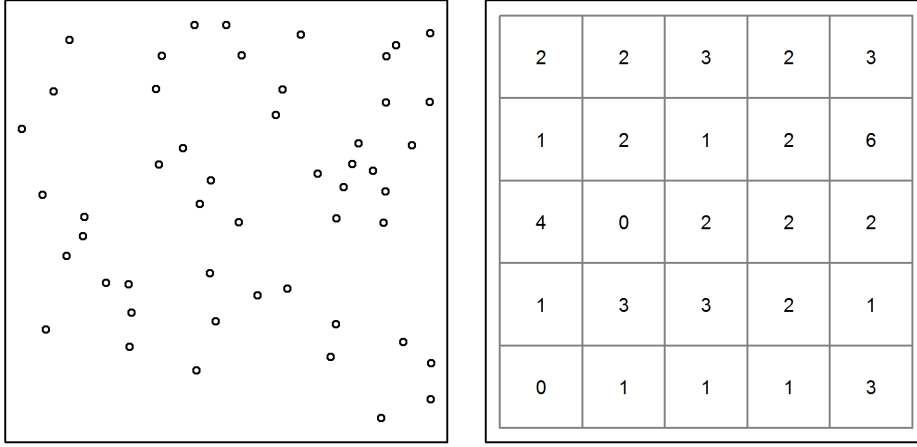


Figure 14.1: Homogeneous binomial point process with  $N=50$  points represented in continuous and discrete space.

Although the Poisson model is typically described in terms of  $\mu$ , the binomial model is not; rather, it is more common to consider a discrete state space, such as a grid with  $K$  pixels. Under the binomial model, the number of points in each region is  $n(B_k) \sim \text{Bin}(N, p_k)$  where  $p_k = A(B)/A(S)$ , ie  $p_k$  is simply the fraction of the state-space area in  $B_k$ . This discrete space representation of the binomial point process is shown in Fig. 14.1. The state-space in this case is the unit square, and thus the probability of a point falling in each of the 25 disjunct regions is  $p_k = 1/25$  and thus the expected counts are simply  $\mathbb{E}(n(B_k)) = Np_k$ . In the figure  $N = 50$  and thus we would expect 2 points per pixel, which happens to be the empirical mean of the data in Fig. 14.1. Note also that these counts are not independent realizations from a binomial distribution since  $\sum_k n(B_k) = N$ . Instead, the model for the entire vector is  $\mathbf{n}(\mathbf{B}) \sim \text{Multinomial}(N, \pi = (p_1, p_2, \dots, p_K))$  (Illian, 2008b). The dependence among counts has virtually no practical consequence when the number of pixels is large. For example, if we have 100 pixels, the number of counts in one pixels tells you very little about the expected count in another pixel. However, if there are only 2 pixels, then clearly the number of points in one pixel tells you exactly how many will occur in the remaining pixel. To gain familiarity with the multinomial distribution and the discrete representation of space, use the `rmultinom` function in **R** to simulate counts similar to those shown in Fig. 14.1, for example using a command such as:

```
n.B_k <- rmultinom(1, size=50, prob=rep(1/25, 25))
matrix(n.B_k, 5, 5)
```

The discrete space representation of the binomial point process is of practical importance when fitting SCR models because spatial covariates are almost always represented in a discrete format, often called “rasters” in GIS-speak. In such cases, we often need to change our definition of the prior for an activity center from  $s_i \sim \text{Uniform}(\mathcal{S})$  to  $s_i \sim \text{Multinomial}(1, \pi)$ . In the latter case, the activity center is simply defined as an integer representing pixel “id”. Note also that the multinomial distribution with an index of 1 (*i.e.* `size=1` in `rmultinom`) is referred to as the categorical distribution, which we will frequently use in the BUGS language.

## 14.2 Inhomogeneous binomial point process

As with the homogeneous model, the inhomogeneous binomial point process model is developed conditional on  $N$ . The primary distinction is that the uniform distribution is replaced with another distribution allowing for the intensity parameter to vary spatially. To arrive at this new distribution, define  $\mu(x, \beta)$  to be a function of spatially-referenced covariates ( $\beta$ ) available at all points of the state space. To be concise we will subsequently drop the vector of coefficients from our notation, and simply use  $\mu(x)$ . Since an intensity must be strictly positive, it is natural to model  $\mu(x)$  using the log-link.

$$\log(\mu(x)) = \sum_{j=1}^J \beta_j v_j(x), \quad x \in \mathcal{S}$$

where  $\beta_j$  is the regression coefficient for covariate  $v_j(x)$ . To be clear,  $v(x)$  is the value of any covariate, such as habitat type or elevation, at location  $x$ . This equation should look familiar because it is the standard linear model used in log-linear GLMs. Note, however, that we have no need for an intercept because it would be confounded with  $N$ . This should be intuitive since an intercept would represent the expected value of  $N$  when  $\beta = 0$ , but we already have a parameter in the model for expected abundance, namely  $\mathbb{E}[N] = \psi M$ . Thus an intercept would be redundant, and without it we are still able to achieve our goal of describing the distribution of  $N$  activity centers as a function of spatial covariates.

