CHAPTER

Closed Population Models

4

In this chapter we introduce ordinary *non-spatial* capture-recapture (CR) models for estimating population size in closed populations. A closed population is one whose size, N, does not change during the study. Two forms of closure are often discussed: demographic closure, meaning that no births or deaths occur, and geographic closure, which states that no individuals move onto or off of the sampled area during the study. Although few populations are actually closed except during very short time intervals, closed population CR models serve as the basis for the development of the rest of the models presented in this book, including the models for open populations discussed in Chapter 16.

We begin with the most basic capture-recapture model, colloquially referred to as "model M_0" (Otis et al., 1978), in which encounter probability is strictly constant in all respects (across individuals, and ~~replicates~~). This allows us to highlight the basic structure of closed population models as binomial GLMs. We then consider some important extensions of ordinary closed population models that accommodate various types of "individual effects"—either in the form of explicit, observed covariates (sex, age, body mass) or unstructured "heterogeneity" in the form of an individual random effect, which represent unobserved or unmeasured covariates. A special type of individual covariate model~~s~~ is distance sampling, which could be thought of as the most primitive spatial capture-recapture model. All of these different types of closed population models are closely related to binomial (or logistic) regression-type models. In fact, when N is known, they are precisely logistic regression models.

We emphasize Bayesian analysis of capture-recapture models and we accomplish this using a method related to classical "data augmentation" from the statistics literature (e.g., Tanner and Wong, 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; Royle and Dorazio, 2012). 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 Kéry and Schaub (2012, chapter 6) for a complementary development of Bayesian analysis of ordinary, i.e., non-spatial closed population models.

Spatial Capture-Recapture. http://dx.doi.org/10.1016/B978-0-12-405939-9.00004-9

4.1 The simplest closed population model: model M_0

To start looking at the simplest capture-recapture model, let's suppose there exists a population of N individuals which we subject to repeated sampling, say over K "occasions," such as trap nights, where individuals are captured, marked, released, 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. As an example, suppose $K = 5$ sampling occasions, then an individual captured during occasions 2 and 3 but not otherwise would have an encounter history $\mathbf{y} = (0, 1, 1, 0, 0)$. Thus, the observation $\mathbf{y}_i$ for each individual ($i = 1, 2, \ldots, N$) is a vector having elements denoted by y_{ik} for $k = 1, 2, \ldots, K$. 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 independent and identically distributed ("iid") Bernoulli random variables and we may write $y_{ik} \sim \text{Bernoulli}(p)$. Consequently, for this very simple model in which p is constant (i.e., there are no individual or temporal covariates that affect p) the original binary detection variables can be aggregated into the total number of encounters for each individual,[1] $y_{i\cdot} = \sum_k y_{ik}$, and the observation model changes from a Bernoulli distribution to a binomial distribution based on a sample of size K. That is

$$y_{i\cdot} = \sum_k y_{ik} \sim \text{Binomial}(K, p)$$

for every individual in the population $i = 1, 2, \ldots, N$, where N is the number of individuals in the population (i.e., population size).

Table 4.1 A toy capture-recapture data set with $n = 6$ observed individuals and $K = 5$ sample occasions. Under a model with constant encounter probability, the binary detection history data can be summarized in the detection frequency (the total number of detections, $y_{i\cdot}$), which is shown in the rightmost column.

	Sample Occasion					
Indiv *i*	**1**	**2**	**3**	**4**	**5**	y_i
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

[1] We use the common "dot notation" to denote having summed over one or more indices of a variable. $y_i = \sum_j y_{ij}$, $y_{\cdot\cdot} = \sum_i \sum_j y_{ij}$, etc.

We emphasize the central importance of the basic Bernoulli encounter model—an individual is either encountered in a sample, or not—which forms the cornerstone of almost all of classical capture-recapture models, including many spatial capture-recapture models discussed in this book. Evidently, the basic capture-recapture model is 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 a slight oversimplification. In fact, we are proceeding here as if we knew N. In practice we don't, of course, and estimating N is actually the central objective. But, by proceeding as if N were known, we can specify a simple model using standard methods that you are already familiar with (i.e., GLMs - see Chapter 3) and then deal with the fact that N is unknown latter.

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

$$\Pr(y_1, \ldots, y_N | p) = \prod_{i=1}^{N} \text{Binomial}(y_i | K, p). \tag{4.1.1}$$

We emphasize that this expression is conditional on N, in which case we get to observe the $y_i = 0$ observations and the resulting data are just iid binomial counts. Because this is a binomial regression model of the variety described in Chapter 3, fitting this model using a **BUGS** engine poses no difficulty.

Equation (4.1.1) can be simplified even further if we reformat the observations as encounter frequencies. Specifically, let n_k denote the number of individuals captured exactly k times after K survey occasions, $n_k = \sum_{i=1}^{N} I(y_i = k)$ where $I()$ is the indicator function evaluating to 1 if its argument is true and 0 otherwise. For sake of illustration, we converted the data from Table 4.1 to this format (Table 4.2). What is important to note is that if we know N, then we know n_0, i.e., the number of individuals not captured. In this case, an alternative and equivalent expression to Eq. (4.1.1) is

$$\Pr(y_1, \ldots, y_N | p) = \prod_{k=0}^{K} \pi_k^{n_k} \tag{4.1.2}$$

where $\pi_k = \Pr(y = k)$ under the binomial model with parameter p and sample size K. The essential problem in capture-recapture, however, is that N is *not* known because the number of uncaptured individuals (n_0) is unknown. Consequently, the observed

Table 4.2 Data from Table 4.1 formatted as capture frequencies. Since N is unknown, the number of individuals not captured (n_0) is also unknown.

	k					
	0	**1**	**2**	**3**	**4**	**5**
Number of individuals captured k times (n_k)	$N - 6$	2	3	1	0	0

capture frequencies n_k are no longer independent because n_0 is a function of the other frequencies, $n_0 = N - \sum_{k=1}^{K} n_k$. Hence, their joint distribution is multinomial (e.g., see Illian et al. (2008, p. 61)):

$$(n_0, n_1, \ldots, n_K) \sim \text{Multinomial}(N, \pi_0, \pi_1, \ldots, \pi_K) \tag{4.1.3}$$

We gave a general overview of the multinomial distribution in Section 2.2. The multinomial distribution is the standard model for discrete responses that can fall into a fixed number ($K+1$ in this case) of possible categories. In the context of capture-recapture, the multinomial posits a population of N individuals with $K+1$ possible outcomes defined by the possible encounter frequencies: encountered $y = 1, 2, \ldots, K$ times, or not encountered at all. These possible outcomes occur with probabilities π_k, which we refer to as "cell probabilities," or in the specific context of capture-recapture, encounter history probabilities.

To fit the model in which N is *unknown*, we can regard n_0 as a parameter and maximize the multinomial likelihood directly (see Section 2.3 on parameter estimation). Direct likelihood analysis of the multinomial model is straightforward, but that is not always sufficiently useful in practice because we seldom are concerned with models for the aggregated encounter history frequencies, which entail that capture probabilities are the same for all individuals. In many instances, including for spatial capture-recapture (SCR) models, we require a formulation of the model that can accommodate individual-level covariates to account for differences in detection among individuals, which we address subsequently in this chapter, and in Chapter 7.

4.1.1 The core capture-recapture assumptions

This basic capture-recapture model—model M_0—comes with a host of specific biological and statistical assumptions. In addition to the basic assumption of population closure, Otis et al. (1978) list the following:

1. animals do not lose their marks during the experiment,
2. all marks are correctly noted and recorded at each trapping occasion, and
3. each animal has a constant and equal probability of capture on each trapping occasion.

The remainder of their classic work is dedicated to relaxing assumption 3. While assumptions 1 and 2 are undoubtedly necessary for inference from basic CR methods to be valid, and while they are also assumed by most of the models we present in the following chapters, we refrain from repeatedly making such statements. Our opinion is that all model assumptions are apparent when a model is clearly specified, and it is both redundant and impossible to list all the things not allowed by the model. For example, closed population models also assume that other sources of error do not occur, but it is not necessary to enumerate each possibility. Rather, it is necessary to make clear statements such as

$$y_i \overset{iid}{\sim} \text{Bernoulli}(p) \quad \text{for } i = 1, \ldots, N.$$

This simple model description carries a tremendous amount of information, and it leaves very little left to say with respect to assumptions. Although we will not always show the iid symbol, it will be assumed unless otherwise noted, and this assumption is critical for valid inference. It implies that the encounter of one individual does not affect the encounter of another individual, and encounter does not affect future encounter. Under this assumption, it is easy to write down the likelihood of the parameters and obtain parameter estimates; however, whether or not it is true depends upon biological and sampling issues. If this assumption is deemed false, the model can be discarded in favor of a more realistic alternative. However, once we have settled on our model, statistical inference proceeds by assuming the model is truth—not an approximation to truth—but actual truth.

