A Brief Survey of Capture-Recapture

Notes by Rachel Fewster, *r.fewster@auckland.ac.nz*
Department of Statistics, University of Auckland, New Zealand

What is capture-recapture?

Capture-recapture describes a suite of methods where we make repeated attempts to capture animals (or in general, objects) from a population. We need to be able to recognise each animal individually. The aim is to *learn enough about the recapture process that we can draw conclusions about the animals that were not captured.*

Capture-recapture data typically consist of *capture histories* which specify a 0 (not caught) or a 1 (caught) for each of k capture occasions. The following capture history across $k = 5$ capture occasions denotes an animal that was caught on occasions 3 and 5:

$$0 \quad 0 \quad 1 \quad 0 \quad 1$$

The full data-set is a data frame where each row denotes the capture history of one animal:

$$\begin{array}{ccccc} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ \vdots & & \vdots & & \vdots \end{array}$$

The population probably contains animals with capture history $000\ldots0$ that were never seen, and of course these cannot be listed in the data.

The precise conclusions we wish to draw about the population differ according to context, with the major distinction being between closed and open population models.

- ***Closed population capture-recapture*** involves populations that are assumed to be *closed* for the duration of the study: no births, deaths, immigration, or emigration. Our interest in closed populations is estimating *population size, N.*

  Using the animals that we do capture, we know how often we missed them on each capture attempt. This enables us to estimate the *capture probability* at each occasion, and hence the number of animals that were missed every time: in other words, the number with capture history $000\ldots0$ that were never seen. Along with the animals that we did see, this gives us the population size, N.

  Models for closed populations are often referred to by well-known abbreviations M_0, M_t, M_b, M_h, M_{tb}, M_{tbh}, ... The letters t, b, and h describe characteristics of the recapture process, described later.

- ***Open population capture-recapture*** methods are used to study populations that do have births, deaths, immigration, and emigration in some combination. The questions of interest here fall into two main categories:

  — ***Survival-only models:*** here we only care about estimating survival; not about birth or population size. This is the province of *Cormack-Jolly-Seber (CJS) models.* Estimation can be done easily by conditioning on first capture and disregarding the animals that are never caught. All we need is that the same survival processes should apply to all animals, whether caught or not.

  The idea of survival-only capture-recapture is that there comes a point where a string of zeros in the capture history is unlikely to be due to repeated capture failure and probably indicates that the animal has died. For example, consider the capture history:

  $$1 \quad 0 \quad 1 \quad 1 \quad 1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0$$

  The initial intensity of captures fizzles out at later times. It's likely that the animal has died, although we don't know how many of the final zeros denote death and how many denote lack of capture while alive. We don't attempt to estimate the exact time of death for each animal, but rather to estimate the survival probability,

  $$\phi = \mathbb{P}(\text{survives to time } t+1 \,|\, \text{alive at time } t).$$

  This relies on having enough information about the recapture process throughout the study for the model to be able to distinguish between lack-of-capture and lack-of-survival.

  — ***Survival, recruitment, and population size:*** as soon as we wish to know about population growth or recruitment, in other words births and immigration, we can no longer condition on first capture, because the capture of new individuals that we didn't know about previously is exactly what informs us about birth and immigration. The same goes for estimating population size: the population probably includes animals that were never caught, so we don't gain information about population size by conditioning on capture.

  Jolly-Seber (JS) models were the original models for estimating recruitment and population size along with survival in open populations. These have been improved over the years and there is now a suite of models in this category: notably *POPAN models*; *Pradel models*; and *robust design* models. We will look at all these later.

  Capture histories inform us about birth in much the same way as they inform us about death. For example, consider the capture history:

  $$0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 1 \quad 0 \quad 1 \quad 1 \quad 1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0$$

  If the recapture probability is sufficiently high, it's unlikely that an animal would avoid capture for all of the first 5 capture occasions, and again for all of the last 5. Thus this capture history would tend to suggest an animal that was born at some point after the start of the study, and died before the end of the study. Again, we don't attempt to deduce an exact time of birth or death, but rather to estimate parameters that control the probability of animals joining the population or leaving the population at each time.

The examples above demonstrate that ***learning about the recapture process*** is pivotal to capture-recapture, because this is how we elicit non-captures despite presence in closed population models, and distinguish between non-capture, death, and not-yet-birth in open populations. This explains the pivotal role of ***capture probabilities*** to capture-recapture inference. Although they are nuisance parameters, they underpin all inference of interest. Unfortunately, inference on the interest parameters is often sensitive to violation of assumptions about capture probability, for example if individuals differ in their catchability (known as ***individual heterogeneity***). This is what makes capture-recapture modelling complicated, and as the models become complicated they also have an awkward tendency to become less-than-well-behaved with regards to statistical properties that we often take for granted when modelling.

Where to find out more

A good place to get a quick overview of any particular class of capture-recapture models is the manual for program Mark (Cooch and White, 2013). The recent book *Analysis of Capture-Recapture Data* by McCrea and Morgan (2015) is another excellent resource, being comprehensive, up-to-date and readable. These resources are good for getting anchored in the literature on most aspects of capture-recapture, with the exception of spatial capture-recapture models which are still too recent to be comprehensively covered. Books about spatially-explicit capture-recapture are on the way, as are a couple of review papers by David Borchers.

Closed population capture-recapture

Most mark-recapture models follow the same general structure, given below. Different models allow different simplifications of this structure in terms of sufficient statistics, and sometimes it takes a bit of work to recognise the resulting likelihood as a special case of the general structure. For this reason it is helpful to start with the most general case and simplify for individual models, rather than starting with the simplest cases and generalising. When building complicated models, it is really helpful to know what the formulation 'must' look like in terms of general structure.

We will start with closed populations because the notation and definitions are quite simple.

Design and assumptions:

- There are k capture occasions: $t = 1, 2, \ldots, k$.
- Capture histories of different animals are independent of each other.
- The population is closed, both geographically (no immigrants / emigrants) and demographically (no births /deaths).
- Animals can be correctly recognised on recapture (no mark loss).

Parameters:

- N is the unknown number of animals in the closed population, and is the parameter of interest.

- $\boldsymbol{\theta}$ is a set of parameters controlling *capture probability*.
  - Define $p_{\boldsymbol{\theta}}$ to be the probability that an animal is ever caught. This has to be the same for all animals, because we know nothing about the animals that were never caught.
  - As an example, Model M_t assumes that all animals have the same capture probability at time t. Then $\boldsymbol{\theta}$ is the vector of capture probabilities for times $t = 1, \ldots, k$: so $\boldsymbol{\theta} = (p_1, \ldots, p_k)$. The probability that an animal is *never* caught is

    $$(1 - p_1)(1 - p_2) \ldots (1 - p_k),$$

    so $p_{\boldsymbol{\theta}}$ is the complement:

    $$p_{\boldsymbol{\theta}} = 1 - \prod_{t=1}^{k} (1 - p_t).$$

Data:

- n is the total number of animals ever caught. (The notation D is sometimes used instead of n, where D denotes the number of *distinct* animals in the sample.)
- $\boldsymbol{x}_i$ is the capture history of animal i, where $\boldsymbol{x}_i$ is a string of 0s and 1s of length k: for example we might have $\boldsymbol{x}_i = 00110$ when $k = 5$. We observe capture histories $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$, and the remaining $N - n$ animals in the population have capture history $00 \ldots 0$.

Likelihood structure

The likelihood of the parameters $(N, \boldsymbol{\theta})$ based on data $(n\,; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ is:

$$\mathcal{L}\,(N, \boldsymbol{\theta}) = \binom{N}{n} (1 - p_{\boldsymbol{\theta}})^{N-n} \prod_{i=1}^{n} \mathbb{P}(\boldsymbol{x}_i\,;\, \boldsymbol{\theta})\,. \tag{1}$$

This likelihood differs from run-of-the-mill statistical models in that it is *not a regular likelihood:* it does not comprise independent, identically distributed observations drawn from a specified probability distribution, due to the leading binomial term. Additionally, the range of the data n is truncated by the parameter N, because we require $n \leq N$. The non-regularity means that we cannot assume that standard maximum likelihood theory will necessarily apply. In particular, we will see below that the asymptotic distribution of $\hat{N}$ is lognormal, rather than normal.