Now that we have a model of the intensity parameter  $\mu(x)$ , we need to develop the associated probability density function to use in place of the uniform prior. Remembering that the integral of a pdf must be unity, we can create a pdf by dividing  $\mu(x)$  by a normalizing constant, which in this case is the integral of  $\mu(x)$  evaluated over the entire state-space. **ANDY, is there a better justification for this?** The probability density function is therefore

$$f(x) = \frac{\mu(x)}{\int_{x \in \mathcal{S}} \mu(x) dx} \quad (14.1)$$

Substituting this distribution for the uniform prior allows us to fit inhomogeneous binomial point process models to spatial capture-recapture data. We can

also use this distribution to obtain the expected number of individuals in any given region. Specifically, the proportion of  $N$  expected to occur in any region  $B$  when heterogeneity in density is present is  $p(B) = \int_B f(x) dx$ . These are also the multinomial cell probabilities if the regions are disjoint and compose the entire state-space.

As a practical matter, note that the integral in the denominator of  $f(x)$  is evaluated over space, and since we almost always regard space as two-dimensional, this is a two-dimensional integral that can be approximated using the methods discussed in refChXXX. These methods include Monte Carlo integration, Gaussian quadrature, etc... Alternatively, if our state-space covariates are in raster format, *i.e* they are in discrete space, the integral can be replaced with a sum over all pixels, which is much more efficient computationally.

We now have all the tools needed to fit inhomogeneous point process (IPP) models. Before doing so, we note that the IPP for the activity centers results in another IPP for the observation process,  $\lambda(x)$ . As a reminder,  $\lambda(x)$  is the expected number of captures for a trap at point  $x$ . As was true for the homogeneous model, this intensity function is a product of the point process intensity and the encounter rate function,  $\lambda(x) = \mu(x)g(x)$ .

In the next section we walk through a few examples, building up from the simplest case where we actually observe the activity centers as though they were data. In the second example, we fit our new model to simulated data in which density is a function of a single continuous covariate. Example three shows an analysis in discrete space using both **secr** (Efford, 2011) and **JAGS** (Plummer, 2003). In the last example, we model the intensity of activity centers for a real dataset collected on jaguars (*Panthera onca*) in Argentina.

## 14.3 Examples

### 14.3.1 Simulation and analysis of inhomogeneous point processes

In SCR models, the point process is not directly observed, but in other contexts it is. Examples include the locations of disease outbreaks or the locations of trees in a forest. Fitting inhomogeneous point process models to such data is straight-forward and illustrates the fundamental process that we will later 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. 14.1) 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, *i.e* the observed number of activity centers.

$$\mathcal{L}(\beta|\mathbf{x}_i) = \prod_{i=1}^R f(x_i)$$

6194 Having defined the likelihood we could choose a prior and obtain the posterior  
 6195 for  $\beta$  using Bayesian methods, or we can find the maximum likelihood estimates  
 6196 (MLEs) using standard numerical methods as is demonstrated below.

6197 First, let's simulate some data. Simulating data under an inhomogeneous  
 6198 point process model is often accomplished using indirect methods such as rejection  
 6199 sampling. Rejection sampling proceeds by simulating data from a standard  
 6200 distribution and then accepting or rejecting each sample using probabilities defined  
 6201 by the distribution of interest. For more information, readers should consult  
 6202 an accessible text such as Robert and Casella (2004). In our example, we  
 6203 simulate from a uniform distribution and then accept or reject using the (scaled)  
 6204 probability density function  $f(x)$ . Note that we first define a spatial covariate  
 6205 (elevation) that is a simple function of the spatial coordinates increasing from  
 6206 the southwest to the northeast of our state-space.<sup>2</sup>

6207 The following **R** commands demonstrate the use of rejection sampling to simulate  
 6208 an inhomogeneous point process for the covariate depicted in Fig. 14.3.1.  
 6209 The code uses the **cuhre** function in the **R2Cuba** package to integrate the intensity  
 6210 function over space (Hahn et al., 2011).

```
6211 # spatial covariate (with mean 0)
6212 elev.fn <- function(x) x[1]+x[2]-1
6213 # intensity function
6214 mu <- function(x, beta) exp(beta*elev.fn(x=x))
6215
6216 # Simulate PP using rejection sampling
6217 set.seed(300225)
6218 N <- 100
6219 count <- 1
6220 s <- matrix(NA, N, 2)
6221 beta <- 2 # parameter of interest
6222 int.mu <- R2Cuba::cuhre(2, 1, mu, beta=beta)$value
6223 elev.min <- elev.fn(c(0,0)) #elev.fn(cbind(0,0))
6224 elev.max <- elev.fn(c(1,1)) #elev.fn(cbind(1,1))
6225 Q <- max(c(exp(beta*elev.min) / int.mu, #2d(beta),
6226           exp(beta*elev.max) / int.mu)) #2d(beta))
6227 while(count <= 100) {
6228   x.c <- runif(1, 0, 1); y.c <- runif(1, 0, 1)
6229   s.cand <- c(x.c,y.c)
6230   pr <- exp(beta*elev.fn(s.cand)) / int.mu #2d(beta)
6231   if(runif(1) < pr/Q) {
6232     s[count,] <- s.cand
6233     count <- count+1
6234   }
6235 }
```

6236 The simulated data are shown in Fig 14.3.1. High elevations are represented  
 6237 by light green and low elevations by dark green. The activity centers of one

---

<sup>2</sup>Such functional forms of covariates are rarely available, which is why continuous spatial covariates are more often measured on a discrete grid.