In spite of the fact that we assume that all models are truth, but we acknowledge that all models are wrong due to their assumptions, assumptions should not be viewed as a necessary evil. In fact, one way to view assumptions is as embodiments of our ecological hypotheses. If we make these assumptions too complex or too specific, then we will never be able to study general phenomena that hold true across space and time. Furthermore, in practice, we will rarely have enough data to estimate the parameters of highly complex models.

4.1.2 Conditional likelihood

We saw that the 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 that can be estimated. This multinomial model, being conditional on N, is sometimes referred to as the "joint likelihood" the "full likelihood," or the "unconditional likelihood" (Sanathanan, 1972; Borchers et al., 2002). 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, if and only if $y > 0$. The observation model is therefore based on $\Pr(y|y > 0)$. For the simple case of model M_0, 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. Both the conditional and unconditional models are legitimate modes of analysis in all capture-recapture types of studies. 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 (described in Section 4.2 below), 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 three distinct formulations of the model—or modes of analysis—for analyzing all capture-recapture models based on the (1) binomial model for the joint or unconditional specification; (2) zero-truncated binomial that arises "conditional on n"; and (3) the zero-inflated

Table 4.3 Modes of analysis of capture-recapture models. Closed population models can be analyzed using the joint or "full likelihood" which contains N as an explicit parameter, the conditional likelihood, which does not involve N, or by data augmentation which replaces N with ψ. Each approach yields a distinct likelihood.

Mode of Analysis	Parameters in Model	Statistical Model
Joint likelihood	p, N	Multinomial with index N
Conditional likelihood	p	Zero-truncated binomial
Data augmentation	p, ψ	Zero-inflated binomial

binomial that arises under data augmentation. Each formulation has distinct model parameters (shown in Table 4.3 for model M_0).

4.2 Data augmentation

We consider a method of analyzing closed population models using parameter-expanded data augmentation (PX-DA), which we abbreviate to "data augmentation" or DA, which is useful for Bayesian analysis and, in particular, analysis of models using the various **BUGS** engines and other Bayesian model fitting 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 in a number of different contexts including individual covariate models (Royle, 2009b), open population models (Royle and Dorazio, 2008, 2012; Gardner et al., 2010a), spatial capture-recapture models (Royle and Young, 2008; Royle et al., 2009a; Gardner et al., 2009), and many others. Kéry and Schaub (2012, Chapters 6 and 10) provide a good introduction to data augmentation in the context of closed and open population models.

Conceptually, the technique of data augmentation represents a reparameterization of the "complete data" model—i.e., that conditional on N. The reparameterization is achieved by embedding this data set into a larger data set having $M > N$ "rows" (individuals) and re-expressing the model conditional on M instead of N. The great thing about data augmentation is that we do not need to know N for this reparameterization. Although this has a whiff of arbitrariness or even outright ad hockery, ~~practically, in the choice of M,~~ it is always possible to choose M easily for a given problem and context, and results will be insensitive to choice of M.[2] 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

[2] Unless the data set is sufficiently small that parameters are weakly identified.

the augmentation. This is achieved by a Bernoulli sampling process that determines whether an individual in M is also a member of N. Inference is focused on estimating the proportion $\psi = E[N]/M$, where ψ is the "data augmentation parameter."

4.2.1 DA links occupancy models and closed population models

There is a close correspondence between so-called "occupancy" models and closed population models (see Royle and Dorazio, 2008, Section 5.6). 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 the sites. This yields encounter data such as that illustrated in the left panel of Table 4.4. The important problem is that a species may occur at a site, but go undetected, yielding an all-zero encounter history for the site, which in the case of occupancy studies, are *observed*. However, some of the ~~zero vectors~~ will typically correspond to sites where the species in fact *does* occur. Thus, while the zeros are observed, there are too many of them and, in a sense, the inference problem is to partition the zeros into "structural" (fixed) and

Table 4.4 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	3	0	1	0	3	0	1	0
4	1	0	1	4	1	0	1	4	1	0	1
5	0	1	1	5	0	1	1	5	0	1	1
.	0	1	1	.	0	1	1	.	0	1	1
.	1	1	1	.	1	1	1	.	1	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

"sampling" (or stochastic) zeros, where the former are associated with unoccupied sites and the latter with occupied sites where the species went undetected. More formally, inference is focused on the parameter ψ, the probability that a site is occupied.

In contrast to occupancy studies, in classical closed population studies, we observe a data set as in the middle panel of Table 4.4 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. This objective (how many sampling zeros?) is precisely the same for both occupancy and CR methods 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 in the sample), whereas a natural choice of M for capture-recapture models may not be obvious. However, the choice of M implies a uniform prior for N on the integers $[0, M]$ (Royle et al., 2007). Thus, 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, where the occupancy or data augmentation parameter (ψ) takes the place of the abundance parameter (N).

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 4.4, and then analyze the augmented data set using an occupancy-type model that includes both "unoccupied sites" (in capture-recapture, augmented individuals that are not members of the real population that was sampled) as well as "occupied sites" at which detections did not occur (in capture-recapture, individuals that are members of the population but ~~that~~ were undetected by sampling). 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-zero rows 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, $z_1, z_2, \ldots, z_M$, 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 $z_i \sim \text{Bernoulli}(\psi)$ where ψ 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 a structural zero in the augmented data set. The zero-inflated binomial model that arises under data augmentation can be formally expressed by the following set of assumptions (we include typical priors for a Bayesian analysis):

$$
\begin{aligned}
y_i | z_i = 1 &\sim \text{Binomial}(K, p) \\
y_i | z_i = 0 &\sim I(y = 0) \\
z_i &\overset{iid}{\sim} \text{Bernoulli}(\psi)
\end{aligned}
$$

$$\psi \sim \text{Uniform}(0, 1)$$
$$p \sim \text{Uniform}(0, 1)$$

for $i = 1, \ldots, M$, where $I(y = 0)$ is a point mass at $y = 0$. It is sometimes convenient to express the conditional-on-z observation model concisely in just one step:

$$y_i | z_i \sim \text{Binomial}(K, z_i\, p)$$

and we understand this to mean, if $z_i = 0$, then y_i is necessarily 0 because its success probability is $z_i\, p = 0$.

Note that, under data augmentation, N is no longer an explicit parameter of this model. In its place, we estimate ψ and functions of the latent variables z. In particular, under the assumptions of the zero-inflated model, $z_i \overset{iid}{\sim} \text{Bernoulli}(\psi)$; therefore, N is a function of these latent variables:

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

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

$$\Pr(y_1, \ldots, y_M | p, \psi) = \prod_{i=1}^{M} \left(\psi \times \text{Binomial}(y_i | K, p) + I(y_i = 0)(1 - \psi)\right) \tag{4.2.1}$$

Interpreted as a likelihood, we can directly maximize this expression to obtain the MLEs of the structural parameters ψ and p or those of other more complex models (e.g., see Royle, 2006). We could estimate these parameters and then use them to obtain an estimator of N using the so-called "Best unbiased predictor" (see Royle and Dorazio, 2012). Normally, however, we will analyze the model in its "conditional-on-z" form using methods of MCMC either in the **BUGS** engines or using our own MCMC algorithms (Chapter 17).

4.2.2 Model M_0 in BUGS

It is helpful to understand data augmentation by seeing what its effect is on implementing model M_0. For this model, in which we can aggregate the encounter data to individual-specific encounter frequencies, the augmented data are given by the vector of frequencies $(y_1, \ldots, y_n, 0, 0, \ldots, 0)$ where the augmented values of $y = 0$ represent the encounter frequency for potential individuals $y_{n+1}, \ldots, y_M$. The zero-inflated model of the augmented data combines the model of the latent variables, $z_i \sim \text{Bernoulli}(\psi)$. The **BUGS** model description of the closed population model M_0 is shown in Panel 4.1. The last line of the model specification provides the expression for computing N from the data augmentation variables z_i. Note that, to improve readability of code snippets (especially of large ones), we will sometimes deviate from

```
model{
p  ~ dunif(0,1)
psi ~ dunif(0,1)

# nind = number of individuals captured at least once
#   nz = number of uncaptured individuals added for DA
for(i in 1:(nind+nz)){
    z[i] ~ dbern(psi)
   mu[i] <- z[i]*p
    y[i] ~ dbin(mu[i],K)
 }

N<-sum(z[1:(nind+nz)])
}
```

PANEL 4.1

BUGS description of model M_0 under data augmentation. Here `y`, `K`, `nind`, and `nz` are provided as data. The population size, N, is computed as a function of the data augmentation variables z.

our standard notation a bit. In this case we use `nind` for n (the number of encountered individuals), and `M = nind + nz` is the total size of the augmented data set. In other cases we might also use `nocc` in place of K and `ntraps` in place of J. We find that word definitions make code easier to understand, especially without having to read surrounding text.