It is perhaps more obvious to see why (1) is the right formulation if we rewrite it as follows:

$$\mathcal{L}\,(N, \boldsymbol{\theta}) = \binom{N}{n} p_{\boldsymbol{\theta}}^{n} (1 - p_{\boldsymbol{\theta}})^{N-n} \prod_{i=1}^{n} \left\{ \frac{\mathbb{P}(\boldsymbol{x}_i\,;\, \boldsymbol{\theta})}{p_{\boldsymbol{\theta}}} \right\}. \tag{2}$$

Conceptually, we first choose n of the N animals to be detected, with probability $p_{\boldsymbol{\theta}}$ each. The likelihood of this step depends upon both N and $p_{\boldsymbol{\theta}}$. We then go through the chosen animals $i = 1, \ldots, n$ one at a time, and factor in the likelihood of the specific capture history of animal i, conditional on animal i being detected. The likelihood factorization amounts to the formulation:

$$\mathbb{P}\left(n, \{\boldsymbol{x}_i\}\right) = \mathbb{P}(n)\, \mathbb{P}\left(\{\boldsymbol{x}_i\} \,|\, n\right) .$$

To check that (2) has the correct structure, imagine that there is only one capture occasion, leading to only two possible capture histories: $x_i = 1$ with probability $p_{\boldsymbol{\theta}}$, or $x_i = 0$ with probability $1 - p_{\boldsymbol{\theta}}$. Then (2) reduces to the standard binomial likelihood,

$$\mathcal{L}\left(N, \boldsymbol{\theta}\right) = \binom{N}{n} p_{\boldsymbol{\theta}}^{n} \left(1 - p_{\boldsymbol{\theta}}\right)^{N-n},$$

which is correct, despite being useless for estimation (as there are two parameters and only one piece of data, n).

Estimation using the full likelihood

The likelihood $\mathcal{L}\left(N, \boldsymbol{\theta}\right)$ in (2) can be maximized with respect to N and $\boldsymbol{\theta}$ to find the maximum likelihood estimates, $\widehat{N}$ and $\widehat{\boldsymbol{\theta}}$. As usual, we typically minimize the negative log-likelihood. Although parameter N (the population size) is strictly-speaking an integer, it is much easier to treat it as a continuous parameter so that the usual optimization functions can be used. The following reference list is useful:

- $\log(\,N!\,) =$ `lgamma(N+1)`, where `lgamma` is the log-gamma function in R, and this relationship extends continuously when N is treated as a continuous parameter. This means that $\log(\,N!\,)$ can be replaced by `lgamma(N+1)` for all $N \in \mathbb{R}^+$ when coding the likelihood.
- There are a couple of ways of computing $\log \binom{N}{n}$ in R:

$$\begin{aligned} \log\left\{\binom{N}{n}\right\} &= \log(\,N!\,) - \log\left\{(N-n)!\right\} - \log\left(n!\right) \\ &= \texttt{lgamma}(\texttt{N}+1) - \texttt{lgamma}(\texttt{N}-\texttt{n}+1) - \texttt{lgamma}(\texttt{n}+1) \\ &= \texttt{lchoose}(\texttt{N},\texttt{n}). \end{aligned}$$

- The first and second derivatives of the log-gamma function are respectively called the digamma and trigamma functions. In particular, the following identities are useful for brave souls who wish to compute analytical likelihood derivatives:

$$\begin{aligned} \frac{\partial}{\partial N} \log\left\{\binom{N}{n}\right\} &= \texttt{digamma}(\texttt{N}+1) - \texttt{digamma}(\texttt{N}-\texttt{n}+1) \\ &= \sum_{r=N-n+1}^{N} \frac{1}{r} \qquad \text{for } N \in \mathbb{R}^+ \text{ and } n \in \mathbb{N}^+ \\ &= \texttt{sum(1/seq(N-n+1,N)) ;} \end{aligned}$$

and

$$\begin{aligned}\frac{\partial^2}{\partial N^2}\log\left\{\binom{N}{n}\right\} &= \texttt{trigamma(N + 1)} - \texttt{trigamma(N - n + 1)}\\ &= \sum_{r=N-n+1}^{N} -\frac{1}{r^2} \qquad \text{for } N \in \mathbb{R}^+ \text{ and } n \in \mathbb{N}^+\\ &= \texttt{-sum(1/seq(N-n+1,N)}^2\texttt{)} \;.\end{aligned}$$

- The gamma and digamma functions have their own Wikipedia pages:

  `https://en.wikipedia.org/wiki/Gamma_function`

  `https://en.wikipedia.org/wiki/Digamma_function`

  Note that the digamma and trigamma functions are derivatives of the *log* of the gamma function.

The non-regularity of the likelihood follows through to a non-standard asymptotic distribution of the MLEs. The following result is proved in Fewster and Jupp (2009):

Result: The MLE of population size, $\widehat{N}$, has an asymptotic **log-normal distribution**; whereas the MLEs of the capture parameters, $\boldsymbol{\theta}$, have asymptotic **normal distributions**.

(Practical experiment suggests that there is only a small difference between CIs for N calculated under lognormal and normal protocols, except perhaps as the sample size transitions from 'nowhere-near-good-enough' to 'nearly-good-enough', at which point the CI coverage properties do seem to be a little better under the lognormal approximation. Nevertheless, we might as well get it right.)

Practical guide to DIY closed-model capture-recapture using the full likelihood

The following steps usually seem to work pretty well.

1. Create a suitably awesome parametrization of the capture history probabilities $\mathbb{P}(\boldsymbol{x}_i\,;\,\boldsymbol{\theta})$ in terms of the parameters $\boldsymbol{\theta}$.

   *(**Note:** I didn't really mean that. As we shall see below, 'awesome' models have a habit of becoming statistically unviable. My real recommendation is to sort out as much of the complexity as possible by sampling design, so that at the modelling stage you can hope to use a very simple model.)*

2. Minimize the negative logarithm of equation (1) or (2) with respect to N and $\boldsymbol{\theta}$.

   (I prefer R functions `nlm` and `nlminb`, with `nlm` seemingly the more reliable. Many people use `optim`, but I've found it very unreliable: just my bad luck or really true? Supplying analytic gradient and hessian functions will greatly enhance the performance of `nlm`, but is often unrealistic. Alternatively, ADMB can give excellent results when the model is sufficiently complicated or slow to merit the extra coding effort. ADMB is not a magic bullet: I've seen it fail where `nlm` succeeds effortlessly; however more often ADMB succeeds brilliantly where `nlm` is either extremely slow or needs extremely good starting values to find the optimum.)

3. Find estimated variances or standard errors of the MLEs by one of the following methods:

   - Use `nlm(..., hessian=TRUE)`: the hessian is returned with the final estimates. The following code is illustrative.

     ```
     mle.res <- nlm(negloglike.func, p=startvals, hessian=T, typsize=startvals)

     ## Variance matrix: the inverse hessian
     varmat <- try(solve(mle.res$hessian))

     ## If inverting the hessian succeeded, the estimated variances
     ## are the diagonal elements:
     if(!inherits(varmat, "try-error"))
             var.est <- diag(varmat)
     ```

     The standard errors are the square root of the variance estimates. Note the use of the argument `typsize` in the code above: this tells the optimization algorithm the 'typical size' of the parameter estimates it is looking for. It is particularly useful to improve convergence properties for capture-recapture models, when one of the parameters (N) is of a different order of magnitude to the others, which are typically the size of probabilities or not much larger.
   - If using `nlminb` or another method that does not return the hessian, use function `hessian` in `library(numDeriv)`.
   - If using ADMB, standard errors are returned as part of the final result.

4. For parameter N, use the following **lognormal confidence interval:**

   $$(\widehat{N}/C,\ \widehat{N}C\,),$$

   where

   $$C = \exp\left[1.96\sqrt{\log\left\{1+\frac{\widehat{\text{var}}(\widehat{N})}{\widehat{N}^2}\right\}}\,\right]$$

   This formula applies to 95% confidence intervals. For a general confidence level, replace 1.96 by `qnorm(1-(1-conf)/2)`, for example the value 1.96 is gained from `qnorm(1-(1-0.95)/2)`.

   For the parameters in $\boldsymbol{\theta}$, use **normal confidence intervals:**

   $$\left(\widehat{\theta} - 1.96\sqrt{\widehat{\text{var}}(\widehat{\theta})}\ ,\quad \widehat{\theta} + 1.96\sqrt{\widehat{\text{var}}(\widehat{\theta})}\right)\ .$$

   Again, for confidence levels other than 95%, replace 1.96 by `qnorm(1-(1-conf)/2)`.

   These are known as ***Wald confidence intervals*** (Fewster and Jupp, 2013).

5. Always check the bias and confidence interval coverage properties of estimators by simulation, using generating parameters that realistically mirror those of the real situation. If confidence interval coverage is poor, the sample sizes might not be large enough for Wald confidence intervals to be effective. Profile likelihood confidence intervals are thought to do better than Wald intervals when the sample sizes are small, and are what I would try first, but they might not be successful either. One other possibility is CIs based on inverting a score test, which are also based on the profile likelihood. I haven't tried these, but I would look into them optimistically if necessary. The three types of confidence interval (in slightly impenetrable language) are listed in Section 3.2 of Fewster and Jupp (2013). At some point, bootstrap becomes the only option. With very small samples, the parametric bootstrap can have better properties than the nonparametric bootstrap, even though it does seem a bit like cheating, because there is not enough variability in the small sample for the nonparametric bootstrap to adequately capture the true variance. Alternatively, if you have informative prior information, go Bayesian to use it and reduce the curse of small samples.