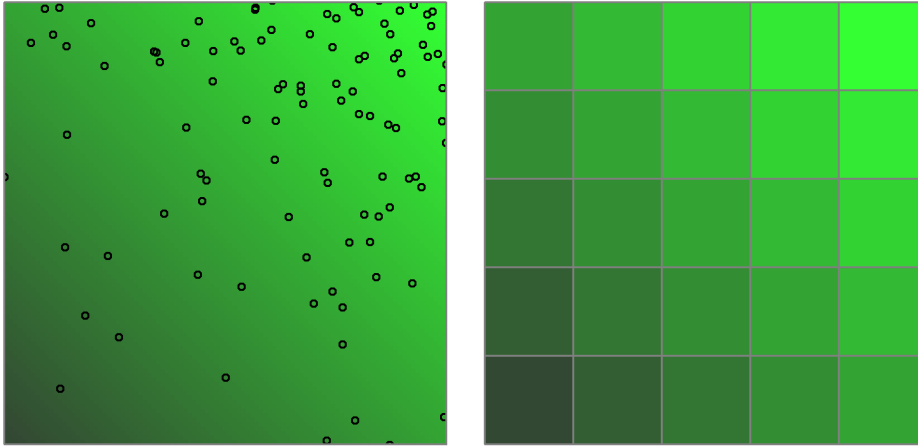


Figure 14.2: An example of a spatial covariate, say elevation, and a realization of an inhomogeneous binomial point process with  $N=100$  and  $\mu(x) = \exp(\beta \text{Elev})$  where  $\beta = 2$ .

hundred animals are shown as points, and it is clear that these simulated animals prefer the high elevations. Perhaps they are mountain goats. The underlying model describing this preference is  $\log(\mu(x)) = \exp(\beta \times \text{Elevation}(x))$  where  $\beta = 2$  is the parameter to be estimated and  $\text{Elevation}(x)$  is a function of the coordinates at  $x$ , as displayed on the map.

Given these points, we will now estimate  $\beta$  by minimizing the negative-log-likelihood using R's `optim` function.

```

6245 # Negative log-likelihood
6246 nll <- function(beta) {
6247   int.mu <- cuhre(2, 1, mu, beta=beta)$value
6248   -sum(beta*elev.fn(s) - log(int.mu))
6249 }
6250 starting.value <- 0
6251 fm <- optim(starting.value, nll, method="Brent",
6252            lower=-5, upper=5, hessian=TRUE)
6253 c(Est=fm$par, SE=sqrt(1/fm$hessian)) # estimates and SEs

```

Maximizing the likelihood took a small fraction of a second, and we obtained an estimate of  $\hat{\beta} = 1.99$ . We could plug in this estimate to our linear model at each point in the state-space to obtain the MLE for the intensity surface.

This example demonstrates that if we had the data we wish we had, *i.e.* if we knew the coordinates of the activity centers, we could easily estimate the parameters governing the underlying point process. Unfortunately, in SCR models, the activity centers cannot be directly observed, but spatial re-captures,

6261 that is captures of individuals at multiple locations in space, provide us with  
6262 the information needed to estimate these latent parameters.

### 6263 14.3.2 Fitting inhomogeneous point process SCR models

#### 6264 Continuous space

6265 One of the nice things about hierarchical models is that they allow us to break  
6266 a problem up into a series of simple conditional relationships. Thus, we can  
6267 simply add the methods described above into our existing MCMC algorithm  
6268 to simulate the posteriors of  $\beta$  conditional on the simulated values of  $\mathbf{s}_i$ . To  
6269 demonstrate, we will continue with the previous example. Specifically, we will  
6270 overlay a grid of traps upon the map shown in Fig. 14.3.1. We will then simulate  
6271 capture histories conditional upon the activity centers shown on the map. Then,  
6272 we will attempt to estimate the activity center locations as though we did not  
6273 know where they were, as is the case in real applications.

6274 Here is some **R** code to simulate the encounter histories under a Poisson  
6275 observation model, which would be appropriate if animals could be detected  
6276 multiple times at a trap during a single occassion.

```
6277 # Create trap locations
6278 xsp <- seq(-0.8, 0.8, by=0.2)
6279 len <- length(xsp)
6280 X <- cbind(rep(xsp, each=len), rep(xsp, times=len))
6281
6282 # Simulate capture histories, and augment the data
6283 ntraps <- nrow(X)
6284 T <- 5
6285 y <- array(NA, c(N, ntraps, T))
6286
6287 nz <- 50 # augmentation
6288 M <- nz+nrow(y)
6289 yz <- array(0, c(M, ntraps, T))
6290
6291 sigma <- 0.1 # half-normal scale parameter
6292 lam0 <- 0.5 # basal encounter rate
6293 lam <- matrix(NA, N, ntraps)
6294
6295 set.seed(5588)
6296 for(i in 1:N) {
6297   for(j in 1:ntraps) {
6298     distSq <- (s[i,1]-X[j,1])^2 + (s[i,2] - X[j,2])^2
6299     lam[i,j] <- exp(-distSq/(2*sigma^2)) * lam0
6300     y[i,j,] <- rpois(T, lam[i,j])
6301   }
6302 }
6303 yz[1:nrow(y),,] <- y # Fill
```