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

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

In this manner, it is straightforward to incorporate covariates on p for both individuals and sampling occasions (see discussion of this below and Chapter 7) as well as to devise other extensions of the model, including models for open populations (Chapter 16).

4.2.3 Remarks on data augmentation

Data augmentation may seem like a strange and mysterious black-box, and likely it is unfamiliar to most people, even to many of those with substantial experience with capture-recapture models. However, it really is just a formal reparameterization of capture-recapture models in which N is marginalized out of the ordinary (conditional-on-N) model (by summation over a binomial prior). As a result, we could refer to the resulting model as the "binomial-integrated likelihood" to reflect that an estimator could be obtained from the ordinary likelihood, integrated over a binomial prior. Other such "integrated likelihood" models are sensible. For example, we could place a Poisson prior on N with mean Λ and marginalize N over the Poisson prior. This produces a likelihood in which Λ replaces N, instead of ψ replacing N. We note that this type of marginalization (over a Poisson prior) is done by the **R** package `secr` for analysis of spatial capture-recapture models (see Section 6.5.3).

We emphasize the motivation for data augmentation being that it produces a data set of fixed size, so that the parameter dimension in any capture-recapture model is also fixed. As a result, MCMC is a relatively simple proposition using standard Gibbs sampling. And, in particular, capture-recapture models become trivial to implement in **BUGS**. Consider the simplest context—analyzing model M_0. In this case, DA converts model M_0 to a basic occupancy model, and the parameters p and ψ have known full conditional distributions (in fact, beta distributions) that can be sampled from directly. Furthermore, the data augmentation variables, i.e., the collection of z's, can be sampled from Bernoulli full conditionals. MCMC is not much more difficult for complicated models—sometimes the hyperparameters need to be sampled using a Metropolis-Hastings step (e.g., Chapter 17), but nothing more sophisticated than that is required.

Potential sensitivity of parameter estimates to M ~~(especially of N)~~ might be cause for some concern. The guiding principle is that it should be chosen large enough so that the posterior for N is not truncated, but it should not be too large due to the increased computational burden. It seems likely that the properties of the Markov chains should be affected by M and so some optimal choice of M might exist (Gopalaswamy, 2012). Formal analysis of this is needed.

There are other approaches to analyzing models with unknown N, using reversible jump MCMC (RJMCMC) or other so-called "trans-dimensional" (TD) algorithms (King and Brooks, 2001; Durban and Elston, 2005; King et al., 2008; Schofield and Barker, 2008; Wright et al., 2009). 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 dimension of the parameter space is a function of N, and will therefore typically vary at each iteration of the MCMC algorithm. 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, 2012;

Royle et al., 2012b; Royle and Converse, in review), which we consider in Chapter 14. Furthermore, data augmentation is often easier to implement than RJMCMC, and the details of the DA implementation are the same for all capture-recapture problems.

4.2.4 Example: Black bear study on Fort Drum

To illustrate the analysis of model M_0 using data augmentation, we use a data set collected at Fort Drum Military Installation in upstate New York. The data were collected by P.D. Curtis and M.T Wegan of Cornell University and their colleagues at the Fort Drum Military Installation. These data have been analyzed in various forms by Wegan (2008), Gardner et al. (2009) and Gardner et al. (2010b). The specific data used here are encounter histories on 47 individuals obtained from an array of 38 baited "hair snares" (Figure 4.1) during June and July 2006. Barbed wire traps were baited and checked for hair samples each week for 8 weeks, thus we distinguished $K = 8$ weekly sample occasions. The data are provided in the **R** package `scrbook`, can be loaded by typing `data(beardata)` at the **R** prompt, and the analysis can be set up and run as follows (also see `?beardata` for the commands to do the analysis). Here, the data were augmented with 128 all-zero encounter histories, resulting in a total data set size of $M = 175$.

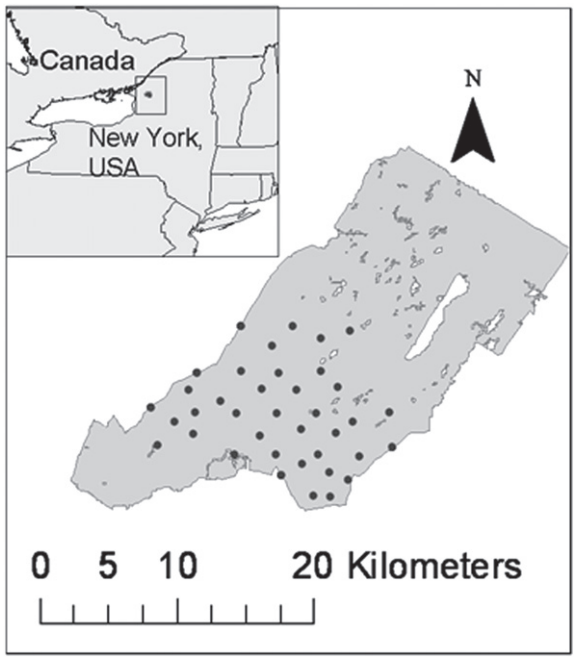

FIGURE 4.1

Fort Drum Black bear study area and the 38 baited hair snare locations operated for 8 weeks during June and July, 2006.

```
> library(scrbook)
> data(beardata)          # load the bear data and extract components
> trapmat <- beardata$trapmat
> nind <- dim(beardata$bearArray)[1]
> K <- dim(beardata$bearArray)[3]
> ntraps <- dim(beardata$bearArray)[2]
> M <- 175
> nz <- M-nind
> Yaug <- array(0, dim = c(M,ntraps,K))

> Yaug[1:nind,,] <- beardata$bearArray
> y <- apply(Yaug,c(1,3),sum)   # summarize by ind x rep
> y[y > 1] <- 1                 # toss out multiple encounters per occasion
                                # b/c traditional CR models ignore space
```

The raw data object, `beardata$bearArray`, is a three-dimensional array `nind` $\times$ `ntraps` $\times$ K of individual encounter events (i.e., $y_{ijk} = 1$ if individual i was encountered in trap j during occasion k, and 0 otherwise). For fitting model M_0 (or M_h, see Section 4.4), it is sufficient to reduce the data to individual encounter frequencies which we have re-labeled "`y`" above. The **BUGS** model file along with commands to fit the model are as follows:

```
> set.seed(2013)                  # to obtain the same results each time
> library(R2WinBUGS)              # load R2WinBUGS, set-up:
> data0 <- list(y = y, M = M, K = K)       # data ....
> params0 <- c('psi','p','N')                 # parameters ....
> zst <- c(rep(1,nind),rbinom(M-nind, 1, .5))     # inits .....
> inits <- function(){   list(z = zst, psi = runif(1), p = runif(1)) }

> cat("
model{

psi ~ dunif(0, 1)
p ~ dunif(0,1)

for (i in 1:M){
   z[i] ~ dbern(psi)
   for(k in 1:K){
     tmp[i,k] <- p*z[i]
     y[i,k] ~ dbin(tmp[i,k],1)
      }
    }
 N<-sum(z[1:M])
}
",file="modelM0.txt")

## Run the model:
> fit0 <- bugs(data0, inits, params0, model.file="modelM0.txt",n.chains = 3,
       n.iter = 2000, n.burnin = 1000, n.thin = 1,debug = TRUE,working.directory = getwd())
```

This produces the following posterior summary statistics:

```
> print(fit0,digits = 2)
Inference for Bugs model at "modelM0.txt", fit using WinBUGS,
 3 chains, each with 2000 iterations(first 1000 discarded)
 n.sims  = 3000 iterations saved
```

```
              mean      sd     2.5%      25%      50%      75%    97.75%  Rhat   n.eff
psi           0.29    0.04     0.22     0.26     0.29     0.31     0.36     1    3000
p             0.30    0.03     0.25     0.28     0.30     0.32     0.35     1    3000
N            49.94    1.99    47.00    48.00    50.00    51.00    54.00     1    3000
deviance    489.05   11.28   471.00   480.45   488.80   495.40   513.70     1    3000

[... some output deleted ...]
```

WinBUGS did well in choosing an MCMC algorithm for this model—we have $\hat{R} = 1$ for all parameters and an effective sample size of 3,000, equal to the total number of posterior samples. We see that the posterior mean of N under this model is 49.94 and a 95% posterior interval is 48, 54. We revisit these data later in the context of more complex models.