Notes: *1. We shall see below a seemingly-innocuous model (M_{tb}) that needs enormous sample sizes before the MLEs and their confidence intervals have good statistical properties, despite only having a few parameters. The concept of 'small' sample sizes is relative to the information content of the data for the parameters of interest. Capture-recapture data contains a lot of information in principle, such as behavioural responses, but some of it is too subtle for realistic sample sizes to tease it apart from other factors.*

2. Profile likelihood confidence intervals take some extra coding: I'm not sure but they might be obtainable from ADMB? Profile likelihood CIs are supplied in program Mark, at least for some models: I'm not clear on how widespread the option is. The Mark manual (Cooch and White, 2013) says both good things and bad things about them: see sections 1.7 for the good and 14.10.1 for the bad.

Estimation using the conditional likelihood

Writing the likelihood in form (2) has the advantage of portraying the capture histories as IID observations from a normalised probability distribution. Explicitly:

$$\begin{aligned}\mathcal{L}(N,\boldsymbol{\theta}) &= \binom{N}{n} p_{\boldsymbol{\theta}}^{n}(1-p_{\boldsymbol{\theta}})^{N-n}\prod_{i=1}^{n}\left\{\frac{\mathbb{P}(\boldsymbol{x}_i\,;\,\boldsymbol{\theta})}{p_{\boldsymbol{\theta}}}\right\}\\ &= \binom{N}{n} p_{\boldsymbol{\theta}}^{n}(1-p_{\boldsymbol{\theta}})^{N-n}\prod_{i=1}^{n}\mathbb{P}(\boldsymbol{x}_i\,|\,\text{animal } i \text{ is detected}\,;\,\boldsymbol{\theta})\,, \qquad (3)\end{aligned}$$

where $\mathbb{P}(\boldsymbol{x}_i\,|\,\text{animal } i \text{ is detected}\,;\,\boldsymbol{\theta})$ is the conditional probability of capture history $\boldsymbol{x}_i$, given detection.

The point here is that the observed capture histories are independent, identically distributed draws from the conditional distribution. The same is not true of the original product term,

$\prod_{i=1}^{n} \mathbb{P}(\boldsymbol{x}_i\,;\,\boldsymbol{\theta})$, because one value of the unconditional probability distribution is unobservable: the zero history $00\ldots0$. With the rewritten formulation, the product term now forms a ***regular, conditional likelihood,*** which we'll call $\mathcal{L}_C$:

$$\mathcal{L}\,(N,\boldsymbol{\theta}) = \binom{N}{n}\, p_{\boldsymbol{\theta}}^{n}\,(1-p_{\boldsymbol{\theta}})^{N-n}\,\mathcal{L}_C(\boldsymbol{\theta})\,, \tag{4}$$

where $\mathcal{L}_C(\boldsymbol{\theta}) = \prod_{i=1}^{n} \{\mathbb{P}(\boldsymbol{x}_i\,;\,\boldsymbol{\theta})/p_{\boldsymbol{\theta}}\}$ is a regular likelihood involving the capture parameters $\boldsymbol{\theta}$ only: it does not involve the pesky non-regular parameter N.

Aside: *Format (3) is a special case of the general formulation*

$$\mathcal{L}\,(N,\boldsymbol{\theta}) = \binom{N}{n}\, p_{\boldsymbol{\theta}}^{n}\,(1-p_{\boldsymbol{\theta}})^{N-n}\prod_{i=1}^{n} f(\boldsymbol{x}_i\,|\,\textit{animal } i \textit{ is detected}\,;\,\boldsymbol{\theta})\,,$$

where $f(\cdot\,;\,\boldsymbol{\theta})$ is any probability density without any missing bits. This formulation encapsulates a range of other models beyond capture-recapture, including distance-sampling models. Fewster and Jupp (2009) call models of this formulation ***binomial detectability models.*** *The same theoretical considerations apply to all models of this structure. The results in F&J2009 generalize those of Sanathanan (1972), which is usually cited for the asymptotic equivalence of $\widehat{N}$ and $\widehat{N}_c$ in a capture-recapture context.*

Because $\mathcal{L}_C$ is a regular likelihood, all the usual maximum likelihood theory can be applied. In principle, we could treat $\mathcal{L}_C$ as 'the' likelihood, and maximize it to find MLEs of $\boldsymbol{\theta}$, which should be well-behaved with asymptotic normal distributions, and then use the resulting estimates to estimate N (see below).

However, while the regularity of $\mathcal{L}_C$ is an advantage, there is a corresponding disadvantage: $\mathcal{L}_C$ is not the full and correct likelihood — at least, not if we assume that the binomial formulation of (2) is correct. When we maximize $\mathcal{L}_C$, we miss out components of the full likelihood $\mathcal{L}$ that involve the parameters $\boldsymbol{\theta}$, as we can see from (4); so $\mathcal{L}_C$ does not in itself encompass all the available information about $\boldsymbol{\theta}$.

The conditional formulation can be used, as we describe next, but it does require an awareness that it is not equivalent to the full likelihood $\mathcal{L}$, despite numerous assertions in the literature that it is. They are *asymptotically* equivalent, but this does not ensure that their finite-sample properties will be similar: sometimes they are, but other times they are not. As we shall see, in poorly-informed models the conditional formulation can go badly wrong: while the full likelihood might perform badly, the conditional likelihood can perform much worse. However, in some contexts the conditional formulation might be the more reasonable choice, for example if we don't trust the binomial formulation (as in distance sampling). In other contexts the conditional formulation is the only viable choice, for example if we are using individual covariates to inform the recapture process.

Practical guide to model-fitting using the conditional likelihood

1. Parametrize the capture history probabilities $\mathbb{P}(\boldsymbol{x}_i\,;\,\boldsymbol{\theta})$ in terms of the parameters $\boldsymbol{\theta}$ as before.

2. Minimize $-\log\left\{\mathcal{L}_C(\boldsymbol{\theta})\right\}$ with respect to the parameters $\boldsymbol{\theta}$:

$$\mathcal{L}_C(\boldsymbol{\theta}) = \prod_{i=1}^{n} \left\{ \frac{\mathbb{P}(\boldsymbol{x}_i\,;\,\boldsymbol{\theta})}{p_{\boldsymbol{\theta}}} \right\} .$$

   Let $\widehat{\boldsymbol{\theta}}_c$ be the MLEs resulting from maximizing $\mathcal{L}_C(\boldsymbol{\theta})$. Parameter N is not involved in this step.

3. Find estimated variances of $\widehat{\boldsymbol{\theta}}_c$ by inverting the hessian matrix of $\mathcal{L}_C$ at the maximum, as for a standard regular likelihood.

4. To estimate parameter N in the conditional formulation, the simplest case is to use

$$\widehat{N}_c = \frac{n}{p_{\widehat{\boldsymbol{\theta}}_c}} .$$

   That is,

$$\widehat{N}_c = \frac{n}{\text{overall estimated probability of capture based on } \widehat{\boldsymbol{\theta}}_c} .$$

   This expression can be justified in a few different ways:

   - $\widehat{N}_c$ is the method-of-moments estimator from the binomial formulation $n \sim \text{Binomial}(N, p_{\boldsymbol{\theta}})$:

$$\mathbb{E}(n) = N p_{\boldsymbol{\theta}} \quad \Rightarrow \quad \widehat{N}_c = \frac{n}{p_{\widehat{\boldsymbol{\theta}}_c}} .$$

   - A two-stage likelihood process. The full likelihood is:

$$\mathcal{L}\left(N, \boldsymbol{\theta}\right) = \binom{N}{n} p_{\boldsymbol{\theta}}^{n} \left(1 - p_{\boldsymbol{\theta}}\right)^{N-n} \mathcal{L}_C(\boldsymbol{\theta}) = \mathcal{L}_B(N, \boldsymbol{\theta})\, \mathcal{L}_C(\boldsymbol{\theta}) .$$

     First find $\widehat{\boldsymbol{\theta}}_c$ by maximizing the $\mathcal{L}_C(\boldsymbol{\theta})$ component with respect to $\boldsymbol{\theta}$ and use it to find $p_{\widehat{\boldsymbol{\theta}}_c}$. Then plug $p_{\widehat{\boldsymbol{\theta}}_c}$ into the component $\mathcal{L}_B(N, \boldsymbol{\theta}) = \binom{N}{n} p_{\boldsymbol{\theta}}^{n} \left(1 - p_{\boldsymbol{\theta}}\right)^{N-n}$, and maximize the $\mathcal{L}_B$ component with respect to N to gain $\widehat{N}_c$. The resulting estimator $\widehat{N}_c = n/p_{\widehat{\boldsymbol{\theta}}_c}$ is correct in the sense that the integer part of $\widehat{N}_c$ is the integer value of N that maximizes $\mathcal{L}_B$. (There is a small offset for continuous N: $\widehat{N}_c$ is actually higher than the value that maximizes $\mathcal{L}_B$ in $\mathbb{R}$, by up to about 0.5 of an animal.)