6304 Now that we have a simulated capture-recapture dataset  $y$ , and we have  
6305 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  $\beta$  conditional on all other parameters in the model. The code to do so is:

```

6306 D1 <- cuhre(2, 1, mu, lower=c(xlims[1], ylims[1]),
6307             upper=c(xlims[2], ylims[2]), beta=beta1)$value
6308 beta1.cand <- rnorm(1, beta1, tune[3])
6309 D1.cand <- cuhre(2, 1, mu, lower=c(xlims[1], ylims[1]),
6310             upper=c(xlims[2], ylims[2]), beta=beta1.cand)$value
6311 ll.beta1 <- sum( beta1*elev.fn.v(S) - log(D1) )
6312 ll.beta1.cand <- sum( beta1.cand*elev.fn.v(S) - log(D1.cand) )
6313 if(runif(1) < exp(ll.beta1.cand - ll.beta1) ) {
6314     beta1<-beta1.cand
6315 }

```

Next, we need to put the new prior on the activity centers:

```

6321 #ln(prior), denominator is constant
6322 prior.S <- beta1*cov(S[i,1], S[i,2]) # - log(D1)
6323 prior.S.cand <- beta1*(Scand[1] + Scand[2]) # - log(D1)
6324 if(runif(1)< exp((ll.S.cand+prior.S.cand) - (ll.S+prior.S))) {
6325     S[i,] <- Scand
6326     lam <- lam.cand
6327     D[i,] <- dtmp
6328 }

```

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 14.2. 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 software<sup>3</sup> **secr** 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 straightforward, and virtually all spatial covariates are defined as such.

<sup>3</sup>It is possible, if somewhat cumbersome, to add new distributions in **BUGS**.

Table 14.1: 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

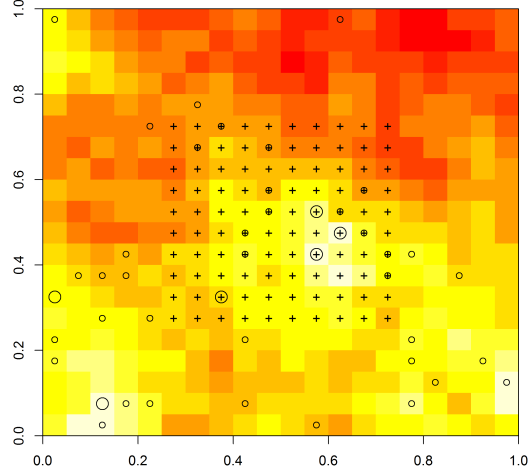


Figure 14.3: Simulated activity centers in discrete space. The spatial covariate, elevation, is highest in the higher areas. Density of activity centers (circles) increases with elevation. Trap locations are shown as crosses.

#### 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 capricaille`. Here we present an analysis of the simulated data shown in the right panel of Fig. 14.3.1. 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.

```

6351 model{
6352   sigma ~ dunif(0, 1)
6353   lam0 ~ dunif(0, 5)
6354   beta ~ dnorm(0,0.1)
6355   psi ~ dbeta(1,1)
6356
6357   for(j in 1:nPix) {
6358     theta[j] <- exp(beta*elevation[j])
6359     probs[j] <- theta[j]/sum(theta[])
6360   }
6361
```

```

6362 for(i in 1:M) {
6363   w[i] ~ dbern(psi)
6364   s[i] ~ dcat(probs[])
6365   x0g[i] <- Sgrid[s[i],1]
6366   y0g[i] <- Sgrid[s[i],2]
6367   for(j in 1:ntraps) {
6368     dist[i,j] <- sqrt(pow(x0g[i]-grid[j,1],2) + pow(y0g[i]-grid[j,2],2))
6369     lambda[i,j] <- lam0*exp(-dist[i,j]*dist[i,j]/(2*sigma*sigma)) * w[i]
6370     y[i,j] ~ dpois(lambda[i,j])
6371   }
6372 }
6373
6374 N <- sum(w[])
6375 Density <- N/1 # unit square
6376 }

```

6377 This model can also be fit in **secr**, which refers to the pixel locations as  
 6378 a “mask”. **R** code to fit the models using **secr** and **JAGS** is available in  
 6379 **scrbook**, see `help(ch9secrYjags)`. Results of the comparison are shown in  
 6380 Table ?? and are very similar as expected.

6381 Density surface maps can be created for fun, and of course to inform man-  
 6382 agement decisions. [describe how to do this]

### 6383 14.3.3 The jaguar data

6384 Estimating density of large felines has been a priority for many conservation  
 6385 organizations, but no robust methodologies existed before the advent of SCR.  
 6386 Distance sampling is not feasible for such rare and cryptic species, and tradi-  
 6387 tional capture-recapture methods yield estimates that are highly sensitive to the  
 6388 subjective choice of the effective survey area. In this example, we demonstrate  
 6389 how readily density can be estimated for a globally imperilled species using  
 6390 SCR. Furthermore, we show how inhomogeneous point process models can be  
 6391 used to test important hypotheses regarding the factors affecting density.  
 6392 [describe study]

Table 14.2: Comparison of **secr** and **JAGS** results

Software	Par	Est.	SD	lower	upper
secr	$N$	49.2803	5.7535	41.0087	64.3879
	$\beta$	2.1772	0.5628	1.0741	3.2804
	$\lambda_0$	0.9203	0.0764	0.7824	1.0825
	$\sigma$	0.0990	0.0038	0.0918	0.1068
JAGS	$N$	48.2072	5.4053	39.0000	60.0000
	$\beta$	2.1026	0.5323	1.0889	3.1506
	$\lambda_0$	0.9328	0.0766	0.7898	1.0921
	$\sigma$	0.1004	0.0041	0.0929	0.1089