In order to obtain an estimate of density, D, we need an area to associate with the estimate of N, and in Chapter 1 we already went through a number of commonly used procedures to conjure up such an area, including buffering the trap array by the home range radius, often estimated by the mean maximum distance moved (MMDM) (Parmenter et al., 2003), 1/2 MMDM (Dice, 1938) or directly from telemetry data (Wallace et al., 2003). Typically, the trap array is defined by the convex hull around the trap locations, and this is what we applied a buffer to. We computed the buffer by using a telemetry-based estimate of the mean female home range radius (2.19 km) (Bales et al., 2005) instead of using an estimate based on our relatively sparse recapture data. For the Fort Drum study, the convex hull has an area of 157.135 km^2, and the buffered convex hull has an area of 277.011 km^2. To create this we used functions contained in the **R** package `rgeos` (Bivand and Rundel, 2011) and created a utility function `bcharea`, which is in our **R** package `scrbook`. The commands are as follows:

```
> library(rgeos)

> bcharea <- function(buff,traplocs){
    p1 <- Polygon(rbind(traplocs,traplocs[1,]))
    p2 <- Polygons(list(p1 = p1),ID = 1)
    p3 <- SpatialPolygons(list(p2 = p2))
    p1ch <- gConvexHull(p3)
    bp1 <- (gBuffer(p1ch, width = buff))
    plot(bp1, col='gray')
    plot(p1ch, border='black', lwd = 2, add = TRUE)
    gArea(bp1)
  }

> bcharea(2.19,traplocs = trapmat)
```

The resulting buffered convex hull is shown in Figure 4.2.

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

```
> summary(fit0$sims.list$N/277.011)

   Min.   1st Qu.   Median     Mean   3rd Qu.     Max.
 0.1697    0.1733   0.1805   0.1803    0.1841   0.2130

> quantile (fit0$sims.list$N/277.011,c(0.025,0.975))

     2.5%       97.5%
0.1696684   0.1949381
```

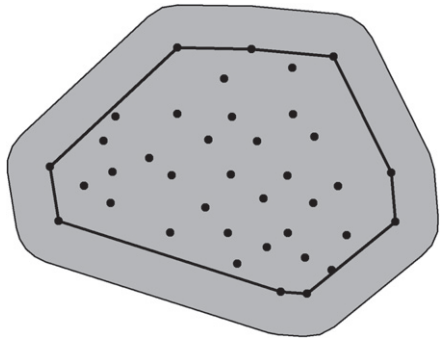

FIGURE 4.2

Convex hull of the bear hair snare array at Fort Drum, NY, buffered by mean female home range radius (2.19 km).

which yields a density estimate of about 0.18 ind/km^2, and a 95% Bayesian confidence interval of (0.170, 0.195). Our estimate of density should be reliable if we have faith in our stated value of the "sampled area." Clearly, though, this is largely subjective, and not something we can formally evaluate (or estimate) from the data based on model M_0.

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 occasions. Sometimes we can't aggregate the encounter data for each individual, such as when encounter probability varies over occasions. Time-varying responses that are relevant in many capture-recapture studies are "effort" such as amount of search time, number of observers, or trap nights, or encounter probability varying over time, as a function of date or season (Kéry et al., 2011) due to species behavior. A common situation in many animal studies 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 animal 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 (a Markovian or ephemeral behavioral response; (Yang and Chao, 2005)), described by the term $\alpha_1 y_{i,k-1}$ or in *any* previous period (persistent behavioral response), described by the term $\alpha_2 x_{ik}$. Because spatial capture-recapture models allow us to include trap-specific covariates, we can describe a third type of behavioral response—a local behavioral response that is trap-specific (Royle et al., 2011b). In this local behavioral response, the encounter probability is modified for an individual trap depending on previous capture in that trap. Models with temporal effects are easy to describe and analyze in the **BUGS** language and we provide a number of examples in Chapter 7 and elsewhere.

4.4 Models with individual heterogeneity

Models in which encounter probability varies by individual have a long history in capture-recapture and, indeed, this so-called "model M_h" ("h" for heterogeneity) is one of the elemental capture-recapture models (Otis et al., 1978). Conceptually, we imagine that the individual-specific encounter probability parameters, p_i, are random variables distributed according to some probability distribution, $[p|\theta]$. We denote this basic model assumption as $p_i \sim [p|\theta]$. 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 non-parametric approaches (Burnham and Overton, 1978; Norris and Pollock, 1996; Pledger, 2004). 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 different heterogeneity 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 (Section 4.5 below). Indeed, spatial capture-recapture models 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, Chapter 6) and Kéry and Schaub (2012, Chapter 6).

We will work with a specific type of model M_h here, which is a natural extension of the basic binomial observation model of model M_0 so that

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

where μ is a fixed parameter (the mean) to be estimated, and η_i is an individual random effect assumed to be normally distributed:

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

We could as well combine these two steps and write $\text{logit}(p_i) \sim \text{Normal}(\mu, \sigma_p^2)$. This "logit-normal mixture" was analyzed by Coull and Agresti (1999) and ~~elsewhere. It is a natural extension of the basic model with constant p, as a GLMM, and similar models occur throughout statistics.~~ It is also natural to consider a beta distribution for p_i (Dorazio and Royle, 2003), and so-called "finite-mixture" models are also popular (Norris and Pollock, 1996; Pledger, 2004). In the latter, individuals are assumed to belong to a finite number of latent classes, each of which has its own capture probability.

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 effective sample area, 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 (e.g., Royle and Young, 2008). Further, suppose that the species exhibits some sense of spatial fidelity in the form of a home range or territory, and individuals move about their home range in some kind of random fashion. Figure 4.3 shows a sample quadrat searched repeatedly over a period of time (home range centroids are given by the solid dots). 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

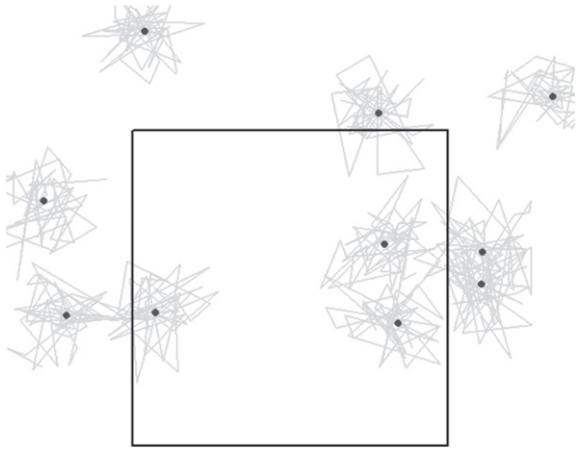

FIGURE 4.3

A quadrat searched for lizards over some period of time. The locations (simulated data) for each of 10 lizards are connected by lines—the dots are the activity centers.

the long-run proportion of times the individual is within the sample plot boundaries, say ϕ. We might model the exposure or *availability* of an individual to capture by supposing that $a_i = 1$ if individual i is available to be captured (i.e., within the survey plot) during any sample, and 0 otherwise. Then, $\Pr(a_i = 1) = \phi$. In the context of spatial studies, it is natural that ϕ should depend on *where* an individual lives, i.e., it should be individual-specific, ϕ_i (Chandler et al., 2011). This system describes, precisely, that of "random temporary emigration" (Kendall et al., 1997), where ϕ_i is the individual-specific probability of being available for capture.

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

$$p_i = p_0 \phi_i$$

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

4.4.1 Analysis of model M_h

If N is known, it is worth taking note of the essential simplicity of model M_h as a binomial GLMM. This type of model is widely applied throughout statistics using standard methods of inference based either on integrated likelihood (Laird and Ware, 1982; Berger et al., 1999), which we discuss in Chapter 6, or standard Bayesian methods. However, because N is not known, inference is somewhat more challenging. We address that here using Bayesian analysis based on data augmentation. 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). 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: $z_i \sim \text{Bernoulli}(\psi)$ and the model of the observations expressed conditional on these latent variables z_i. For $z_i = 1$, we have a binomial model with individual-specific p_i:

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

and otherwise $y_i | z_i = 0 \sim I(y = 0)$, i.e., a point mass at $y = 0$. Further, we prescribe a distribution for p_i. Here we assume

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

For prior distributions we assume $p_0 = \text{logit}^{-1}(\mu) \sim \text{Uniform}(0, 1)$ and, for the standard deviation $\sigma \sim \text{Uniform}(0, B)$ for some large B. Another common default prior is to assume $\tau = 1/\sigma^2 \sim \text{Gamma}(.1, .1)$.