   - $\widehat{N}_c$ is often described as a ***Horvitz-Thompson-like estimator.*** This way of thinking is the most useful for generalizing to more complicated models. Horvitz-Thompson estimators come from survey sampling theory: the idea is to adjust for cases where some units in a population are sampled with lower probability than others. The estimators used in capture-recapture and distance sampling are called Horvitz-Thompson-*like* because we don't *know* what the probability of sampling each unit (animal) is: we have estimated it. In a nutshell, the idea is that for every animal that you captured with probability $p_{\boldsymbol{\theta}}$, there were actually $1/p_{\boldsymbol{\theta}}$ animals out there in the population. For example, for every animal you captured with probability 1/3, there were three in the population. Thus, for every n animals you captured with probability $p_{\boldsymbol{\theta}}$ each, there were $n/p_{\boldsymbol{\theta}}$ in the population: so $\widehat{N}_c = n/p_{\widehat{\boldsymbol{\theta}}_c}$. The reason why this way of thinking is useful is because we can extend it to the case where we estimate *different* capture probabilities for different animals. For example, if animal i has covariates $\boldsymbol{z_i}$, and we estimate a specific capture probability $p(\boldsymbol{z_i})$ for this covariate combination, then for every animal captured with probability $\boldsymbol{z_i}$ we estimate that there were $1/p(\boldsymbol{z_i})$ animals out there in the population with those same covariates. This leads to the more complicated Horvitz-Thompson-like estimator of $\widehat{N}_{HT} = \sum_i 1/p(\boldsymbol{z_i})$.

5. We now need to estimate the variance of $\widehat{N}_c$. Typically, papers that use conditional estimation devote some effort to deriving the appropriate variance estimator, because $\widehat{N}_c$ is not encompassed by the hessian matrix gained from maximizing $\mathcal{L}_C$. Reading between the lines in equations (13) to (17) of Fewster and Jupp (2009), the following estimator can be gleaned and is reasonably straightforward to obtain:

$$\widehat{\text{var}}\left(\widehat{N}_c\right) = \exp\left\{2\log\widehat{N}_c + \texttt{varlog}\right\}\left\{\exp\left(\texttt{varlog}\right) - 1\right\}$$

$$\text{where}\quad \texttt{varlog} = \widehat{\text{var}}\left(\log\widehat{N}_c\right) = \hat{\boldsymbol{q}}^T\,\widehat{V}_c\,\hat{\boldsymbol{q}} + \frac{(1-p_{\widehat{\boldsymbol{\theta}}_c})}{\widehat{N}_c\,p_{\widehat{\boldsymbol{\theta}}_c}}$$

$$\text{where}\quad \hat{\boldsymbol{q}} = \left.\frac{d\log p_{\boldsymbol{\theta}}}{d\boldsymbol{\theta}}\right|_{\widehat{\boldsymbol{\theta}}_c} = \left.\left(\frac{\partial\log p_{\boldsymbol{\theta}}}{\partial\theta_1}\,,\ \frac{\partial\log p_{\boldsymbol{\theta}}}{\partial\theta_2}\,,\ \ldots\,,\ \frac{\partial\log p_{\boldsymbol{\theta}}}{\partial\theta_p}\right)^T\right|_{\widehat{\boldsymbol{\theta}}_c}\ ;$$

where $\widehat{V}_c$ is the estimated variance matrix gained from inverting the hessian of the $\mathcal{L}_C$ maximization. The first line above is simply the variance of a lognormal random variable for which the mean of the log is taken to be $\log(\widehat{N}_c)$ and the variance of the log is `varlog`. The other two lines use equations (13) to (17) of F&J(2009) to infer the transformation from $\widehat{V}_c$ to $\widehat{\text{var}}\left(\log\widehat{N}_c\right)$.

6. Finally use the **lognormal confidence interval** for N (replacing 1.96 by the appropriate quantity if required):

$$(\widehat{N}/C,\ \widehat{N}C\,),$$

where

$$C = \exp\left[1.96\sqrt{\log\left\{1 + \frac{\widehat{\text{var}}(\widehat{N}_c)}{\widehat{N}^2}\right\}}\,\right]$$

and use **normal confidence intervals** for parameters $\boldsymbol{\theta}$:

$$\left(\widehat{\theta}_c - 1.96\sqrt{\widehat{\text{var}}(\widehat{\theta}_c)}\ ,\quad \widehat{\theta}_c + 1.96\sqrt{\widehat{\text{var}}(\widehat{\theta}_c)}\right)\ ,$$

where $\widehat{\text{var}}(\widehat{\theta}_c)$ is gained from the diagonal elements of $\widehat{V}_c$.

7. Test estimator properties by simulation as usual.

Why use the conditional method?

Yang and Chao (2005) suggest four reasons why the conditional approach $(\widehat{N}_c, \widehat{\boldsymbol{\theta}}_c)$ might be preferred over maximizing the full likelihood to gain $(\widehat{N}, \widehat{\boldsymbol{\theta}})$. Some of these reasons are more persuasive than others.

1. **Covariate models.** As described above, if capture probability depends upon individual covariates, the conditional approach enables us to estimate N without having to make assumptions about the distribution of the covariate values in the population at large (noting that we are unable to observe the covariates of those animals not captured).

   This is the most persuasive reason for using the conditional approach, in my opinion.

2. **The two approaches are asymptotically equivalent.**

   Yes, they are, but subsequent work has shown that their finite-sample properties can be very different, and not favourable towards the conditional model in instances examined (F&J2009).

3. **The conditional MLE can be thought of as a Horvitz-Thompson estimator.**

   True, but not clear why this is a particular advantage!

4. **The conditional MLE is scale-invariant, whereas the full MLE is not.** This means that, if every animal's capture history in the study were replaced by (say) 10 copies of itself, then the resulting estimate $\widehat{N}_c$ would be 10 times the estimate gained from the original data. This property applies to $\widehat{N}_c$ but not to the full-likelihood estimator $\widehat{N}$.

   Although it seems intuitive, it's not clear that scale-invariance is really an advantage. It could equally mean that if $\widehat{N}_c$ starts out bad in the initial sample, it doesn't get better in a sample size 10 times larger!

I shall add another possible reason for using the conditional MLE rather than the full MLE, although the precise details of this reason are hazy:

5. The conditional MLE might be more suitable if it is hard to justify binomial sampling of individuals, as assumed by the binomial term in the full likelihood.

   This is considered to be the case in distance sampling, in which the conditional approach is invariably used, because of the spatial layout of the population. For capture-recapture I am a little hazy about the potency of this argument: perhaps without the binomial sampling assumption the assumptions underpinning $\widehat{N}_c$ are invalid anyway?

Classification of closed capture-recapture models

Otis et al. (1978) proposed the following abbreviations for closed population capture-recapture models that remain the primary classification of model types today. The different models pertain to different influences on the *recapture* process.

Throughout, let $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ik})$ be the capture history of animal i at times $t = 1, \ldots, k$. Each $x_{it} = 0$ or 1 according to non-capture or capture of animal i at time t.

Model M_0

Model M_0 is the simplest capture-recapture model, and is too simplistic to be of much practical use. All animals have the same probability p of capture on all capture occasions. The parameter θ is therefore simply $\theta = p$. The probability of capture history $\boldsymbol{x}_i$ is:

$$\mathbb{P}(\boldsymbol{x}_i\,;\,\boldsymbol{\theta}) = \mathbb{P}(\boldsymbol{x}_i\,;\,p) = \prod_{t=1}^{k} p^{x_{it}}(1-p)^{1-x_{it}}\,.$$

For example, capture history $\boldsymbol{x}_i = 00101$ has probability $(1-p)(1-p)p(1-p)p$.

The overall probability of capture is $p_{\boldsymbol{\theta}} = 1 - (1-p)^k$.

Model M_t

Model M_t corresponds to Time-dependence on capture. All animals are assumed to have the same probability of capture at each capture occasion, but the probability can vary across capture occasions (i.e. vary over time). Model M_t is the simplest model likely to be useful in practice.

Under model M_t, all animals have probability p_t of being captured at time t, where $p_1, \ldots, p_k$ are estimated as free parameters. The capture parameter is $\boldsymbol{\theta} = (p_1, \ldots, p_k)$. The probability of capture history $\boldsymbol{x}_i$ is:

$$\mathbb{P}(\boldsymbol{x}_i\,;\,\boldsymbol{\theta}) = \prod_{t=1}^{k} p_t^{x_{it}}(1-p_t)^{1-x_{it}}\,.$$

For example, capture history $\boldsymbol{x}_i = 00101$ has probability $(1-p_1)(1-p_2)p_3(1-p_4)p_5$.

The overall probability of capture is

$$p_{\boldsymbol{\theta}} = 1 - \prod_{t=1}^{k}(1-p_t).$$

If the assumptions of model M_t can be met, it is a very good model to use. It is well-informed by the data and doesn't suffer from hidden pitfalls (as far as I am aware). For situations where sampling is expensive or difficult and sample sizes are small, model M_t could be the only model that the data can adequately support. To reduce the number of parameters, and therefore improve precision, some of the capture probabilities p_t are often constrained to be equal, perhaps based on the amount of sampling effort on the different occasions, or perhaps by trying several possibilities and using AIC or AICc to select the most parsimonious.

