Lab 7 – Closed capture models AND advanced covariate application in RMark

Below I've provided just the briefest of background information on closed capture models – for a much more thorough (to the tune of 43 pages) treatment see Chapter 14 of the Gentle Introduction, which was authored by Paul Lukacs from the University of Montana, and Chapter 8 of Powell and Gale.

Exercise 1 - This exercise gives a brief introduction into the analysis methods collectively referred to as closed capture, which is one of the more foundational mark-recapture analysis types and is typically used to derive estimates of abundance (N). As the name implies, a closed capture model assumes closure – no animals enter (through birth or immigration) or leave (through death or emigration) the sampling area in between successive sampling events. If a population is closed and we sample it by capturing animals, we can estimate abundance based on the number we catch and the probability that any individual is captured based on the following simple equation:

$$\widehat{N} = \frac{n}{\hat{p}}$$

So how do we derive estimates of p? From an introductory ecology course you may remember the foundational and most basic method for estimating abundance using the Lincoln-Peterson estimator, were we capture, mark, and release some number of individuals (n1), and then during a secondary sample we capture another round of animals (n2) where some proportion of those individuals were previously marked (m2): The Lincoln-Peterson estimator gives us an estimate of abundance, n-hat, as:

$$\widehat{N} = \frac{n_1 n_2}{m_2}$$

where the probability of capture, p, is estimated as m2/n2. As Lukacs points out in Chapter 14, the process of estimating abundance in a closed capture framework is really the process of estimating detection probabilities, and using these to derive abundance. In a nutshell, and just like the open capture CJS models we've worked with so far, closed capture models use data derived from the encounter histories fit to a model likelihood to estimate N directly (this is referred to as a full likelihood – see 14.3 in GI), or N can be estimated as a derived parameter under a conditional likelihood. In either case, we can think of N as being estimated based on the total number of animals that were captured and marked during the course of the study (M_t), and the probability that an animal is captured and marked at least once, which we will term *p**. In a two occasion study, N could be estimated as

$$\widehat{N} = \frac{M_t}{(1 - (1 - p_1)(1 - p_2))}$$

where the (1-(1-p)(1-p)) term defines p* (as a side note, this is a pretty useful probability statement for any situation where you want to estimate whether something is likely to happen ≥1 time). This statement basically says that if during a course of a study we encounter and mark 100 animals, and the probability of capturing an animal in order to mark it is 0.5, then the total abundance in the area we are sampling is estimated as 200 animals. Importantly, this estimate is only relevant to the animals that were actually available for capture, and it only extends to an estimate of population size insomuch as every member of the population was available to sample. For this reason you'll often see these estimates of N expressed per unit area (i.e. density) based on the size of the sample area.

Basic assumptions of closed capture models are as follows:

1. The population is closed – no births, deaths, emigration or immigration occur during the sampling period. The practical implication of this is that you will need to craft a study design and sampling scheme that works with the biology of your organism. If we trapped bog lemmings once in June and again in August, we would probably violate this assumption. If we did the same for moose, we may have less of an issue.
2. Marks are not lost, and they are not overlooked by observers (e.g. if your summer intern forgets to check for ear tags).
3. All animals are equally likely to be captured, regardless of previous capture history.

Meeting assumptions #1 and #2 are largely a function of study design (and proper training of interns), but in practice assumption #3 is regularly violated under a simple Lincoln-Peterson approach because individuals often deviate in their capture likelihood for reasons that cannot be addressed through sampling design. For this reason a suite of more complex closed capture model iterations exist and are implemented in MARK. These models will incorporate a new parameter that describes the recapture probability (c), which defines the probability that a previously marked animal will be recaptured after its initial marking within a particular trapping session. Including c in the model likelihood allows us to estimate variation in capture probability between marked and unmarked animals which could result from, for example, trap happy or trap shy behaviors. The classic variations on closed-capture models that relax the assumptions of equal detection include the following (with the standard notation from Otis et al. 1976 that you'll commonly see used):

M0 – the simplest model, all assumptions above are required.

Mt – Capture probabilities may vary among occasions. Assumes equal probabilities among individuals within occasions.

Mb – Allows a behavioral response between animals that were previously marked, and those newly captured.

Mbt – Allows a behavioral response and a temporal response.

Mh – Full heterogeneity is allowed among all individuals (difficult to achieve).

Other biologically meaningful (and measurable) sources of heterogeneity in detection can be modelled as group effects (e.g. sex or habitat type). One situation in which assumption 3 is violated that is not so easily dealt with is the case of spatial heterogeneity among individuals. For example, let's say we set a 25x25 trapping grid for marten with traps spaced at 1 km intervals. Animals whose home ranges are close to the center of the grid are going to have lower capture probability than an animal on the edge of the grid, simply by virtue of the fact that the former are more likely to encounter a trap. To accommodate these cases we will use spatial capture recapture models, which incorporate space use explicitly as a source of heterogeneity, and which we will cover later in the course.

Part 1. Running closed capture models in RMark

You will find the step-by-step instructions for this lab contained in the RStudio Script posted on blackboard. Also there you'll find the data file for the lab (Lab7Data.csv), which is simulated data with 3 capture occasions and three groups labeled A, B, and C. For the purpose of this lab, let's assume that these are mark-recapture data from a fish species, how about Arctic Char (Salvelinus alpinus) that are captured using shore-based fyke nets during 3 consecutive days during their spawning season. The fish are marked with floy tags. The sampling occurs in three different lakes, which constitute the groups. The goal with this exercise will be to generate the most reliable estimates of abundance for each of these three lakes that is possible, and we will use the classic forms of the closed capture models and model-averaging to accomplish this.

Homework

Your homework for this lab will mirror the analysis using a different dataset, and there is one additional twist. The homework data file is simulated and based on 4 groups and 4 sampling occasions. We'll assume these data represent four of the remaining populations of American Burying Beetles (*Nicrophorus americanus*) which are a federally endangered beetle that were once widespread but now occupy a very limited range. Yes, mark recapture can be used on invertebrates, too. You task is to estimate the current population size for each population, AND as an added twist, I would like you to calculate an estimate of p* for each population that is based on model-averaged estimates of detection probability. I haven't given you code to do this explicitly, but as a reminder for above the equation for p* is given as

$$p^* = 1 - (1 - p^n)$$

Where n is the number of sampling occasions. If you have unique values for p during each of the n occasions, than the equation is based on the product of all n values for 1-p:

$$p^* = 1 - \prod_{1:n} (1 - p_i)$$

For this assignment please upload your R code, estimates for population size of each group, and estimates of p* for each group.