27. Estimating Abundance

Chris Sutherland & Andy Royle

1. Introduction

Fundamental to much of applied ecology, including species management and conservation, is the ability to reliably estimate population size, or *abundance*. This is particulary true for reptiles given growing evidence of declines globally (Gibbons et al. 2000, Reading et al. 2010) and high levels of data deficiency (Böhm et a.l 2013). The major challenge arises because rarely are all the individuals in a population encountered during a survey, and that the resulting *counts*, i.e. the number of individuals encountered, C, represents only some fraction of the true abundance, N. Although the distinction between counts and true abundance is an important one, the two are intuitively related and, by repeatedly sampling and marking individuals in a population, capture-recapture methods provide a formalization of this relationship generating estimates of the true population size.

In this chapter, we provide a non-technical overview of 'closed population capture-recapture' models, a class of well established models that are widely applied in ecology (Borchers et al. 2002, Williams et al. 2001), and regularly adopted for studies of reptiles (Mazerolle et al. 2007), to estimate abundance from counts of marked individuals by accounting for imperfect detection. We first describe some classic closed population models for estimating abundance (Otis et al., 1978), then consider some recent extensions that provide a spatial context for the estimation of abundance, and therefore density, D (Spatial or spatially-explicit Capture-Recapture: Efford 2004, Royle et al. 2014), and finally provide an example of estimating abundance and density of reptiles using an artificial cover object survey of slow worms, *Anguis fragilis*.

2. Closed Population Capture-Recapture

2.1. Sampling a population

The primary objective in a closed population capture-recapture study is to estimate the abundance of a population of interest. The population of N individuals is subjected to repeated sampling for a specified number of occasions, say K where, in the first sampling occasion, all captured individuals are marked an released, and then at each subsequent sampling occasion the detection of marked

individuals is recorded and unmarked individuals are marked. Identifying a focal population, the spatial extent of the population is implicitly defined and the method of capture depends on the species in question and available resources. For reptiles, survey methods that allow individuals to be captured and marked include, for example, visual searches within a defined area (Zylstra et al. 2010), cage traps (Tyrrell et al. 2009, Christy et al. 2010) or pitfall traps, and the use of artificial cover objects (CHRIS?) (see also Chapter X). Once captured, individuals can be uniquely identified using either natural markings that can be used to determine individual identity (), using tags or colored markings () or physical marking such as toe clipping () (see also Chapter X).

Such repeated sampling results in individual encounter histories that, for each of the $i = 1, \ldots, n$ individuals encountered, describes whether or not individuals were detected in each of the K occasions. For example, in a $K = 4$ occasion capture-recapture study, an individual with an encounter $y_i = (0101)$ was encountered $y_i = 2$ times; first in occasion 2, and then again in occasion 4, and was not encountered in occasions one or three. In Table 1 we provide an example of encounter history data for a $K = 4$ occasion capture-recapture study during which $n = 8$ individuals were captured.

Table 1: An example of an encounter history for a $K = 4$ occasion capture-recapture study during which $n = 8$ individuals were detected.

	Occasion				
Individual	1	2	3	4	y
1	0	1	0	1	2
2	1	0	0	1	2
3	0	1	1	1	3
4	1	1	0	0	2
5	1	0	0	1	2
6	0	0	1	1	2
7	0	0	1	0	1
n=8	1	0	0	1	2

Estimating abundance using encounter history data collected using the general sampling scheme we have described above is basically the process of estimating how many individuals were *missed*, i.e., how many individuals have encounter history $y_i = 0$. The ability to do so requires that the following basic assumptions are met:

1. the population is closed to demographic processes and to movement

2. individual marks can be identified unambiguously and are not lost

3. individuals are equally likely to be captured

A 'closed' population is one that experiences no additions or subtractions for the duration of the study, and whose size is therefore assumed to be fixed during

sampling. Defining a sampling period over which the assumption of closure can be satisfied means that an individual detected at least once during the study was present for the entire study, and therefore, failure to detect that individual in any occasion was due to imperfect detection. This highlights the importance of the second assumption – that individuals are identified unambiguously – because misidentification would lead to erroneous encounter histories that don't reflect the true process of encountering individuals. The third assumption is less important as we will see later but satisfying this assumption means that we can employ the simplest formulation of a capture-recapture model, model M_0.