4.4.2 Analysis of the Fort Drum data with model M_h

Here we provide an analysis of the Fort Drum bear survey data using the logit-normal heterogeneity model, and we used data augmentation to produce a data set of $M = 700$ individuals. We have so far mostly used **WinBUGS** but we are now transitioning to **JAGS** run from within **R** using the packages `R2jags` (Su and Yajima, 2011) or `rjags` (Plummer, 2009). The function `jags` from the `R2jags` package runs essentially like the `bugs` function which we demonstrate here for setting up and running model M_h for the Fort Drum bear data:

```
[... get data as before ....]

> set.seed(2013)

> cat("
model{
 p0 ~ dunif(0,1)           # prior distributions
 mup <- log(p0/(1-p0))
 sigmap ~ dunif(0,10)
 taup <- 1/(sigmap*sigmap)
 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],K)    # observation model
 }

 N<-sum(z[1:(nind + nz)])
}
",file="ModelMh.txt")

> data1 <- list(y = y, nz = nz, nind = nind, K = K)
> params1 <- c('p0','sigmap','psi','N')
> inits <- function(){ list(z = as.numeric(y>=1), psi=.6, p0 = runif(1),
          sigmap = runif(1,.7,1.2),lp = rnorm(M,-2)) }
> library(R2jags)
> wbout <- jags(data1, inits, params1, model.file = "modelMh.txt", n.chains = 3,
            n.iter = 1010000, n.burnin = 10000, working.directory = getwd())
```

We provide an **R** function `modelMhBUGS` in the package `scrbook` that will fit the model using either **JAGS** or **WinBUGS** as specified by the user. In addition, for fun, we construct our own MCMC algorithm using a Metropolis-within-Gibbs algorithm for model M_h in Chapter 17, where we also develop MCMC algorithms for spatial capture-recapture models. Using `modelMhBUGS`, we ran 3 chains of 1 *million* iterations (mixing is poor for this model and this data set), which produced the posterior distribution for N shown in Figure 4.4. Posterior summaries of parameters are given in Table 4.5.

We used $M = 700$ for this analysis and we note that while the posterior mass of N is concentrated away from this upper bound (Figure 4.4), the posterior has an extremely long right tail, with some MCMC draws at the upper boundary $N = 700$,

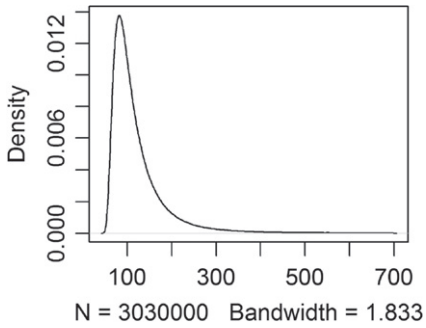

FIGURE 4.4

Posterior of N for Fort Drum black bear study data under the logit-normal version of model M_h.

Table 4.5 Posterior summaries from model M_h fitted to the Fort Drum black bear data. Results were obtained using **JAGS** running 3 chains, each with 10,10,000 iterations, discarding the first 10,000 for a total of three *million* posterior samples.

Parameter	Mean	SD	2.4%	50%	97.5%	Rhat	n.eff
p_0	0.072	0.056	0.002	0.060	0.203	1.008	540
σ_p	2.096	0.557	1.215	2.025	3.373	1.003	820
ψ	0.176	0.101	0.084	0.147	0.458	1.006	650
N	122.695	69.897	62.000	102.000	319.000	1.006	630

suggesting that an even higher value of M may be called for. To characterize the posterior distribution of density, we produce the relevant summaries of the posterior distribution of $D = N/277.11$ (recall the buffered area of the convex hull is 277.11 km^2):

```
> summary(wbout$sims.list$N/277.11)

      Min.   1st Qu.   Median     Mean   3rd Qu.     Max.
    0.1696    0.2959   0.3681   0.4428    0.4944   2.5260

> quantile(wbout$sims.list$N/277.11,c(0.025,0.50,0.975))

       2.5%         50%       97.5%
  0.2237379   0.3680849   1.1511674
```

Therefore, the point estimate, characterized by the posterior median, is around 0.37 bears/km^2 and a 95% Bayesian credible interval is (0.224, 1.151).

4.4.3 Comparison with MLE

The posterior of N is highly skewed; therefore, we see that the posterior mean ($N = 122.7$) is considerably higher than the posterior median ($N = 102$). Further, it may be surprising that these posterior summaries do not compare well with the MLE. To obtain the MLE of $\log(n_0)$, the logarithm of the number of uncaptured individuals, we used the **R** code contained in Panel 6.1 from Royle and Dorazio (2008). We found $\widehat{\log(n0)} = 3.86$ and therefore $\hat{N} = \exp(3.86) + 47 = 94.47$, which is larger than the mode shown in Figure 4.4. To see this, we compute the posterior mode, by finding the posterior value of N with the highest mass. Because N is discrete, we can use the `table()` function in **R** and find the most frequent value.[3] If we want to smooth out some of the Monte Carlo error a bit, we can use a smoother of some sort applied to the tabled posterior frequencies of N. Here we use a smoothing spline (**R** function `smooth.spline`) with the degree of smoothing chosen by cross-validation (the `cv = TRUE` argument):

```
> N <- table(jout$BUGSoutput$sims.list$N)
> xg <- as.numeric(names(N))

> sp <- smooth.spline(xg,N,cv = TRUE)

> sp

Call:
smooth.spline (x  = xg, y  = N, cv  = TRUE)

Smoothing Parameter spar = 0.09339815 lambda = 8.201724e-09 (17 iterations)
Equivalent Degrees of Freedom (Df): 121.1825
Penalized Criterion: 2544481
PRESS: 5903.4
```

We obtain the mode of the smoothed frequencies as follows:

```
sp$x[sp$y==max(sp$y)]
[1] 82
```

We don't dwell too much on the difference between the MLE and features of the posterior, but we do note here that the posterior distribution for the parameters of this model, for the Fort Drum data set, is very sensitive to the prior distributions. In the present case, the use of a Uniform(0, 1) prior for $p_0 = \text{logit}^{-1}(\mu)$ is somewhat informative—in particular, it is not at all "flat" on the scale of μ, and this affects the posterior. We generally recommend use of a Uniform(0, 1) prior for $\text{logit}^{-1}(\mu)$ in such models. That said, we were surprised at this result, and we experimented with other prior configurations including putting a flat prior on μ directly. This kind of small sample instability has been widely noted in model M_h (Fienberg et al., 1999; Dorazio and Royle, 2003), as has extreme sensitivity to the specific form of model M_h (Link, 2003). In summary, while the mode is well defined, the data set is relatively sparse and hence inferences are poor and sensitive to model choice.

[3] For a continuous random variable we can use the function `density()` to smooth the posterior samples and obtain the mode.

4.5 Individual covariate models: toward spatial capture-recapture

A standard situation in capture-recapture is when a covariate that is thought to influence encounter probability is measured for each individual. These are often called "individual covariate models" but, in keeping with the classical nomenclature on closed population models, Kéry and Schaub (2012) referred to this class of models as "model M_x" (the x here being an explicit covariate). As with other closed population models, we begin with the basic binomial observation model:

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

To model the covariate, we use a logit model for encounter probability of the form:

$$\text{logit}(p_i) = \alpha_0 + \alpha_1 x_i, \tag{4.5.1}$$

where x_i is the covariate value for individual i and the parameters $\boldsymbol{\alpha} = (\alpha_0, \alpha_1)$ are the regression coefficients. Classical examples of covariates influencing detection probability are type of animal (juvenile/adult or male/female), a continuous covariate such as body mass, or a discrete covariate such as group or cluster size. For example, in models of aerial survey data, it is natural to model the detection probability of a group as a function of the observation-level individual covariate, "group size" (Royle, 2008; Langtimm et al., 2011).

Model M_x is 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 can be described precisely as a form of model M_x, where the covariate describes *where* the individual is located in relation to the trapping array. Specifically, SCR models *are* individual covariate models, but where the individual covariate is only observed imperfectly (or partially observed) for each captured individual. Unlike model M_h, in SCR models (and model M_x) we do have some direct information about the latent variable, which comes from the spatial locations/distribution of individual recaptures.

Traditionally, estimation of N in model M_x is achieved using methods based on ideas of unequal probability sampling (i.e., Horvitz-Thompson estimation[4]; Huggins (1989); Alho (1990); Borchers et al. (2002)). An estimator of N is

$$\hat{N} = \sum_{i=1}^{n} \frac{1}{\tilde{p}_i},$$

where $\tilde{p}_i$ is the probability that individual i appeared in the sample. This quantity is $\tilde{p}_i = \Pr(y_i > 0)$ and, in closed population capture-recapture models, it can be computed as:

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

[4]For a quick summary of the idea see: http://en.wikipedia.org/wiki/Horvitz-Thompson_estimator.

where p_i is a function of parameters α_0 and α_1 according to Eq. (4.5.1). In practice, parameters are estimated from the conditional likelihood of the observed encounter histories which is, for observation y_i,

$$\mathcal{L}_c(\boldsymbol{\alpha}|y_i) = \frac{\text{Binomial}(y_i|\boldsymbol{\alpha})}{\tilde{p}_i}. \tag{4.5.2}$$

This derives from a straightforward application of the law of total probability. Conceptually, we partition $\Pr(y)$ according to $\Pr(y) = \Pr(y|y > 0)\Pr(y > 0) + \Pr(y|y = 0)\Pr(y = 0)$. For any positive value of y the second term is necessarily 0, and so we rearrange to obtain $\Pr(y|y > 0) = \Pr(y)/\Pr(y > 0)$ which, in the specific case where $\Pr(y)$ is the binomial probability mass function (pmf), produces Eq. (4.5.2).