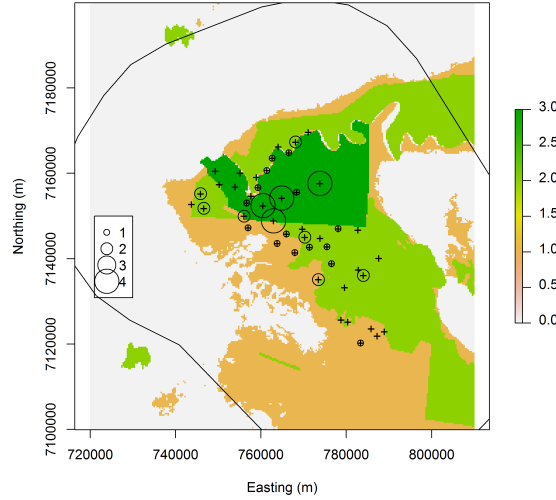


Figure 14.4: Jaguar detections

A few aspects of this design are noteworthy. First, the dimensions and configuration of the trap array differed among the regions of the trap array. This fact alone could explain variation in the number of animals exposed to sampling, which would have no biological meaning. Furthermore, the area of inference is an irregular polygon that was not sampled uniformly. Only by estimating density can we hope to extrapolate our estimates from the sampled region to get what we are after. In this case, this is readily accomplished since the entire state-space can be classified as one of the 3 levels of protection from poaching. Of course, it general it is always preferable to sample more uniformly throughout the area of interest in case some unmeasured covariate biases the extrapolation.

To assess the influence of poaching on jaguar density, we considered 2 metrics of poaching pressure, one political and one continuous measure of accessibility (Fig xxx).

## 14.4 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 yields a model of 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.

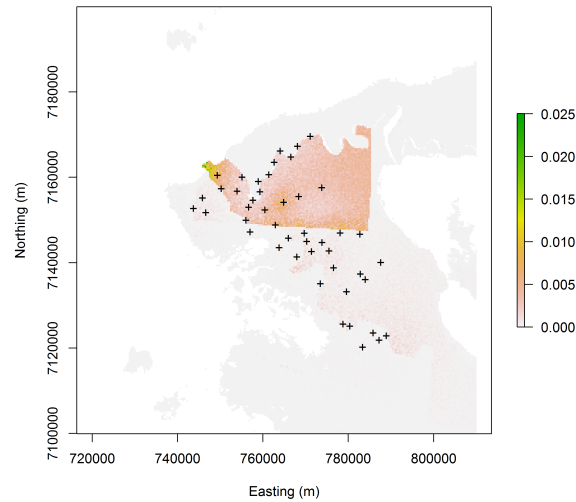


Figure 14.5: Estimated density surface for the jaguar dataset

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**.

All the examples in this section included a single state-space covariate, but this was for simplicity only. Including multiple covariates poses no additional challenges. Likewise, additional model structure such sex-specific encounter rate parameters or behavioral responses can be accommodated.

## 14.5 Other ideas

Should have some discussion on some ideas for building flexible models. Might be cool to use the Ickstadt/Wolpert as a model for the inhomogeneous point process. Don't have to do it, just mention it. Also some kind of a spline model or similar.





## 6426 Chapter 15

# 6427 Inhomogeneous Point 6428 Process



## 6429 Chapter 16

## 6430 Open models



# 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.
- Berger, J. O., Liseo, B., and Wolpert, R. L. (1999), “Integrated likelihood methods for eliminating nuisance parameters,” *Statistical Science*, 1–22.
- 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.

- 6458 Burnham, K. P. and Overton, W. S. (1978), “Estimation of the size of a closed  
6459 population when capture probabilities vary among animals,” *Biometrika*, 65,  
6460 625.
- 6461 Casella, G. and George, E. I. (1992), “Explaining the Gibbs sampler,” *American*  
6462 *Statistician*, 46, 167–174.
- 6463 Chandler, R. and Royle, J. (2012), “Spatially-explicit models for inference about  
6464 density in unmarked populations,” *Biometrics (in review)*.
- 6465 Chandler, R., Royle, J., and King, D. (2011), “Inference about density and  
6466 temporary emigration in unmarked populations,” *Ecology*, 92, 1429–1435.
- 6467 Converse, S. J. and Royle, J. A. (2010), “Missing,” *Missing*, Missing.
- 6468 Coull, B. A. and Agresti, A. (1999), “The Use of Mixed Logit Models to Reflect  
6469 Heterogeneity in Capture-Recapture Studies,” *Biometrics*, 55, 294–301.
- 6470 Dawson, D. and Efford, M. (2009), “Bird population density estimated from  
6471 acoustic signals,” *Journal of Applied Ecology*, 46, 1201–1209.
- 6472 Dice, L. R. (1938), “Some census methods for mammals,” *Journal of Wildlife*  
6473 *Management*, 2, 119–130.
- 6474 Dorazio, R. and Royle, J. (2003), “Mixture models for estimating the size of a  
6475 closed population when capture rates vary among individuals,” *Biometrics*,  
6476 351–364.
- 6477 Dorazio, R. M. (2007), “On the choice of statistical models for estimating oc-  
6478 currence and extinction from animal surveys,” *Ecology*, 88, 2773–2782.
- 6479 Durbin and Elston (2012), “Missing,” *Missing*, Missing.
- 6480 Efford, M. (2004), “Density estimation in live-trapping studies,” *Oikos*, 106,  
6481 598–610.
- 6482 — (2011), “secre-spatially explicit capture-recapture in R,” .
- 6483 Efford, M., Dawson, D., and Robbins, C. (2004), “DENSITY: software for  
6484 analysing capture-recapture data from passive detector arrays,” *Animal Bio-*  
6485 *diversity and Conservation*, 217–228.
- 6486 Fienberg, S. E., Johnson, M. S., and Junker, B. W. (1999), “Classical multilevel  
6487 and Bayesian approaches to population size estimation using multiple lists,”  
6488 *Journal of the Royal Statistical Society of London A*, 163, 383–405.
- 6489 Gardner, B. (2009), “missing,” *Missing*, missing.
- 6490 Gardner, B., Royle, J., Wegan, M., Rainbolt, R., and Curtis, P. (2010), “Esti-  
6491 mating black bear density using DNA data from hair snares,” *The Journal of*  
6492 *Wildlife Management*, 74, 318–325.