Model M_b

Model M_b corresponds to Behavioural response to capture. The idea is that the experience of being captured on a previous occasion might influence the animal's catchability on future occasions: in other words, animals become *trap-happy* or *trap-shy*.

In practice, time-dependence in capture is very likely, so the influence of behaviour is likely to be considered alongside time effects. We use notation M_{tb} to describe the model in which the probability of initial capture varies over time, and the catchability of each animal changes after its first capture relative to animals that have not yet been caught.

In the extreme, model M_{tb} would include a different parameter for first capture at every sampling occasion, $p_1, \ldots, p_k$, and for recapture at every sampling occasion past the first, $c_2, \ldots, c_k$. The capture parameter is $\boldsymbol{\theta} = (p_1, \ldots, p_k, c_2, \ldots, c_k)$. The probability of a capture history $\boldsymbol{x}_i = 00101$ would be:

$$\mathbb{P}(00101) = (1 - p_1)(1 - p_2)p_3(1 - c_4)c_5.$$

However, this saturated model is non-identifiable without further constraints on the parameters (Otis et al., 1978). This is known because the minimal sufficient statistic has dimension $2t - 1$, but the number of parameters in the saturated model is $2t$, including parameter N.

Chao et al. (2000) use a conditional MLE approach to fit the following submodel of M_{tb}. (They justify the choice of conditional approach on the basis of scale-invariance.) The behavioural response is encapsulated by a single parameter ϕ that multiplies the initial-capture probability at each time. That is,

$$c_t = \phi p_t \quad \text{for } t = 2, \ldots, k.$$

The case $\phi > 1$ represents a trap-happy response, and $\phi < 1$ represents trap-shy response.

The capture parameter becomes $\boldsymbol{\theta} = (p_1, \ldots, p_k, \phi)$, and the probability of capture history $\boldsymbol{x}_i = 00101$ is:

$$\mathbb{P}(00101) = (1 - p_1)(1 - p_2)p_3(1 - \phi p_4)\phi p_5.$$

The overall probability of capture in Model M_{tb} relates to the first capture of each animal only, so it doesn't involve ϕ:

$$p_{\boldsymbol{\theta}} = 1 - \prod_{t=1}^{k}(1 - p_t).$$

Even though model M_{tb} only contains one more parameter than model M_t, its practical performance can be very poor. The problem is that a low recapture rate could be due to either high initial capture probabilities and a low ϕ, or to low initial capture probabilities. The addition of parameter ϕ therefore introduces substantial variance to the initial capture probabilities $p_1, \ldots, p_k$. Moreover, the overall capture probability $p_{\boldsymbol{\theta}}$ *only* involves $p_1, \ldots, p_k$, so it is not counterbalanced by ϕ in any way, as ϕ oscillates from large (forcing $p_1, \ldots, p_k$ to be small) to small (forcing $p_1, \ldots, p_k$ to be large). This induces high variance in the critical parameter $p_{\boldsymbol{\theta}}$, and the problem is worse for the conditional MLE approach which allows $p_{\boldsymbol{\theta}}$ to be estimated without being tempered by N. Overall this can result in extremely high variance in $\widehat{N}_c$. The full likelihood approach can also give very high variance in $\widehat{N}$, but it is not quite as severe as the conditional likelihood approach. More details are in the powerpoint `Conditional_and_Unconditional_Estimation.pptx` and in the R code file `Closed_Populations.R`.

My recommendation is to avoid model M_{tb} unless you have very juicy sample sizes, and always to test models by simulation when they rely on teasing apart subtle signals in the recapture data.

Model M_h

Model M_h refers to individual Heterogeneity in capture probability: that is, different animals in the population are intrinsically less likely to be captured than others. These models are the hardest to fit and assess. Unfortunately, individual heterogeneity is a real effect, especially for intelligent animals, and it does have a non-trivial impact on population size estimates if left unaccounted for.

If heterogeneity corresponds to a fixed, time-constant attribute of individuals (individuals are by nature capture-averse or capture-willing), failure to account for it causes abundance estimates to be negatively biased. We can see this by imagining that the population corresponds to two groups: one of individuals that are commonly captured, and the other of individuals that are never captured. The ensuing abundance estimate might be accurate as an assessment of how many 'catchable' animals there are in the population, but it will entirely disregard the uncatchable ones, and will therefore underestimate the true population size.

Other issues of heterogeneity correspond to changes through an individual's life stages: for example a single individual might be more catchable when it is reproducing than it is at other times. This case was considered by Carroll et al. (2013) for southern right whales on calving grounds, at which females are more catchable in their calving years than at other times. If the catchability of an animal changes over time, the blanket rule that 'heterogeneity causes negative bias' no longer applies: the bias can probably go in any direction.

The major problem with heterogeneity is that we know nothing about the animals we can't catch. The whole business of fitting and evaluating models becomes very difficult when we need to distinguish between non-capture due to individual variability, and non-capture due to us simply not being very good at capturing things.

McCrea and Morgan (2015) give a good summary of heterogeneity issues in their Section 3.5, including appropriate references. Briefly, the main approaches to dealing with heterogeneity are:

- Finite mixture models (Pledger, 2000): usually fitted with two mixture components. Animals are imagined to belong to one of two groups representing ‘more catchable’ and ‘less catchable’ animals, and the mixture probability is estimated alongside capture probabilities for each group.
- Infinite mixture models (e.g. Morgan and Ridout, 2008): here we imagine the capture probability p_i for animal i to be itself a random variable with (for example) a beta distribution, so the resulting distribution of the number of captures of animal i is a compound distribution such as the beta-binomial, for which the probabilities can be expressed in closed form. The likelihood is then maximized with respect to the parameters of the beta distribution. This can also be conceptualized as a random-effect model. Morgan and Ridout (2008) also combine this beta-binomial distribution in a two-point mixture with an ordinary binomial, which has the advantage of containing the single binomial model M_0, Pledger’s two-point binomial, and the pure beta-binomial model as special cases, and can help to temper the extremities of the individual models on troublesome data sets.
- Covariate models, in which capture probabilities are related to measurable individual covariates such as weight or sex. Here, a logistic formulation is typically used, following Huggins (1989):

  $$\mathbb{P}(x_{it} = 1 \mid \boldsymbol{z}_i) = \left\{1 + \exp(-\boldsymbol{\beta}^T \boldsymbol{z}_i)\right\}^{-1} ,$$

  where $\boldsymbol{z}_i$ is a vector of individual covariates of animal i, and the regression coefficients $\boldsymbol{\beta}$ are estimated. A good summary, plus a nice treatment of what to do when covariates are missing, is given by Lee et al (in press).

  Covariate models are almost always fitted using the conditional likelihood $\mathcal{L}_C$, with the Horvitz-Thompson-like estimator $\widehat{N}_{HT}$ to estimate N:

  $$\widehat{N}_{HT} = \sum_{i=1}^{n} \frac{1}{\widehat{p}(\boldsymbol{z}_i)} ,$$

  where $\widehat{p}(\boldsymbol{z}_i)$ is the overall probability that animal i was ever captured, and is gained from the conditional MLEs $\hat{\boldsymbol{\beta}}$.

  As far as I know, covariate models are the only way that the every-effect model M_{tbh} can be fitted. Without covariates, the different influences of behaviour, individual heterogeneity, and time become confounded.

All heterogeneity models are vulnerable to thorny issues. Covariate models suffer from the obvious problem of model misspecification, as well as the concern that the conditional likelihood formulation, and any covariate combinations that induce very low estimates of $\widehat{p}(\boldsymbol{z}_i)$, could create very high variance in estimating N. Mixture models have various issues of parameter identifiability, and it can be difficult to get likelihoods to converge. Although I have not experimented with any of these models directly, I would guess that they perform well in data-rich settings where data are drawn from the model to be fitted, but it could be a challenge to apply them confidently to real populations. Having said this, they seek to address an important, non-ignorable problem, and it is not clear how else the problem might be addressed.

Note: I've focused on the classical modelling framework throughout. There is considerable potential for Bayesian models, and if there is a reason to use informative priors this can help to mitigate the problems of weak identifiability that plague complicated capture-recapture models. For a brilliant application of the Dirichlet process prior to account for individual heterogeneity, see Ford (2013).

Open population capture-recapture

Once the principles underlying closed-population capture-recapture are understood, methods for open population models follow quite naturally. The primary points of difference are:

- Parameters of interest, which might or might not include population size, N. Traditionally, survival was the chief focus of open-population capture-recapture. Methods for simultaneous estimation of survival and population growth, with or without N, have gradually become more prominent. Extensions to models for immigration and emigration are natural, and might involve sampling multiple locations to investigate movement between them.

- If population size N is of interest, it has to be carefully formulated — because in an open population it is always changing.

- Time dependence in capture histories. An animal can only be captured at time t if it is alive at the time.

- Plethora of different fitting methods. The typical options in wide use seem a little haphazard for those of us with delicate statistical sensibilities. In particular, there seems to be an avoidance of N in widely-used open-population models: some papers seem to go to great analytical length to avoid N. To me, this seems to be either at odds with the modern computing power at our disposal, or just odd. The experiences we have just had with closed populations suggest that omitting N might be a risk rather than a benefit. This is a speculation that warrants more investigation.