Here, we take a formal model-based approach to Bayesian analysis of such models based on the joint likelihood using data augmentation (Royle, 2009b). Classical likelihood analysis of the so-called "full likelihood" is covered by Borchers et al. (2002). For Bayesian analysis of model M_x, because the individual covariate is unobserved for the $n_0 = N - n$ uncaptured individuals, we require a model to describe variation in x among individuals, essentially allowing the sample to be extrapolated to the population. For example, if we have a continuous trait measured on each individual, then we might assume that x 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 following form, for each $i = 1, 2, \ldots, M$:

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

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

Here we consider a special type of model M_x that is particularly relevant to spatial capture-recapture. Intuitively, some measure of distance from home range center to traps for an individual should be a reasonable covariate to explain heterogeneity in encounter probability, i.e., individuals with more exposure to traps should have higher encounter probabilities and vice versa. So we can imagine *estimating* such a quantity, say average distance from home range center to "the trap array", and then using it as an individual covariate in capture-recapture models. 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 from home range center to edge of the trapping array", where the home range center is estimated by the average capture location. This is intuitively appealing because we can imagine, in some kind of an ideal situation where we have a dense grid of traps over some geographic region, that the average location of capture would be a decent estimate (heuristically) of an individual's home range center. We provide an example of this type of approach using a fully model-based analysis of the version of model M_x 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 will resolve the problem of having an ill-defined sample area. 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 the scalar individual covariate x_i to be the distance from the average encounter location of individual i, say $\mathbf{s}_i$, to the centroid of the trap array, $\mathbf{x}_0$: $x_i = ||\mathbf{s}_i - \mathbf{x}_0||$. Note that $||\mathbf{u}||$ is standard notation for Euclidean norm or magnitude of the vector $\mathbf{u}$, and we use it throughout the book. In practice, people have used distance from edge of the trap array but that is less easy to quantify, as "edge" itself is not precisely defined. Conceptually, individuals in the middle of the array should have a higher probability of encounter and, as x_i increases, p_i should therefore decrease. Note that we have defined $\mathbf{s}_i$ in terms of a sample quantity—the observed mean encounter location—which, while ad hoc, is consistent with the use of individual covariate models in the literature. For an expansive, dense trapping grid 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.

A key point is that $\mathbf{s}_i$ is missing for each individual that is not encountered and consequently x_i is also missing. Therefore, it is a latent variable, and we need to specify a probability distribution for it. As a measurement of distance we know it must be positive-valued, and it seems sensible that an individual located extremely far from the array of traps would not be captured. Therefore, let's assume that x_i is uniformly distributed from 0 to some large number, say B, beyond which it would be difficult to imagine an individual being captured by the trap array:

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

where B is a specified constant, which we may choose to be arbitrarily large. For example, B should be at least a home range diameter past the furthest trap from the centroid of the array.

4.5.2 Example: Fort Drum black bear study

We have to do a little bit of data processing to fit this individual covariate model to the Fort Drum data. We need to compute the individual covariate x_i (distance from the centroid of the trapping array) using the **R** function `spiderplot` provided in

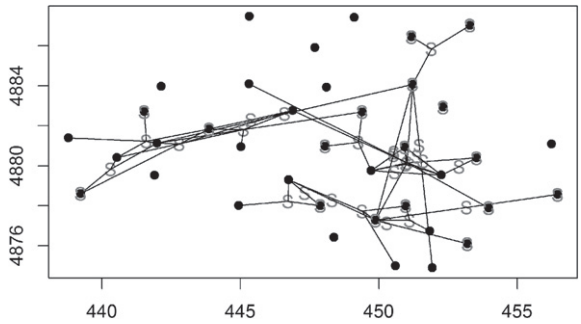

FIGURE 4.5

Spider plot of the Fort Drum black bear study data. The black dots represent the 47 trap locations with the "S" symbols being the average capture location of each bear, i.e., its estimated home range center. All traps in which a bear was captured are connected to its estimated home range center with a line.

`scrbook`. This function also produces the keen plot shown in Figure 4.5, which we call a "spider plot." The **R** commands for obtaining the individual covariate "distance from trap centroid" (the variable `xcent` returned by `spiderplot`) and making the spider plot are as follows:

```
> library(scrbook)
> data(beardata)
> toad<-spiderplot(beardata$bearArray,beardata$trapmat)
> xcent<-toad$xcent
```

For the analysis of these data using the individual covariate "distance from centroid" we used $x_i \sim \text{Uniform}(0, B)$ with $B = 11.5$ km, which is about the distance from the array center to the furthest trap. Once we choose a value for B, the direct implication is that the population size parameter, N, applies to the area within 11.5 kms of the trap centroid. Therefore, the model associates a precise area within which the population of N individuals resides. We will see shortly that N does, in fact, scale with our choice of B to reflect the changing area over which the N individuals of the model reside. The **BUGS** model specification and **R** commands to package the data and fit the model are as follows:

```
cat("
model{
   p0 ~ dunif(0,1)              # prior distributions
   alpha0 <- log(p0/(1-p0))
   psi ~ dunif(0,1)
   beta ~ dnorm(0,.01)

>for(i in 1:(nind + nz)){
   xcent[i] ~ dunif(0,B)
   z[i] ~ dbern(psi)            # DA variables
```

Table 4.6 Posterior summaries from the individual covariate model (model M_x) with covariate "distance from the centroid of the trap array," fitted to the Fort Drum black bear data. Results were obtained using **WinBUGS** running 3 chains, each with 11,000 iterations, discarding the first 1,000 for a total of 30,000 posterior samples.

Parameter	Mean	SD	2.4%	50%	97.5%	Rhat	n.eff
p_0	0.54	0.07	0.40	0.54	0.67	1	1100
ψ	0.34	0.05	0.25	0.34	0.44	1	3500
N	58.92	5.49	50.00	58.00	71.00	1	1900
β	−0.25	0.06	−0.36	−0.25	−0.12	1	780

```
   lp[i] <- alpha0  + beta*xcent[i]   # individual effect
   logit(p[i]) <- lp[i]
   mu[i] <- z[i]*p[i]
   y[i] ~ dbin(mu[i],K)            # observation model
  }

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

>data2 <- list(y = y,nz = nz, nind = nind, K = K, xcent = xcent,B = 11.5)
>params2 <- c('p0','psi','N','beta')
>inits <- function() {list(z = zst, psi = psi, p0 = runif(1), beta = rnorm(1) ) }
>fit2 <- bugs(data2, inits, params2, model.file="modelMcov.txt",
             n.chains = 3, n.iter = 11000, n.burnin = 1000, n.thin = 1)
```

This produces the posterior summary statistics in Table 4.6. We note that the estimated N is much lower than obtained by model M_h but there is a good explanation for this which we discuss in the next section. That issue notwithstanding, it is worth pondering how this model could be an improvement (conceptually or technically) over some other model/estimator including M_0 and M_h considered previously. Well, for one, we have accounted formally for heterogeneity due to spatial location of individuals relative to exposure to the trap array, characterized by the centroid of the array. Moreover, we have done so using a model that is based on an explicit mechanism, as opposed to a phenomenological one such as model M_h. In addition, and importantly, using our new model, *the estimated N applies to an explicit area which is defined by our prescribed value of B*. 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 located less than B from the centroid of the trap array. As such, the implied "effective area" of the trap array for a given B is a precisely defined quantity—the area of a circle with radius B.

4.5.3 Extension of the model

The model developed in the previous section is not a very good model for one important reason: imposing a uniform prior distribution on x implies that density is *not*

constant over space. In particular, this model implies that density *decreases* as we move away from the centroid of the trap array. That is, $x_i \sim \text{Uniform}(0, B)$ implies constant N in each distance band from the centroid but obviously the *area* of each distance band is increasing. This is one reason we have a lower estimate of density than that obtained previously from model M_h (Section 4.4.2) and also why, if we were to increase B, we would see density continue to decrease.

Fortunately, we are not restricted to use this specific distribution for the individual covariate. Clearly, it is a bad choice and, therefore, we should think about whether we can choose a better distribution for x—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 abundance should be proportional to the amount of area in each successive distance band as you move farther away from the centroid, so that density is *constant*. In fact, theory exists that tells us we should choose $[x] = 2x/B^2$. This can be derived by noting that $F(x) = \Pr(X < x) = (\pi x^2)/(\pi B^2)$. Then, $f(x) = dF/dx = 2x/(B^2)$. This is a sort of triangular distribution in density induced because the incremental area in each additional distance band increases linearly with radius (i.e., distance from centroid). This can be verified empirically as follows:

```
> 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 directly in **BUGS** but there is not a built-in way to do so. However, we can implement a discrete version of the pdf.[5] To do this, we break the interval $[0, B]$ into L distance classes of width δ, with probabilities proportional to $2x$. In particular, if we denote the cut-points by $g_1 = 0, g_2, \ldots, g_{L+1} = B$ and the interval midpoints are $m_i = g_{i+1} - \delta$. Then the interval probabilities are, approximately,[6] $p_i = \delta(2m_i/B^2)$, which we can compute once and then pass to **BUGS** as data. The **R** commands for doing all of this (noting that we have already loaded and processed the Fort Drum bear data) are given in the following **R/BUGS** script:

```
> delta <-.2
> xbin <- xcent%/%delta  + 1        # Put x in bins
> midpts <- seq (delta,Dmax,delta)
> xprobs <- delta*(2*midpts/(B*B))
> xprobs <- xprobs/sum (xprobs)

> cat ("
```

[5] We might also be able to use what is referred to in **WinBUGS** jargon as the "zeros trick" (see *Advanced BUGS tricks* in the manual) although we haven't pursued this approach.

[6] This is just length × width, the area of small rectangles approximating the integral.