2.2. Estimating abundance using model M_0

Under model M_0, the encounter probability for each individual, p_i, is assumed to be the same for all individuals in the population, i.e., $p_i = p$. Then, whether or not we encounter an individual $i = 1, 2, \ldots, N$ during sampling occasion k, y_{ik}, is a Bernoulli trial (a "coin flip") with constant probability p, which we can write formally as:

$$y_{ik} \sim \text{Bernoulli}(p).$$

i.e., there are no individual or temporal covariates that affect p. The basic idea of all closed population capture-recapture methods is that the pattern of detections in the encounter histories of individuals observed at least once provides information about individual detectability which, or detection probability, p which in turn, can be used to estimate the number of individuals *not* encountered. The underlying concept can be understood by recognizing that the observed number of individuals n is related to the total population size N by the expression:

$$E(n) = N\tilde{p}$$

where $E()$ denotes statistical expectation and $\tilde{p}$ is the probability that an individual is captured *at all* during the study. In a study of K survey occasions $\tilde{p}$ is directly related to the basic parameter p by the formula

$$\tilde{p} = 1 - (1 - p)^K$$

This expression applies to model M_0 but the expression relating p to $\tilde{p}$ is different depending on the specific capture-recapture being considered. In general the parameter p can be estimated from the observed encounter histories and, in turn, this is used to estimate $\tilde{p}$ and then finally we estimate N by $\hat{N} = n/\tilde{p}$.

XXXXX need to work on this next paragraph XXXX

Thus, for example, consider our individuals in the above table where we see 8 individuals sampled over 4 periods producing a total of 16 captures. Thus, on average, the capture probability for these 8 individuals was 0.50. We note that this is actually *not* an unbiased estimator of the probability of capture because we only observe , in our data set, captured individuals and they necessarily have at least one encounter event in their history. So, under the binomial model, we have some all-zero encounter histories that we don't observe and so our

naive estimate of 0.50 is , in fact, biased high. The true value should be much smaller. Part of the technical ideas underling closed capture-recapture models is to account for this bias in the observed sample and produce the correct unbiased estimate of p. One way to do this is using the so-called conditional likelihood and another approach is to use the 'full likelihood' which actually has a few different variations. In any case, without giving these details (See Royle et al. 2014 sec. xxxx or Borchers et al...) the conceptual motivation for capture-recapture is to estimate the probability of capturing an individual *at all* that exists in this population from the observed encounter histories. Call this probability $p_{cap} = (1 - p)^4$ where p is the per-occasion capture-probability, a fundamental parameter of the model. We would estimate p using the condietional likelihood or full likelihood approaches and then just plug in p into the expression for p_{cap} and we find that $p = 0.40$ (or whatever it is) and therefore $p_{cap} = .8$ (or whatever it is).

The conditional probability of encounter 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 2nd 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
XXX end XXXXX

Thus, in model M_0 there are two parameters of interest that are estimated using individual encounter histories: p, the probability of encountering an individual during a sampling occasion; and $n0$, the number of individuals present in the population that went unobserved (note that total population size is $N = n+n0$).

2.3. Variation in p: beyond M_0

The assumption of equal capture probability is a rather restricting one and there are many situations under which the capture probabilities would be expected to vary. For these situations, model M_0 is not appropriate although and Otis et al. (1978) described a family of models that can be used to deal with most sources of variation in individual encounter probabilities:

M_0 Capture probability is the same for all individuals

M_t Capture probability is the same for all individuals, but varies between sampling occasions (**t**ime) (note that we use k for the time index in our development).

M_b Capture probabilities vary depending on whether or not individuals have been captured previously (**b**ehavioral response)

M_h Capture probabilities vary among individuals (individual **h**eterogeneity).