The open-population models that seem to be in popular use are flavours of the following types:

- Cormack-Jolly-Seber models (CJS) for estimating survival only (Cormack, 1964; Jolly, 1965; Seber, 1965).

- POPAN models, which include estimation of survival and recruitment. The standard reference is Schwarz and Arnason (1996). I think POPAN models have largely supplanted the traditional Jolly-Seber models (Jolly, 1965; Seber, 1965), though the occasional JS model fit might still be found. These models include N via what is called a *superpopulation parameter,* N_s. The superpopulation comprises *all animals ever exposed to capture*, and this definition needs to be understood carefully.

- Pradel models (Pradel, 1996), which are essentially POPAN models without N. These use a rather clever trick to fit both survival and recruitment without involving N, and are well-used because (I think) they are the only models on offer in program MARK that allow recruitment to be modelled through the annual growth rate, λ. Although the trick is clever, I am not convinced it is necessary — see above — and it would be interesting to compare performance of the Pradel fits with POPAN-lambda fits as developed in the esteemed software CAPOW (Fewster et al., sometime soon).

- Robust design models, also called Pollock's robust design (Pollock, 1982; Williams et al., 2002). Here, sampling takes place in two phases: *primary samples* are the usual, large-scale capture occasions (for example, annual sampling), and the population is open between these periods. Additionally, there are *secondary samples* that are taken within each primary period, during which the population is (typically) assumed to be closed. For example, sampling might take place over a 20-day period every year, in which case each day could be considered a secondary sampling occasion, and each year could be considered a primary sampling occasion. Potentially, the secondary periods could add considerable information. I have not explored robust design models, but I imagine that they will perform well as long as the recapture rate is sufficiently high between the secondary samples within the primary samples: this might be asking quite a lot.

Cormack-Jolly-Seber (CJS) models

The CJS model investigates survival only. It conditions on the first capture of each animal, which means that there is no scope for including animals that are never caught, so there is no scope for estimating N. Additionally, by conditioning on the first capture of each animal, there is no scope for estimating the entry process by which new animals join the population, in other words no possibility of estimating recruitment. Instead, the conceptual model is that ***we*** are the ones releasing the animals into the population to 'see how long they last'.

We need a new parameter, ϕ_t, that describes the probability of survival from time t to time $t+1$:

$$\phi_t = \mathbb{P}(\text{animal survives to time } t+1 \,|\, \text{animal is alive at time } t)\,.$$

Note: Throughout, we refer to 'survival' to mean 'staying in the population', and 'death' to mean 'leaving the population', whether by true death or permanent emigration. We cannot distinguish between these two scenarios in the CJS model. The parameters ϕ_t are strictly referred to as ***apparent survival*** probabilities, rather than true survival probabilities, to emphasize the difference.

The capture history probabilities, $p_2, \ldots, p_k$, are defined as before. Note that there is no probability for time $t = 1$, because there are no recaptures at time 1: only first-captures, which are conditioned upon. The parameter vector is $\boldsymbol{\theta} = (\phi_1, \ldots, \phi_{k-1}, p_2, \ldots, p_k)$. However, the saturated model is not identifiable: there is no information by which we can separate ϕ_{k-1} and p_k. For other times t, we can separate ϕ_{t-1} and p_t because an animal undetected at time t might pop up at a time later than t, and this reveals it was still alive at time t but went undetected. Formally, we can tell that ϕ_{k-1} and p_k are confounded because the only possible capture records at time k for an animal previously captured are 1 (with probability $\phi_{k-1}p_k$) or 0 (with probability $1-\phi_{k-1}+\phi_{k-1}(1-p_k) = 1-\phi_{k-1}p_k$). So the two parameters ϕ_{k-1} and p_k only ever appear in the likelihood as the product $\phi_{k-1}p_k$, so only the product can be estimated.

This raises a couple of important points about open population capture-recapture models in general:

- ***Parameter identifiability*** is a big issue. Sometimes, it is quite obvious that some parameters will not be identifiable, for example ϕ_{k-1} and p_k above. However, sometimes confounding can be very subtle. The phenomenon of parameters that can't be estimated is often called ***parameter redundancy*** in a capture-recapture context, and identifying combinations of redundant parameters can sometimes be done using heavy-duty symbolic algebra to find the model rank. Most of the work on parameter redundancy in capture-recapture models comes from Byron Morgan and associates, and an introduction to the topic is in Chapter 10 of McCrea and Morgan (2015).

  Note: There is some subtle difference in terminology between parameter redundancy and non-identifiability, but I am not sure what it is.

- The second important point is that ***identifiability depends upon the data.*** A combination of parameters that is identifiable in principle might not be identifiable for a particular data-set, because there just happen to be no observations in the data by which those parameters can be separated. This means that sometimes a model will unexpectedly fail on a particular data set: again, it can be difficult to identify what the precise 'problem' with the data or the model is.

- In practice, these problems are usually dealt with by fitting parsimonious models (e.g. constraining $\phi_1 = \phi_2 = \ldots = \phi_{k-1}$) and looking out for parameters that have given boundary estimates (0 or 1 for probabilities) or seem to be estimated with unusually high variance. Such parameters can be constrained equal to another parameter, in particular one that seems to have been estimated without difficulty.

Capture history probabilities in the CJS model are straightfoward to obtain, but they require a recurrence relation to deal with final 0s. The recurrence relation is simply a partitioning over the multiple possibilities of non-survival, or survival but non-capture, from the last sighting onwards.

Consider the capture history below:

$$\boldsymbol{x}_i = (\ 0\quad 0\quad 1\quad 0\quad 1\quad 1\quad 0\quad 0\quad 0\ )$$

Remembering that we ***condition*** on the first capture at time 3, the capture history probability does not involve any of ϕ_1, ϕ_2, p_1, p_2, or p_3. We obtain:

$$\mathbb{P}(\boldsymbol{x}_i) = \phi_3\,(1-p_4)\,\phi_4\,p_5\,\phi_5\,p_6\,\mathbb{P}(\text{never seen again after occasion 6}\,|\,\text{alive at occasion 6});$$

Now let

$$\chi_t = \mathbb{P}(\text{never seen again after occasion t}\,|\,\text{alive at occasion t})\,.$$

Clearly, $\chi_k = 1$: animals are never seen after occasion k, because k is the last capture occasion. Note that the χ_t parameters are cleverly defined to condition on being *alive* at time t, not on being *seen* at time t. This is what enables us to establish the recurrence relation. Moving on, the probability an animal is never seen again after time $k-1$, given that it is alive at $k-1$, is:

$$\chi_{k-1} = 1 - \phi_{k-1} + \phi_{k-1}(1 - p_k),$$

to account for the two possibilities that the animal failed to survive from $k-1$ to k, or that it did survive from $k-1$ to k but then was not seen at time k.

The pattern is established with the next record:

$$\chi_{k-2} = 1 - \phi_{k-2} + \phi_{k-2}(1 - p_{k-1})\chi_{k-1},$$

because the possibilities for an animal never being seen again after time $k-2$, despite being alive at time $k-2$, are:

- failed to survive from time $k-2$ to time $k-1$: probability $1 - \phi_{k-2}$;
- survived from time $k-2$ to time $k-1$, was not seen at time $k-1$, and was never seen again after time $k-1$: probability $\phi_{k-2}(1 - p_{k-1})\chi_{k-1}$.

In general,

$$\chi_t = 1 - \phi_t + \phi_t(1 - p_{t+1})\chi_{t+1},$$

with boundary condition $\chi_k = 1$. This recursion can easily be programmed into the likelihood function to find the χ_t parameters for all times $t = 1, \ldots, k$. Thus our final expression for the probability of capture history

$$\boldsymbol{x}_i = (\begin{array}{ccccccccc} 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \end{array})$$

is:

$$\mathbb{P}(\boldsymbol{x}_i) = \phi_3\,(1 - p_4)\,\phi_4\,p_5\,\phi_5\,p_6\,\chi_6.$$

Survey gaps are readily incorporated into this model. If annual survival is modelled as a constant, in other words $\mathbb{P}(\text{ survive to year } r+1 \mid \text{alive in year } r\ ) = \phi$ for all r, then the probability of surviving a gap of g years is ϕ^g instead of ϕ. Capture history probabilities, including the computation of the χ_t, are adjusted. For example, if there is a gap of g years from capture occasion t to capture occasion $t+1$, we have:

$$\chi_t = 1 - \phi^g + \phi^g(1 - p_{t+1})\chi_{t+1}.$$

Aside: *My understanding is that it can be quite tricky to ensure that MARK is doing the right thing when it comes to survey gaps. MARK does do the right thing, but the user often doesn't. Users need to take care when setting up their survey and formulation in MARK, but it seems to be tricky to get it right. This issue occurred with one of my collaborators and a survival model with differential survival at age 1: the apparent discrepancies between my R code and MARK results were finally resolved as being due to a model misspecification in MARK.*

The CJS model can be fitted simply by maximizing $\prod_{i=1}^{n} \mathbb{P}(\boldsymbol{x}_i \,|\, \boldsymbol{\theta})$ where the capture history probabilities $\mathbb{P}(\boldsymbol{x}_i)$ are given above. Closed-form estimators are also available: see Section 4.3.2 in McCrea and Morgan (2015). Variance estimates may be gained from the inverse hessian.