- 6493 Gelfand, A. and Smith, A. (1990), “Sampling-based approaches to calculating  
6494 marginal densities,” *Journal of the American statistical association*, 85, 398–  
6495 409.
- 6496 Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004), *Bayesian*  
6497 *data analysis, second edition.*, Boca Raton, Florida, USA: CRC/Chapman  
6498 & Hall.
- 6499 Gelman, A., Meng, X. L., and Stern, H. (1996), “Posterior predictive assessment  
6500 of model fitness via realized discrepancies,” *Statistica Sinica*, 6, 733–759.
- 6501 Geman, S. and Geman, D. (1984), “Stochastic relaxation, Gibbs distributions,  
6502 and the Bayesian restoration of images,” *IEEE Transactions on Pattern Anal-*  
6503 *ysis and Machine Intelligence*, PAMI-6, 721–741.
- 6504 Genz, A. S., Meyer, M. R., Lumley, T., and Maechler, M. (2007), “The adapt  
6505 Package. R package version 1.0-4,” .
- 6506 Gilks, W. and Wild, P. (1992), “Adaptive rejection sampling for Gibbs sam-  
6507 pling,” *Applied Statistics*, 41, 337–348.
- 6508 Gilks, W. R., Thomas, A., and Spiegelhalter, D. J. (1994), “A Language  
6509 and Program for Complex Bayesian Modelling,” *Journal of the Royal Sta-*  
6510 *tistical Society. Series D (The Statistician)*, 43, 169–177, ArticleType: pri-  
6511 mary\_article / Issue Title: Special Issue: Conference on Practical Bayesian  
6512 Statistics, 1992 (3) / Full publication date: 1994 / Copyright 1994 Royal  
6513 Statistical Society.
- 6514 Gopalaswamy (2012), “Missing,” *Missing*, missing.
- 6515 Gopalaswamy, A. M., Royle, A. J., Hines, J., Singh, P., Jathanna, D., Kumar,  
6516 N. S., and Karanth, K. U. (2011), *A Program to Estimate Animal Abun-*  
6517 *dance and Density using Spatially-Explicit Capture-Recapture*, r package ver-  
6518 sion 1.0.4.
- 6519 Hahn, T., Bouvier, A., and Kieu, K. (2011), “Package ‘R2Cuba’ R package  
6520 version 1.0-6,” .
- 6521 Hastings, W. (1970), “Monte Carlo sampling methods using Markov chains and  
6522 their applications,” *Biometrika*, 57, 97–109.
- 6523 Hawkins, C. and Racey, P. (2005), “Low population density of a tropical forest  
6524 carnivore, *Cryptoprocta ferox*: implications for protected area management,”  
6525 *Oryx*, 39, 35–43.
- 6526 Hestbeck (1991), “Missing,” *Missing*, Missing, Missing.
- 6527 Huggins, R. M. (1989), “On the statistical analysis of capture experiments,”  
6528 *Biometrika*, 76, 133.
- 6529 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.
- 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*, 60, 65–81.
- Laird, N. M. and Ware, J. H. (1982), “Random-effects models for longitudinal  
data,” *Biometrics*, 963–974.
- Langtimm (2010), “Missing,” *Missing*, missing.



