

Projects

Counts versus capture histories

Most of the literature on capture-recapture methods expect that data is collected through repeated surveys, generating binary capture histories for individual animals. For example, an animal encountered in the second and last survey out of five would generate a capture history $y = (0, 1, 0, 0, 1)$. Estimating the size of the population is then essentially equivalent to estimating the number of animals with capture history $(0, 0, 0, 0, 0)$, i.e. the animals that are never encountered.

The Swedish brown bear survey takes place during approximately 11 weeks in autumn. Samples are collected continuously, and individual capture histories are generated by splitting the survey-period into weeks. In a particular week, an individual is denoted as captured if it is encountered at least once during the week.

There are at least three issues with this approach

- the division into weeks is essentially arbitrary
- there may be information lost in dichotomizing the weekly number of encounters
- it requires weekly estimates of capture probabilities (many parameters to estimate)

You can find the code generating the results of the 2020 survey at <https://github.com/mskoldSU/BjornYZ2020>. A description of the methodology can be found in Kindberg et.al. (2011).

Approach using total counts

An alternative approach is to consider the sequence of encounters of a particular animal as a (possibly inhomogenous) Poisson process in time. The total number of encounters of an animal will then follow a Poisson distribution. Since we do not observe the animals with zero encounters, the observed counts will follow a zero-truncated Poisson distribution with probability mass function

$$P(Y = y) = p_\lambda(y)/(1 - p_\lambda(0)), \quad y = 1, 2, \dots,$$

where $p_\lambda(y) = \exp(-\lambda)\lambda^y/y!$, $y = 0, 1, \dots$, is the probability mass function of the $\text{Poisson}(\lambda)$ -distribution. The expected proportion of the total population that are not encountered is $p_\lambda(0) = \exp(-\lambda)$, and we may estimate the total population size as $\hat{N} = n/(1 - \exp(-\hat{\lambda}))$, where n is the number of unique individuals captured.

Individual heterogeneity

In a “perfect” survey, all individuals have the same probability of being captured. Unfortunately, there are strong evidence that capture probabilities vary among individuals. This may be due to

- Geographic variation in effort: The survey is based on volunteers, we do not know how much effort is spent searching for samples and how this varies geographically.
- Geographic variation in capture probability: Most samples are found on roads or paths, animals with home-ranges in areas with a higher density of roads and paths are more likely to be found.
- Variation in genotyping success: Some animals may be easier to identify in the DNA-analysis, depending on e.g. their diet.
- Movement: Some animals may spend part of their time outside the surveyed area, and hence be less likely to be found.
- ...

In a Poisson-distribution, the mean and variance are the same (λ). Individual heterogeneity will manifest itself as an overdispersion in observed counts, i.e. a variance larger than the mean. This is often handled using a so-called *mixture of Poisson distributions*: if we denote by n_i the number of times individual i is captured, then we assume $n_i \sim \text{Poisson}(\lambda_i)$, where $\lambda_i \sim P_\theta$, independently for all i and where P_θ is a distribution that describes the individual variation in catchability. Common choices of P_θ are

- A $\text{Gamma}(\alpha, \beta)$ distribution (in this case one can show that n_i will follow a Negative Binomial distribution).
- A $\text{LogNormal}(\mu, \sigma^2)$ distribution.
- A discrete mixture, e.g. $\lambda_i \in (\theta_1, \theta_2)$, $P(\lambda_i = \theta_1) = 1 - P(\lambda_i = \theta_2) = \theta_3$ (this approach is taken for capture probabilities in the current approach).

Projects

Below I list four project ideas together with a suggested student. Whether you choose to focus on the statistical theory/methods or the more practical consequences for the monitoring program is up to you.

Project: Individual heterogeneity and identifiability (Peter)

Read the article by Link (2003) who argues that it is difficult to distinguish between different models of individual heterogeneity and that they may produce very different results. In this paper, a model with binary capture histories is used. Verify that the results transfers to the over-dispersed zero-truncated Poisson case and illustrate using data from the bear monitoring programme.

Project: Capture histories or counts (Ossian)

Compare the currently used approach (weekly capture-histories) with the above approach of using full counts (zero-truncated Poisson). Both in terms of performance when the Poisson assumption is correct and how sensitive they are to heterogeneity. Also, how does the choice of a “week” affect the current approach? It seems likely that the capture-history approach converge to the approach of full counts as the length of the capture-periods (the “weeks”) approach zero, can you verify this?

Project: Sensitivity to DNA genotyping errors (Herman)

Due to genotyping errors, a scat sample could possibly be mistakenly identified as a capture of a new individual when it is in fact a recapture of an individual already known. If this mistake happens with some (likely to be small) probability p , we expect to see an inflation in the number of individuals captured only once. The resulting distribution of the number of observed counts per individual will be a zero-truncated one-inflated Poisson distribution (if we ignore individual heterogeneity). Use this to investigate sensitivity of population estimates to p . Can we extend the approach to account for individual heterogeneity?

Project: Bears across borders (Martin)

Since bears may have large home ranges, it is often unclear what county (län) they “belong” to. This has important consequences for management, since hunting quotas are set at the county level. If an animal frequently crosses the border between county A and B, it is likely to be captured both when A and B are surveyed. This means it may be counted towards the population in A one year and towards the population in B another, hence counting towards both hunting quotas. This project will try to evaluate the bias caused by double-counting animals within the Swedish survey, and consider methods for counting fractions of animals to correct for this bias. For example based on estimates of how large proportion of an animals home-range belongs to the surveyed county.