POPAN models

POPAN models derive from the superpopulation framework of Schwarz and Arnason (1996), which is itself an extension of Crosbie and Manly (1985). It is called the POPAN model because Arnason and Schwarz also developed software called POPAN for 'Population Analysis'. The POPAN-style model can be fitted in MARK.

(I'm unclear whether the POPAN software is accessed by MARK, or whether it has been rewritten for MARK, but the end result is the same.)

The big step forward here is that POPAN models include both abundance and recruitment. They are treated cleverly by the introduction of the idea of a *superpopulation*, which has size N_s and contains *all animals ever exposed to capture.* Note that this means that animals that are born and died in a survey gap are never exposed to capture, so they are not included in the superpopulation. This makes the notion of a superpopulation quite hard to link to a biological reality — not least because the superpopulation size depends upon our own choice of survey times — but from a statistical perspective it creates a parsimonious fitting framework. Parameters that do have proper biological interpretations, such as the number alive at any survey time, can then be derived from the fitted parameters including N_s.

One particularly useful property of the superpopulation formulation is that it establishes a full likelihood framework, and furthermore the likelihood has the same structure as equation (2). (As Fewster and Jupp (2009) would say, it is a *binomial detectability model.*) This means that all the theory we examined earlier applies directly: for example, the asymptotic lognormality of $\widehat{N}_s$, the asymptotic normality of $\widehat{\boldsymbol{\theta}}$, and the relationship between conditional likelihood estimators and full likelihood estimators.

In its saturated version, the POPAN model includes the following parameters:

- Capture probabilities $p_1, \ldots, p_k$ as usual. Probability p_t is the conditional probability that an animal is captured at time t, given that it is alive at time t.
- Survival probabilities $\phi_1, \ldots, \phi_{k-1}$, as for the CJS model. Probability ϕ_t is the conditional probability that an animal survives to time $t+1$, given that it is alive at time t.
- The superpopulation parameter, N_s: the number of animals ever exposed to capture by being alive and in the population at a survey time.
- A new set of parameters that describe the probability of animals *entering* the population for the first time at each time $t = 1, 2, \ldots, k$. These parameters are inconveniently known as $p_{ent(1)}, \ldots, p_{ent(k)}$, where p_{ent} is described as 'probability of entry'. I don't like this terminology because it sounds too much like 'probability of survival'. However,

while survival probabilities $\phi_1, \ldots, \phi_{k-1}$ are represented by free parameters that need not bear any relationship to each other, the $p_{ent(1)}, \ldots, p_{ent(k)}$ parameters by contrast measure proportions of the same thing and therefore must sum to 1.

Explicitly, each of the N_s animals in the superpopulation is first exposed to capture at one of the surveys $1, 2, \ldots, k$. We can think of the animals lining up one by one to have their 'survey of entry' decided. For each animal i, we select one of the surveys $1, 2, \ldots, k$ to be the survey of entry for animal i, with probabilities $p_{ent(1)}, p_{ent(2)}, \ldots, p_{ent(k)}$ respectively. Thus, $p_{ent(t)}$ is the probability that survey t is selected to be the survey of entry for each animal, and the p_{ent} parameters must satisfy

$$p_{ent(1)} + p_{ent(2)} + \ldots + p_{ent(k)} = 1.$$

An improved terminology, in my view, is to say that the p_{ent} probabilities are *entry proportions* rather than *probability of entry.*

As with the CJS model, the saturated POPAN model is not identifiable. The final parameters ϕ_{k-1} and p_k are confounded, as before, because we can't distinguish non-capture from non-survival at the end of the study. Similarly, the same effect happens at the other end: we can't distinguish non-capture from non-birth at the start of the study, so $p_{ent(1)}$ and p_1 are also confounded. It is likely that there will be further issues of identifiability for any given data set, so a more parsimonious model is preferable, for example with $\phi_1 = \ldots = \phi_k = \phi$.

Care needs to be taken with the $p_{ent(t)}$ parameters: for example, if there are uneven gaps between surveys then it is not reasonable to assume that the $p_{ent(t)}$ parameters should be constant, because in a real population new animals (births) will accrue over the gaps. The $p_{ent(1)}$ parameter is likely to be much larger than others, because all the animals alive at the first survey are first exposed to capture at that time, whereas the other $p_{ent(t)}$ parameters only describe new births since the time of the first survey.

Capture history probabilities are calculated as for the CJS model, but this time we need a recurrence relation at both ends of the observed captures, to describe the various possibilities of not-yet-birth and birth-but-no-capture for a string of zeros at the start of the study.

For example, the probability of capture history

$$\boldsymbol{x}_i = (\ 0 \ \ 0 \ \ 1 \ \ 0 \ \ 1 \ \ 1 \ \ 0 \ \ 0 \ \ 0\ )$$

is:

$$\mathbb{P}(\boldsymbol{x}_i) = \Big\{\, p_{ent(1)}(1-p_1)\phi_1(1-p_2)\phi_2 + p_{ent(2)}(1-p_2)\phi_2 + p_{ent(3)} \Big\}\, p_3\phi_3\,(1-p_4)\,\phi_4\, p_5\,\phi_5\, p_6\, \chi_6\,,$$

where χ_6 follows the same recurrence relation as for the CJS model:

$$\chi_t = 1 - \phi_t + \phi_t(1-p_{t+1})\chi_{t+1}\,; \qquad \chi_k = 1.$$

In practice, we calculate the term in braces by setting up a similar recurrence relation to that used for χ_t.

Finally, the parameters are N_s and $\boldsymbol{\theta} = (\phi_1, \ldots, \phi_{k-1}, p_{ent(2)}, \ldots, p_{ent(k)}, p_1, \ldots, p_k)$, or rather a more parsimonious version of the latter. (Note that the p_{ent} parameters sum to 1, so there are only $k-1$ free parameters in the saturated model.)

The full likelihood is exactly analogous to (2), but with the superpopulation size N_s replacing N, and with the relatively complicated capture history probabilities described above in $\mathbb{P}(\boldsymbol{x}_i\,;\,\boldsymbol{\theta})$:

$$\mathcal{L}\left(N_s, \boldsymbol{\theta}\right) = \binom{N_s}{n} p_{\boldsymbol{\theta}}^{n} \left(1 - p_{\boldsymbol{\theta}}\right)^{N_s - n} \prod_{i=1}^{n} \left\{ \frac{\mathbb{P}(\boldsymbol{x}_i\,;\,\boldsymbol{\theta})}{p_{\boldsymbol{\theta}}} \right\} .$$

The probability that an animal is ever seen, $p_{\boldsymbol{\theta}}$, is most easily given by its complement: the probability that an animal is never seen:

$$1 - p_{\boldsymbol{\theta}} = p_{ent(1)} \left(1 - p_1\right) \chi_1 + p_{ent(2)} \left(1 - p_2\right) \chi_2 + \ldots + p_{ent(k)} \left(1 - p_k\right) \chi_k .$$

We can derive the *expected numbers alive* at times $1, 2, \ldots, k$ from the model parameters. Let $EN_1, \ldots, EN_k$ be the expected numbers of animals exposed to capture at times $1, 2, \ldots, k$. These are *expected* numbers in the sense that the model doesn't prescribe that they should be the exact numbers: for example, although proportion $p_{ent(1)}$ of the N_s animals enter the population at time 1 *on average,* the number need not be $N_s\, p_{ent(1)}$ exactly, and is in fact a Binomial$(N_s, p_{ent(1)})$ random variable. We can gain expected numbers by the following equations:

$$\begin{aligned} EN_1 &= N_s\, p_{ent(1)} \\ EN_t &= \phi_{t-1}\, EN_{t-1} + N_s\, p_{ent(t)} , \end{aligned}$$

because the expected numbers alive at time t comprise those surviving from time $t-1$, plus new entrants at time t.

If we wish to include the variability in $N_1, \ldots, N_k$, this amounts to incorporating demographic variability (variance in births and deaths) within the model. A boostrap approach is probably the best way to achieve this.

CAPOW software for POPAN models

The POPAN models available in MARK only model the p_{ent} parameters as free parameters. It is natural to wish to constrain them to a growth curve, so that we can estimate annual rate of population growth, λ. The definition of λ is:

$$\lambda = \frac{EN_{t+1}}{EN_t} ,$$

where this relationship is assumed to be constant for $t = 1, \ldots, k-1$ if we are estimating a single population growth rate over the period. If we are doing this, it also makes sense to constrain $\phi_t = \phi$ to be constant for $t = 1, \ldots, k-1$, as it would be a rather odd model to allow ϕ to vary over time but for the birth rate to exactly cancel out any variability and remain constant. The annual birth rate is readily shown to be $\lambda - \phi$.