- 6564 Le Cam, L. (1990), "Maximum likelihood: an introduction," *International Sta-*  
6565 *tistical Review/Revue Internationale de Statistique*, 153–171.
- 6566 Lewin-Koh, N. J., Bivand, R., contributions by Edzer J. Pebesma, Archer, E.,  
6567 Baddeley, A., Bibiko, H.-J., Dray, S., Forrest, D., Friendly, M., Giraudoux, P.,  
6568 Golicher, D., Rubio, V. G., Hausmann, P., Hufthammer, K. O., Jagger, T.,  
6569 Luque, S. P., MacQueen, D., Niccolai, A., Short, T., Stabler, B., and Turner,  
6570 R. (2011), *maptools: Tools for reading and handling spatial objects*, r package  
6571 version 0.8-10.
- 6572 Link, W. A. (2003), "Missing," *missing*, missing.
- 6573 Link, W. A. and Barker, R. J. (2009), *Bayesian Inference: With Ecological*  
6574 *Applications*, London, UK: Academic Press.
- 6575 Liu and Wu (1999), "Parameter expansion for data augmentation," *J Am Stat*  
6576 *Assoc*, 94, 1264–1274.
- 6577 MacEachern, S. and Berliner, L. (1994), "Subsampling the Gibbs sampler,"  
6578 *American Statistician*, 188–190.
- 6579 MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Royle, J. A., and  
6580 Langtimm, C. A. (2002), "Estimating site occupancy rates when detection  
6581 probabilities are less than one," *Ecology*, 83, 2248–2255.
- 6582 Mackenzie, D. I. and Royle, J. (2005), "Designing occupancy studies: general  
6583 advice and allocating survey effort," *Journal of Applied Ecology*, 42, 1105–  
6584 1114.
- 6585 Magoun, A. J., Long, C. D., Schwartz, M. K., Pilgrim, K. L., Lowell, R. E.,  
6586 and Valkenburg, P. (2011), "Integrating motion-detection cameras and hair  
6587 snags for wolverine identification," *The Journal of Wildlife Management*, 75,  
6588 731–739.
- 6589 McCarthy, M. A. (2007), *Bayesian Methods for Ecology*, Cambridge: Cambridge  
6590 University Press.
- 6591 McCullagh, P. and Nelder, J. (1989), *Generalized linear models*, Chapman &  
6592 Hall/CRC.
- 6593 Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., Teller, E., et al.  
6594 (1953), "Equation of state calculations by fast computing machines," *The*  
6595 *journal of chemical physics*, 21, 1087–1092.
- 6596 Metropolis, N. and Ulam, S. (1949), "The Monte Carlo method," *Journal of the*  
6597 *American Statistical Association*, 44, 335–341.
- 6598 Millar, R. (2009), "Comparison of hierarchical Bayesian models for overdis-  
6599 persed count data using DIC and Bayes' Factors," *Biometrics*, 65, 962–969.

- 6600 Mollet, P., Kery, M., Gardner, B., Pasinelli, G., and A, R. J. (2012), “Popu-  
6601 lation size estimation for capercaillie (*Tetrao urogallus* L.) using DNA-based  
6602 individual recognition and spatial capture-recapture models,” *missing*, miss-  
6603 ing, missing.
- 6604 Neal, R. (2003), “Slice sampling,” *Annals of Statistics*, 31, 705–741.
- 6605 Nelder, J. and Wedderburn, R. (1972), “Generalized linear models,” *Journal of*  
6606 *the Royal Statistical Society. Series A (General)*, 370–384.
- 6607 Norris III, J. L. and Pollock, K. H. (1996), “Nonparametric MLE under two  
6608 closed capture-recapture models with heterogeneity,” *Biometrics*, 639–649.
- 6609 Pebesma, E. and Bivand, R. (2011), *Package ‘sp’*, r package version 0.9-91.
- 6610 Pledger, S. (2000), “Unified maximum likelihood estimates for closed capture-  
6611 recapture models using mixtures,” *Biometrics*, 434–442.
- 6612 Plummer, M. (2003), “JAGS: A program for analysis of Bayesian graphical mod-  
6613 els using Gibbs sampling,” in *Proceedings of the 3rd International Workshop*  
6614 *on Distributed Statistical Computing (DSC 2003)*. March, pp. 20–22.
- 6615 — (2009), “rjags: Bayesian graphical models using mcmc. R package version  
6616 1.0. 3-12,” .
- 6617 — (2011), *rjags: Bayesian graphical models using MCMC*, r package version  
6618 3-5.
- 6619 Plummer, M., Best, N., Cowles, K., and Vines, K. (2006), “CODA: Convergence  
6620 Diagnosis and Output Analysis for MCMC,” *R News*, 6, 7–11.
- 6621 Robert, C. P. and Casella, G. (2004), *Monte Carlo statistical methods*, New  
6622 York, USA: Springer.
- 6623 — (2010), *Introducing Monte Carlo Methods with R*, New York, USA: Springer.
- 6624 Roberts, G. O. and Rosenthal, J. S. (1998), “Optimal scaling of discrete ap-  
6625 proximations to Langevin diffusions,” *Journal of the Royal Statistical Society:*  
6626 *Series B (Statistical Methodology)*, 60, 255–268.
- 6627 Royle, J. (2006), “Site occupancy models with heterogeneous detection proba-  
6628 bilities,” *Biometrics*, 62, 97–102.
- 6629 — (2009), “Analysis of capture-recapture models with individual covariates us-  
6630 ing data augmentation,” *Biometrics*, 65, 267–274.
- 6631 Royle, J. and Dorazio, R. (2006), “Hierarchical models of animal abundance and  
6632 occurrence,” *Journal of Agricultural, Biological, and Environmental Statis-*  
6633 *tics*, 11, 249–263.
- 6634 — (2008), *Hierarchical modeling and inference in ecology: the analysis of data*  
6635 *from populations, metapopulations and communities*, Academic Press.