Variations of these different models exist. For example, the usually application of model M_t involves occasion-specific parameters p_k but we can also consider systematic variation in detection probability that results from explicit

covariates. For example, we could model season variation in p_k using a quadratic polynomial in Julian day, J_k, such as:

$$logit(p_k) = \alpha_0 + \alpha_1 J_k + \alpha_2 J_k^2$$

and estimate the parameters α_0, α_1 and α_2 instead of fully occasion specific parameters p_k. The behavioral response model is usually parameterized as a permanent change in p for individuals subsequent to their initial capture. This could be the result of 'trap happiness' due to having baited traps (uncommon in reptile studies) or it could be the result of 'trap shyness' due to aversion to handling. Sometimes a transient or ephemeral behavioral response may be sensible (Yang and Chao 2005). Under this type of model the response to initial capture only lasts for a brief period after initial capture. The heterogeneity model, model M_h, has many different versions in existance. Norris and Pollock (1996) formulated the model in terms of a finite mixture or latent class model in which each individual in the population belongs to a finite (and small) number of classes represented by distinct values of p (see also Pledger 2000). Dorazio and Royle (2003) considered continuous mixtures such as the beta-binomial and logit-normal models (see also Coull and Agresti 1999).

2.4. Removal Sampling

2.5 Individual covariate models and distance sampling

Model M_h is the standard closed population model when *unexplained* individual heterogeneity in capture-probability exists. However an important related class of models are models in which individual heterogeneity can be explained by explicit individual covariates. These are often called "individual covariate models" but, in keeping with the classical nomenclature on closed population models, Kery and Schaub (2012) referred to these models as "model M_x", the x representing some explicit covariate (of course multiple covariates are allowed). 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 2009, Langtimm et al. 2011).

The basic encounter model for model M_x is the same as our other closed models, the Bernoulli encounter model:

$$y_i \sim \text{Bernoulli}(p_i).$$

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

$$\text{logit}(p_i) = \alpha_0 + \alpha_1 x_i \tag{1}$$

where x_i is the covariate value for individual i and the parameters α_0 and α_1 are the parameters to be estimated.

Traditionally, estimation of N in model M_x is achieved using methods based on ideas of unequal probability sampling (i.e., Horvitz-Thompson estimation). This idea was developed independently by Huggins (1989) and Alho (1990). The estimator of N is given as a derived parameter:

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

where $\tilde{p}_i$ is the probability that individual i appeared in the sample. This is related to the more fundamental parameters α in the model for detection probability according to:

$$\tilde{p}_i = 1 - (1 - p_i)^K$$

where p_i is a function of parameters α_0 and α_1. In practice, parameters are estimated from the conditional-likelihood of the observed encounter histories.

An alternative formulation of model M_x is the "full likelihood" which requires that we put a model on the individual covariate x allowing for the sample not only of the encounter histories but also of the covariate 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:

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

If the covariate was group size then, naturally, some discrete probability mass function would be needed. Inference for individual covariate models from the standpoint of the full likelihood is discssed in Royle (2009), Kery and Schaub (2012), etc..

Individual covariate models are important in practice for the simple reason that heterogeneity exists in almost every capture-recapture study due to the spatial organization of traps and of individuals in the population (see next section). Thus they were adopted historically to account for spatial structure in capture-recapture (Boulanger and McLellan 2001, Karanth and Nichols 1998). For this purpose an individual covariate is created which describes *where* the individual is located in relation to the trapping array. This approach leads naturally to more recent spatial capture-recapture models described in the next section.

3. Spatial Capture-Recapture

One of the main deficiencies with classical closed population models is that they do not permit direct estimation of animal *density* because, in almost all practical field applications, it is not possible to precisely define the area sampled by a set of trapping devices. This is because individuals being captured move about space and can be captured without the biologists knowing from whence those individuals originated or how much space they are using. Newly developed *spatial capture-recapture* (SCR) models (also called spatially-explicit capture-recapture, SECR) provide a technical framework for dealing with this problem

(Efford 2004, Borchers and Efford 2008, Royle and Young 2008, Royle et al. 2014).