The CAPOW software (Fewster et al., sometime soon) parametrizes the POPAN model in terms of λ, including adjustments that are needed for unequal survey gaps. This simply involves constraining the parameters $p_{ent(1)}, \ldots, p_{ent(k)}$ to their values along a smooth growth curve. See the powerpoint `Capow.pptx` for more details.

CAPOW maximizes the full likelihood of the POPAN model, with respect to parameters $N_s, \lambda, \phi, p_1, \ldots, p_k$. The saturated POPAN-$\lambda$ model is identifiable. CAPOW also allows fitting of the most general POPAN model with parameters N_s and $\boldsymbol{\theta} = (\phi_1, \ldots, \phi_{k-1}, p_{ent(2)}, \ldots, p_{ent(k)}, p_1, \ldots, p_k)$. When it suspects non-identifiability of a model, it does display a warning, but it allows you to plough on regardless to see what happens. The first appearence of the POPAN-λ model was Carroll et al. (2013), and the preliminary coding for CAPOW resulted from a humpback whale power analysis in Carroll et al. (2015).

Pradel models

As far as I'm aware, Pradel-based models (Pradel, 1996) are the only way that MARK users can currently fit open population models parametrized in terms of annual growth rate, λ, and survival probability ϕ. (This would be possible for POPAN models too, but MARK does not provide the λ-parametrization.) Pradel models might be better described as a fitting method. They rely on the observation that, just as the CJS model estimates survival by moving forward through a capture history having conditioned upon its start-point, one might equally reverse the capture history, condition on the last capture, and estimate something related to recruitment by moving backwards. Essentially, we condition on the animal being captured at its last known occasion, and 'let it go' backwards to see how long it 'survives', which is clearly related to when it was born. Pradel (1996) then derives an ingenious set of results that enable ϕ and λ parameters to be fitted together in the same model, conditional on the animal being captured at some point in the study. The simultaneous model does not estimate any sort of abundance, N, but it does use the idea of $(N_1, \ldots, N_k)$, and how they must relate to λ and ϕ, to derive the expressions that enable λ and ϕ to be fitted together. I suspect that the resulting fitting method is similar (or perhaps identical) to use of the conditional likelihood $\mathcal{L}_C$ based on the POPAN likelihood, but I haven't checked this out.

Pradel models are popular because MARK users want to estimate λ as well as ϕ, and the Pradel models are the only way of doing so. However, users also want to estimate N at the same time, and currently the only way they can do this with standard capture-recapture data is to use the Pradel model to estimate λ, then separately to use the POPAN model to estimate N_s with the p_{ent} parametrization. Alternatively, if their data are suitable for a robust design, they can estimate N and λ together through the Pradel-robust-design methods.

Although it is ingenious, I am sceptical that there is any advantage to using a Pradel model once the λ-POPAN model is available (via CAPOW!) But then, I am a natural sceptic, and a little biased.

Robust design

Pollock's robust design (Pollock, 1982; Williams et al., 2002) offers a way of getting the best of both open and closed population models, by conducting mini closed-population capture-recapture experiments within each of the primary survey sessions. For example, the population might be sampled once per year for 20 days, in which case the primary sampling periods are the annual occasions, and the secondary sampling periods are the 20 days within the year. The idea is that the secondary samples span a sufficiently short time that it is reasonable to assume that the population is closed, so population size can be estimated for each period separately, whereas the changes between primary periods reveal information about survival and population growth, perhaps modelled using a Pradel formulation (which is provided by MARK).

Robust design models typically include an extra layer of parameters for *temporary emigration.* This is effectively a way of dealing with a particular type of individual heterogeneity in capture probability: the issue that an animal might be wholly unavailable for sampling in one or more of the primary periods, because it is not in the population exposed to capture. It has emigrated, but there is the chance that it might come back for a later primary period. (If there is no chance it will come back, it would be classified as a non-survival rather than an emigrant.)

Suppose that there are 4 primary periods, and 10 secondary periods. A capture history for a single animal might look like this:

$$0100101101 \quad - \quad 0000000000 \quad - \quad 1011101010 \quad - \quad 0001111001$$

Given the intensity of captures in the first, third, and fourth periods, the complete non-capture in the second period is surprising, and suggests evidence that the animal was not available for capture in the second period. Consequently, new parameters describing *temporary emigration* are introduced to account for this.

Choices for modelling temporary emigration include:

- **Random emigration:** the probability that each animal is unavailable for capture in primary period t (temporarily emigrated) is γ_t;
- **Markovian emigration:** the probability that an animal is unavailable for capture in primary period t depends upon whether it was available for capture in primary period $t-1$: this involves two parameters, γ'_t and γ''_t, describing the probability of unavailability in period t conditional on being unavailable or available in period $t-1$, respectively.

There are also versions of the model involving behavioural response to capture: a robust-design version of model M_{tb}.

While I don't have direct experience with the robust design framework, it seems like a good idea, with its performance presumably being highly dependent upon having a suitably high recapture rate between the secondary sampling sessions, such that the all-zero capture history within a period can be easily distinguished from a typical capture history.

References

Carroll, E. L., Childerhouse, S. J., Fewster, R. M., Patenaude, N. J., Steel, D., Dunshea, G., Boren, L., and Baker, C. S. Accounting for female reproductive cycles in a superpopulation capture-recapture framework: application to southern right whales (*Eubalaena australis*). *Ecological Applications* 23, 1677-1690, 2013.

Carroll, E. L., Brooks, L., Baker, C. S., Burns, D., Garrigue, C., Hauser, N., Jackson, J. A., Poole, M. M., and Fewster, R. M. Assessing the design and power of capture-recapture studies to estimate demographic parameters for the Endangered Oceania humpback whale population. *Endangered Species Research* 28, 147-162, 2015.

Chao, A., Chu, W., and Hsu, C.-H. Capture-recapture when time and behavioral response affect capture probabilities. *Biometrics* 56, 427-433, 2000.

Cooch, E., and White, G. *Program Mark, a gentle introduction.*
Online at `http://www.phidot.org/software/mark/docs/book/`. Retrieved 2013.

Cormack, R. Estimates of survival from the sighting of marked animals. *Biometrika* 51, 429-438, 1964.

Crosbie, S. F., and Manly, B. F. Parsimonious modelling of capture-mark-recapture studies. *Biometrics* 41, 385-398, 1985.

Fewster, R. M., and Jupp, P. E. Inference on population size in binomial detectability models. *Biometrika* 96, 805-820, 2009.

Fewster, R. M., and Jupp, P. E. Information on parameters of interest decreases under transformations. *Journal of Multivariate Analysis* 120, 34-39, 2013.

Fewster, R. M., Oh, J., and Carroll, E. L. Capow - a system for capture-recapture power analysis and model exploration. *Soon to be submitted, Somwhere.*

Ford, J. *Incorporating individual heterogeneity into mark-recapture models.* PhD thesis, University of Tasmania, 2013.

Huggins, R. M. On the statistical analysis of capture experiments. *Biometrika* 76, 133-140, 1989.

Jolly, G. M. Explicit estimates from capture-recapture data with both death and immigration — Stochastic model. *Biometrika* 52, 225-247, 1965.

Lee, S.-M., Hwang, W.-H., and de Tapsoba, J. D. Estimation in closed capture-recapture models when covariates are missing at random. *Biometrics,* in press, 2016.

McCrea, R. S., and Morgan, B. J. T. *Analysis of capture-recapture data.* Chapman and Hall, 2015.

Morgan, B. J. T. and Ridout, M. S. A new mixture model for capture heterogeneity. *Applied Statistics* 57, 433-446, 2008.

Otis, D. L., Burnham, K. P., White, G. C. and Anderson, D. R. Statistical inference from capture data on closed animal populations. *Wildlife Monographs* 62, 1-135, 1978.

Pledger, S. Unified maximum likelihood estimates for closed capture-recapture models using mixtures. *Biometrics* 56, 434-442, 2000.

Pollock, K. H. A capture-recapture design robust to unequal probability of capture. *Journal of Wildlife Management* 46, 752-757, 1982.

Pradel, R. Utilization of capture-mark-recapture for the study of recruitment and population growth rate. *Biometrics* 52, 703-709, 1996.

Sanathanan, L. Estimating the size of a multinomial population. *Annals of Mathematical Statististics* 43, 142-152, 1972.

Schwarz, C. J., and Arnason, A. N. A general methodology for the analysis of capture-recapture experiments in open populations. *Biometrics* 52, 860-873, 1996.

Seber, G. A. F. A note on the multiple recapture census. *Biometrika* 52, 249-259, 1965.

Williams, B. K., Nichols, J. D., and Conroy, M. J. *Analysis and management of animal populations.* Academic Press, San Diego, 2002.

Yang, H.-C. and Chao, A. Modeling animals' behavioral response by Markov chain models for capture-recapture experiments. *Biometrics* 61, 1010-1017, 2005.