- 6636 — (2010), “Missing,” *Missing*, missing, missing.
- 6637 — (2011), “Missing,” *Missing*, missing, missing.
- 6638 Royle, J., Dorazio, R., and Link, W. (2007), “Analysis of multinomial models  
6639 with unknown index using data augmentation,” *Journal of Computational  
6640 and Graphical Statistics*, 16, 67–85.
- 6641 Royle, J. and Dubovsky, J. (2001), “Modeling spatial variation in waterfowl  
6642 band-recovery data,” *The Journal of wildlife management*, 726–737.
- 6643 Royle, J., Karanth, K., Gopalaswamy, A., and Kumar, N. (2009), “Bayesian  
6644 inference in camera trapping studies for a class of spatial capture-recapture  
6645 models,” *Ecology*, 90, 3233–3244.
- 6646 Royle, J. and Link, W. (2006), “Generalized site occupancy models allowing for  
6647 false positive and false negative errors,” *Ecology*, 87, 835–841.
- 6648 Royle, J. and Nichols, J. (2003), “Estimating abundance from repeated presence-  
6649 absence data or point counts,” *Ecology*, 84, 777–790.
- 6650 Royle, J. A. (2004a), “Generalized estimators of avian abundance from count  
6651 survey data,” *Animal Biodiversity and Conservation*, 27, 375–386.
- 6652 — (2004b), “Missing,” *Missing*, missing.
- 6653 — (2008), “Modeling individual effects in the Cormack–Jolly–Seber model: a  
6654 state–space formulation,” *Biometrics*, 64, 364–370.
- 6655 — (2010), “missing,” *missing*, missing.
- 6656 Royle, J. A., Converse, S., and Link, W. (2011a), “Data augmentation for struc-  
6657 tured populations,” *unpublished*.
- 6658 Royle, J. A., Kéry, M., and Guélat, J. (2011b), “Spatial capture-recapture mod-  
6659 els for search-encounter data,” *Methods in Ecology and Evolution*, 1–10.
- 6660 Royle, J. A., Magoun, A. J., Gardner, B., Valkenburg, P., and Lowell, R. E.  
6661 (2011c), “Density estimation in a wolverine population using spatial capture-  
6662 recapture models,” *The Journal of Wildlife Management*, 75, 604–611.
- 6663 Royle, J. A. and Young, K. V. (2008), “A Hierarchical Model For Spatial  
6664 Capture-Recapture Data,” *Ecology*, 89, 2281–2289.
- 6665 Sanathanan, L. (1972), “Estimating the size of a multinomial population,” *The  
6666 Annals of Mathematical Statistics*, 142–152.
- 6667 Schofield and Barker (missing), “Missing,” *Missing*, missing.
- 6668 Sepúlveda, M., Bartheld, J., Monsalve, R., Gómez, V., and Medina-Vogel, G.  
6669 (2007), “Habitat use and spatial behaviour of the endangered Southern river  
6670 otter (*Lontra provocax*) in riparian habitats of Chile: conservation implica-  
6671 tions,” *Biological Conservation*, 140, 329–338.

- 6672 Sillett (2011), “Missing,” *Missing*, missing.
- 6673 Spiegelhalter, D., Thomas, A., Best, N., and Lunn, D. (2003), *WinBUGS User*  
6674 *Manual Version 1.4*.
- 6675 Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002),  
6676 “Bayesian measures of model complexity and fit,” *Journal of the Royal Sta-*  
6677 *tistical Society. Series B, Statistical Methodology*, 583–639.
- 6678 Stabler, B. (2006), *shapefiles: Read and Write ESRI Shapefiles*, r package version  
6679 0.6.
- 6680 Sturtz, S., Ligges, U., and Gelman, A. (2005), “R2WinBUGS: A Package for  
6681 Running WinBUGS from R,” *Journal of Statistical Software*, 12, 1–16.
- 6682 Su, Y.-S. and Yajima, M. (2011), *R2jags: A Package for Running jags from R*,  
6683 r package version 0.02-17.
- 6684 Tanner, M. A. and Wong, W. H. (1987), “The calculation of posterior distribu-  
6685 tions by data augmentation,” *J Am Stat Assoc*, 82, 528–540.
- 6686 Thomas, A., O’Hara, B., Ligges, U., and Sturtz, S. (2006), “Making BUGS  
6687 Open,” *R News*, 6, 12–17.
- 6688 Trolle, M. and Kéry, M. (2005), “Camera-trap study of ocelot and other secretive  
6689 mammals in the northern Pantanal,” *Mammalia*, 69, 409–416.
- 6690 Tyre, A. J., Tenhumberg, B., Field, S. A., Niejalke, D., Parris, K., and Poss-  
6691 ingham, H. P. (2003), “Improving precision and reducing bias in biological  
6692 surveys: estimating false-negative error rates,” *Ecological Applications*, 13,  
6693 1790–1801.
- 6694 Wegan (missing), “missing,” *missing*, missing.
- 6695 Yang, H. C. and Chao, A. (2005), “Modeling Animals’ Behavioral Response by  
6696 Markov Chain Models for Capture–Recapture Experiments,” *Biometrics*, 61,  
6697 1010–1017.
- 6698 Zuur, A., Ieno, E., Walker, N., Saveliev, A., and Smith, G. (2009), *Mixed effects*  
6699 *models and extensions in ecology with R*, Springer Verlag.

# Index

<sup>6700</sup> bracket notation, 8

# Index

<sup>6701</sup> bracket notation, 8