The sampling scheme for a *spatial* capture-recapture analysis is the same as described above, i.e., there is a population of N individuals, but now we consider each individual having an activity center that has X and Y coordinates ($\mathbf{s}_i = (s_{i,X}, s_{i,Y})$). Now the goal is to estimate the number of individuals (or activity centers) within a region of interest which we refer to as a *state-space*, or $\mathcal{S}$, which is to say we wish to estimate density: $D = N/||\mathcal{S}||$, where $||\mathcal{S}||$ is the area of $\mathcal{S}$. We assume that these activity centers are distributed uniformly throughout across space:

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

As before, the population is subjected to sampling using some trapping devices (for convenience, we will refer to these as 'traps'). However, we explicitly acknowledge both how many traps there are: $j = 1, ..., J$ traps, and the locations of each of the traps, which we will call these locations x_j. The acknowledgement of the spatial structure of the traps means observations can be spatially indexed so encounter histories describe *who* (i), *when* (k), and importantly, *where* (j) individuals were located, i.e., $y_{i,j,k}$. Typically, these observations are assumed to be binomially distributed with sample size K (the number of sampling occasions):

$$y_{i,j} \sim \text{Binomial}(K, p_{i,j}),$$

where $p_{i,j}$ is the probability of encountering individual i in trap j, which depends on the distance between the trap location (x_j) and the individuals activity center (s_i) as follows:

$$p_{i,j} = p_0 \times e^{-\alpha_1 \mathrm{d}(x_j, s_i)^2}. \tag{2}$$

This is refered to as the 'bivariate normal' encounter model where $\text{logit}(p_0) = \alpha_0$ is the baseline encounter probability, which is the probability of encountering an individual at its' activity center, $\alpha_1 = 1/(2\sigma^2)$ describes the distance over which detection declines, and $\mathrm{d}(x_j, s_i)$ is the Euclidean distance between trap j and the activity center of individual i. In a spatial capture-recapture analysis, the parameters to be estimated are α_0 and α_1.

SCR models address the density estimation problem by parameterizing the model directly in terms of individual activity centers s_i and prescribing the state-space $\mathcal{S}$. The inference problem then reduces to estimating the number of such activity centers in the well-defined area $\mathcal{S}$, i.e., density. While SCR models resolve this problem, they also enable researchers to study many aspects of spatial ecology from individual encounter history data, including resource selection or space usage (Royle et al. 2013), landscape connectivity (Royle et al. 2013b; Sutherland et al. 2014), spatial variation in density (Borchers and Efford 2008, Royle et al. 2013), and movement or dispersal (Schaub and Royle 2014, Ergon and Gardner 2014, Royle et al. *in review*).

4. Software

CPCR:
MARK
Rmark
unmarked for fitting certain types of hierarchical capture-recapture models.
BUGS/JAGS
SCR:
Density superceded by secr
oSCR

5. Slow worm example

We provide an example here using a study of the slow worm from (Anguis fragilis) at Mueterschwanderberg, canton Nidwalden, Switzerland (Meier 2012, thesis). A more detailed SCR analysis of the data from that study is given by Meier et al. (in prep). Here we analyze a portion of the data from one of the trapping grids, shown in Fig. XXXX.

0.1 Summary

MARK

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 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 etal, 2007; royle 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 (Ch. 6 kery schaub 2011) for an accessible and complementary development of Bayesian analysis of ordinary, i.e., nonspatial closed population models.
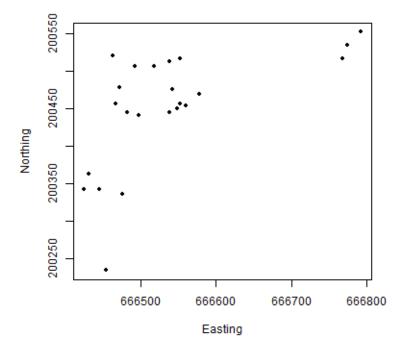
Figure 1: array of artificial cover boards used in the slow worm study from Meier (2012).