```
model{
p0 ~ dunif (0,1)                              # Prior distributions
alpha0 <- log (p0/(1-p0))
psi ~ dunif (0,1)
beta ~ dnorm (0,.01)

for (i in 1: (nind + nz)){
  xbin[i] ~ dcat (xprobs[])
  z[i] ~ dbern (psi)                          # DA variables
  lp[i] <- alpha0  + beta*xbin[i]*delta
  logit (p[i]) <- lp[i]
  mu[i] <- z[i]*p[i]
  y[i] ~ dbin (mu[i],K)                       # Observation model
 }

N <- sum (z[1: (nind + nz)])                  # N is derived
}
",file="modelMcov.txt")
```

In the model description, the variable x (observed distance from centroid of the trap array) has been rounded or binned (placed into a distance bin) so that the discrete version of the pdf of x can be used, as described previously. The new variable labeled `xbin` is then the *integer category* in units of δ from 0. Thus, to convert back to distance in the expression for `lp[i]`, `xbin[i]` has to be multiplied by δ. To fit the model, keeping in mind that the data objects required below have been defined in previous analyses of this chapter, we do this:

```
> data2 <- list (y = y, nz = nz, nind = nind, K = K, xbin = xbin, xprobs = xprobs,
                 delta = delta)
> params2 <- c('p0','psi','N','beta')
> inits <- function() { list (z = z, psi = psi, p0 = runif (1), beta = rnorm(1) ) }
> fit <- bugs (data2, inits, params2, model.file="modelMcov.txt",
          working.directory = getwd (), debug = FALSE, n.chains = 3,
          n.iter = 11000, n.burnin = 1000, n.thin = 2)
```

By specification of B, this model induces a clear definition of area in which the population of N individuals resides. The parameter N of the model is the population size that applies to the particular value of B and, as such, we will see that N scales with our choice of B. This might be disconcerting to some—we can get whatever value of N we want by changing B! However, it is intuitively reasonable that, as we increase the area under consideration, there should be more individuals in it. Fortunately, we find empirically, that while N is highly sensitive to the prescribed value of B, density appears invariant to B as long as B is sufficiently large. We fit the model for a set of values of B from $B = 12$ (restricting values of x to be in close proximity to the trap array) on upto 20. The results are given in Table 4.7.

We see that the posterior mean and SD of density (individuals/km^2) appear insensitive to choice of B once we reach about $B = 17$ or so. The estimated density of $0.25/\text{km}^2$ is actually quite a bit lower than we reported using model M_h for which no relevant "area" quantity is explicit in the model (and so we had to make it up). Using MLEs of N in conjunction with buffer strips (see Table 1.1) our estimates were in

Table 4.7 ~~Analysis~~ of Fort Drum bear hair snare data using the individual covariate model, for different values of B, the upper limit of the uniform distribution of 'distance from centroid of the trap array'.

B	Density (post. mean)	Posterior SD
12	0.230	0.038
15	0.244	0.041
17	0.249	0.044
18	0.249	0.043
19	0.250	0.043
20	0.250	0.044

the range of 0.32–0.43. On the other hand, our estimate of $\hat{D} = 0.25$ here (based on the posterior mean) is higher than that reported from model M_0 using the buffered area ($\hat{D} = 0.18$). There is no basis really for comparing or contrasting these various estimates. In particular, application of models M_0 and M_h are distinctly *not* spatially explicit models—the area within which the population resides is not defined under either model. There is therefore no reason at all to think that the estimates produced under either closed population model, based on a buffered "trap area," are justifiable by 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 uses an explicit model for "distance from centroid" which 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 area, which seems like a desirable feature.

4.5.4 Invariance of density to *B*

Under model M_x, and under models that we consider in later chapters, a general property of the estimators is that while N increases with the prescribed area of the model (defined by B in this model), we expect that density estimators should be invariant to this area. In the model used above, we note that $\text{Area}(B) = \pi B^2$ and $\mathbb{E}(\text{N}(B)) = \lambda \text{Area}(B)$ and thus $\mathbb{E}(\text{Density}(B)) = \lambda$, i.e., constant. This should be interpreted as the *prior* density. Absent data, realizations under the model will have density, λ, regardless of what B is prescribed to be. As we verified empirically above, posterior summaries of density are also invariant to B as long as the prescribed area is sufficiently large.

4.5.5 Toward fully spatial capture-recapture models

While the use of an individual covariate model resolves two important problems inherent in almost all capture-recapture studies (induced heterogeneity and absence

of a precise relationship between N and area), it is not ideal for all purposes 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, so that we cannot use this model to model trap-specific effects (e.g., trap effort or type). Moreover, we applied this model to "data" being the average observed encounter location, and equated that summary to the home range center $\mathbf{s}_i$. 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, which might not be reasonable in practice. Additionally, this approach also ignored the variable precision with which each $\mathbf{s}_i$ is estimated. Finally, it ignores that estimates of $\mathbf{s}_i$ around the "edge" (however we define that) are biased because the observations are truncated—we can only observe locations interior to the array.

However, there is hope to extend this model in order to resolve these remaining deficiencies. In the next chapter we provide a further extension of this individual covariate model that definitively resolves the *ad hoc* nature of the approach we took here. In that chapter we build a model in which $\mathbf{s}_i$ 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 $\mathbf{s}_i$ as a latent variable, simply by adding a prior distribution for $\mathbf{s}_i$. This is actually easier, and less *ad hoc* in a number of respects, and you should try it out.

4.6 Distance sampling: a primitive SCR model

Distance sampling is a class of methods for estimating animal density from measurements of distance from an observer to individual animals (or groups). The basic assumption is that detection probability is a function of distance. Distance sampling is one of the most popular methods for estimating animal abundance (Burnham et al., 1980; Buckland et al., 2001; Buckland, 2004) because, unlike ordinary closed population models, distance sampling provides explicit estimates of *density*. In terms of methodological context, the distance sampling model is a special case of a closed population model with an individual covariate. The covariate in this case, x, is the distance between an individual's location say $\mathbf{u}$ and the observation location or transect. In fact, distance sampling is precisely an individual-covariate model, except that observations are made at only $K = 1$ sampling occasion. Distance sampling eliminates the need to explicitly identify individuals (except they need to be *distinguished* from other individuals) repeatedly and so distance sampling can be applied to unmarked populations. 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 (Hedley et al., 1999; Royle et al., 2004; Johnson, 2010; Niemi and Fernández, 2010; Sillett et al., 2012).

As with other models we've encountered in this chapter, the distance sampling model, under data augmentation, includes a set of M zero-inflation variables z_i and

a binomial observation model expressed conditional on **z** (binomial for $z_i = 1$, and fixed zeros for $z_i = 0$). In distance sampling we pay for having only a single sample occasion (i.e., $K = 1$) by requiring constraints on the model of detection probability, normally imposed as the assumption that detection probability is 1.0 when distance equals 0. A standard model for detection probability is the "half-normal" model:

$$p_i = \exp(-\alpha_1 x_i^2)$$

for $\alpha_1 > 0$, where x_i denotes the distance at which the ith individual is detected relative to some reference location. This encounter probability model is more often written with $\alpha_1 = 1/2\sigma^2$. If $K > 1$ then an intercept in this model, say α_0, is identifiable and such models are usually called "capture-recapture distance sampling" (Alpízar-Jara and Pollock, 1996; Borchers et al., 1998).

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). Specification of this distance sampling model in the **BUGS** language is shown in Panel 4.2, taken from Royle and Dorazio (2008).

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

$$p_i = h(||\mathbf{u}_i - \mathbf{x}_0||, \alpha_1)$$

along with

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

```
alpha1 ~ dunif(0,10)           # Prior distributions
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 function
   logp[i] <-   - alpha1*(x[i]*x[i])
   mu[i] <- z[i]*p[i]
   y[i] ~ dbern(mu[i])         # Observation model
 }

N <- sum(z[1:(nind+nz)])       # N is a derived parameter
D <- N/striparea               # D = N/total area of transects
```

PANEL 4.2

Distance sampling model in **BUGS** for a line transect situation, using a half-normal detection function.

where $\mathbf{x}_0$ is a fixed point (or line) and $\mathbf{u}_i$ is the individual's location, which is observed for the sample of 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, we provided the analogous argument in the previous section.

The preceding paragraph makes it clear that distance sampling is a special case of spatial capture-recapture models, such as those derived from model M_x of the previous section, where the encounter probability is related directly to *distance*, which is a reduced information summary of *location*, $\mathbf{u}$. ~~Some~~ intermediate forms of SCR/DS models can be described (Royle et al., 2011a). In the context of our general characterization of SCR models (Chapter 2.6), we suggested that every SCR model can be described, conceptually, by a hierarchical model of the form:

$$[y|\mathbf{u}][\mathbf{u}|\mathbf{s}][\mathbf{s}].$$

Distance sampling ignores the part of the model pertaining to $\mathbf{s}$, and deals only with the model components for the observed data $\mathbf{u}$.[7] Thus, we are left with a hierarchical model of the form

$$[y|\mathbf{u}][\mathbf{u}].$$

In contrast, as we will see in the next chapters, many SCR models (Chapter 5) ignore $\mathbf{u}$ and condition on $\mathbf{s}$, which is not observed:

$$[y|\mathbf{s}][\mathbf{s}]$$

Since $[\mathbf{u}]$ and $[\mathbf{s}]$ are both assumed to be uniformly distributed, these are equivalent models! The main differences have to do with interpretation of model components and whether or not the 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 trying to collect distance sampling data on species such as jaguars or 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 and then apply SCR models. Furthermore, as we will see in Chapter 15, SCR models can make use of distance data, allowing us to study distribution, movement, and density. Thus, SCR models are more general and versatile than distance sampling models, and can accommodate data from virtually all animal survey designs.

4.6.1 Example: Sonoran desert tortoise study

We illustrate the application of distance sampling models using data on the Sonoran desert tortoise (*Gopherus agassizii*), shown in Figure 4.6, collected along transects in

[7] Equivalently, we could also say that $[\mathbf{u}]$ in the distance sampling model is $[\mathbf{u}] = \int [\mathbf{u}|\mathbf{s}][\mathbf{s}]d\mathbf{s}$.

FIGURE 4.6

Desert tortoise in its native habitat (*Photo credit: Erin Zylstra, Univ. of Arizona*).

southern Arizona (see Zylstra et al. (2010) for details). The data are from 120 square transects having four 250-m sides, although we ignore this detail in our analysis here and regard them as 1 km transects, and we pooled the detection data from all 120 transects. The histogram of encounter distances from the 65 encountered individuals is shown in Figure 4.7.

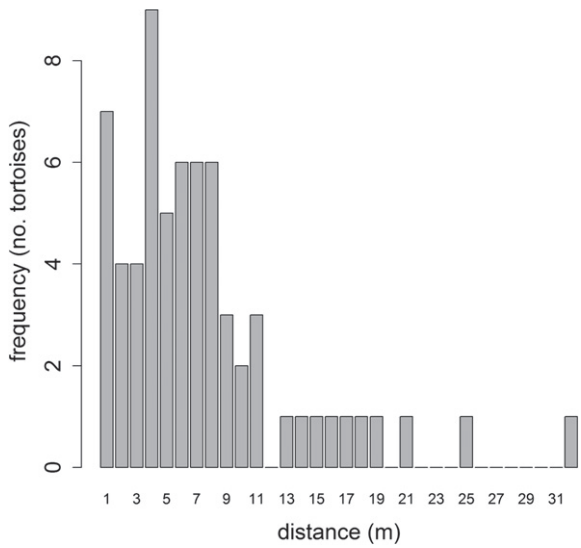

FIGURE 4.7

Distance histogram of $n = 65$ Sonoran desert tortoise detections from a total of 120 km of survey transect.

Table 4.8 Posterior summaries from the Sonoran desert tortoise distance sampling data. Results were obtained using **WinBUGS** running 3 chains, each with 3,000 iterations and the first 1,000 discarded, thinning by 2.

Parameter	Mean	SD	2.4%	50%	97.5%	Rhat	n.eff
α_1	0.01	0.00	0.00	0.01	0.01	1.02	130
σ	9.12	0.77	7.77	9.07	10.77	1.02	130
N	516.67	54.71	415.00	516.00	632.00	1.02	100
D	0.54	0.06	0.43	0.54	0.66	1.02	100
ψ	0.61	0.07	0.49	0.61	0.75	1.02	96

Commands for reading in and organizing the data for analysis using **WinBUGS** are given in the help file `?tortoise` provided with the `scrbook` package. To compute density, the total sampled area of the transects `striparea` is input as data, and computed as: 120 (transects) multiplied by the length (1,000 m) and half-width ($B = 40$ m), then multiplied by 2, and divided by 10,000 to convert to units of individuals/ha. We also provide commands for analyzing the data with `unmarked` (Fiske and Chandler, 2011) using hierarchical distance sampling models (Royle et al., 2004).

Posterior summaries for the tortoise data are given in Table 4.8. Estimated density (posterior mean) is 0.54 individuals/ha and the estimated scale parameter of the distance function (posterior mean) is $\sigma = 9.12$ m. The posterior mass of the data augmentation parameter ψ is located away from the upper bound $\psi = 1$ and so the degree of data augmentation appears sufficient.

4.7 Summary and outlook

Traditional closed population capture-recapture models are closely related to binomial generalized linear models. Indeed, the only real distinction is that in capture-recapture models, the population size parameter N is unknown. This requires special consideration in the analysis of capture-recapture models. The classical approach to inference recognizes that the observations don't have a standard binomial distribution but, rather, a truncated binomial (from which the so-called *conditional likelihood* derives) since we only have encounter frequency data on observed individuals. If instead we analyze capture-recapture models using data augmentation, which arises under a discrete-Uniform(0, M) prior for N, the observations can be modeled using a zero-inflated binomial distribution. The analysis of such zero-inflated models is practically convenient, especially using the **BUGS** variants.

Spatial capture-recapture models considered in the rest of the chapters of this book are closely related to individual covariate models (model M_x). Spatial capture-recapture models arise naturally by defining individual covariates based on observed locations of individuals—we can think of using some function of mean encounter location as an individual covariate. We did this in a novel way, by using distance to the centroid of the trapping array as a covariate. We analyzed the *full likelihood* using

data augmentation, and placed a prior distribution on the individual covariate which was derived from an assumption that individual locations are, a priori, uniformly distributed in space. This assumption provides for invariance of the density estimator to the choice of area (induced by maximum distance from the centroid of the trap array). The model addressed some important problems in the use of closed population models: it allows for heterogeneity in encounter probability due to the spatial juxtaposition of individuals with the array of traps, and it also provides a direct estimate of density because area is a feature of the model (via the prior on the individual covariate). The model is still not completely general, however, because it does not make full use of the spatial encounter histories, which provide direct information about the locations and density of individuals.

A specific individual covariate model that is in widespread use is classical distance sampling. The model underlying distance sampling is precisely a special kind of SCR model—but one without replicate samples. Understanding distance sampling and individual covariate models more broadly provides a solid basis for understanding and analyzing spatial capture-recapture models. In fact if, instead of placing an explicit model on *distance* in the classical distance sampling model, we were to place the prior distribution on *location*, $\mathbf{s}$, of each individual, then the form of the distance sampling model more closely resembles the SCR model we introduce in the next chapter.

This page is intentionally left blank

Non Print Items

Abstract: In this chapter we introduce traditional (non-spatial) closed-population capture-recapture models for estimating abundance, emphasizing analysis in a Bayesian framework. If population size (N) is known, these models resemble simple logistic regressions, where the observations ($0 =$ not captured or $1 =$ captured on each occasion) are Bernoulli trials with detection probability p. Usually, though, we are interested in estimating N, and we can do so using data augmentation, which involves adding a large number of all-zero encounter histories to the n observed encounter histories, and estimating how many of these hypothetical individuals are part of the population but were never observed. This reformulation of the capture-recapture model facilitates Bayesian analysis and the inclusion of individual covariates, and thus we use data augmentation frequently throughout the rest of the book. To demonstrate data augmentation, we present examples and **BUGS** code for non-spatial capture-recapture models incorporating different sources of variation in p. The drawback of non-spatial estimates of abundance is that they are not linked to a specific area and so ad hoc approaches are necessary to define an area in order to estimate density. We present some common approaches to do so and discuss their shortcomings. We show that individual covariate models, where the covariate is some description of an animal's location in space, such as average capture location, are a step towards fully spatial capture-recapture models. The last section of this chapter highlights the parallels of distance sampling and spatial capture-recapture. Providing background on capture-recapture in general, the chapter facilitates understanding of spatial capture-recapture models, which are fully introduced in the following chapter.

Keywords: Behavioral response, Data augmentation, Density estimation, Detection probability, Distance sampling, Effective area sampled, Heterogeneity model, Individual covariate model, MMDM, Model Mh, Non-spatial capture-recapture