Estimating abundance in small populations using pedigree reconstruction

Sarah Croft[1], Richard Barker[1], Mik Black[2], Jamie Sanderlin[3] and Matt Schofield[1]

[1]Department of Mathematics and Statistics,
University of Otago

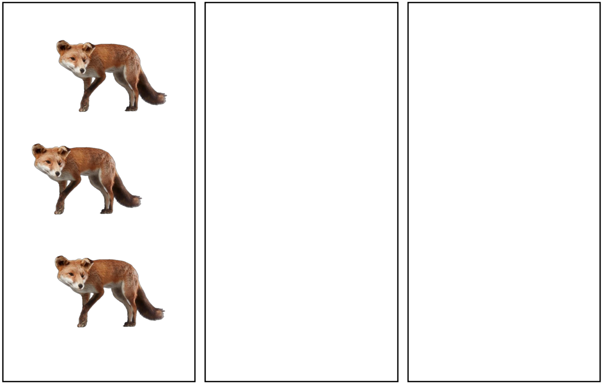
[2]Department of Biochemistry,
University of Otago

[3]Rocky Mountain Research Station,
USDA Forest Service

Biometrics in the Bush Capital, November 2025

Mark-Recapture

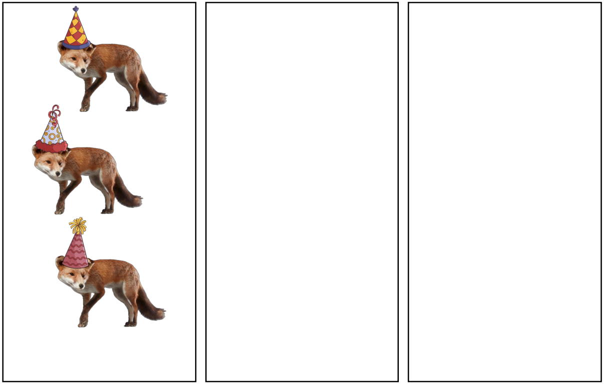
Suppose we have three time periods and capture the following foxes in each period:



[1] Image copyright: Museums Victoria

Mark-Recapture

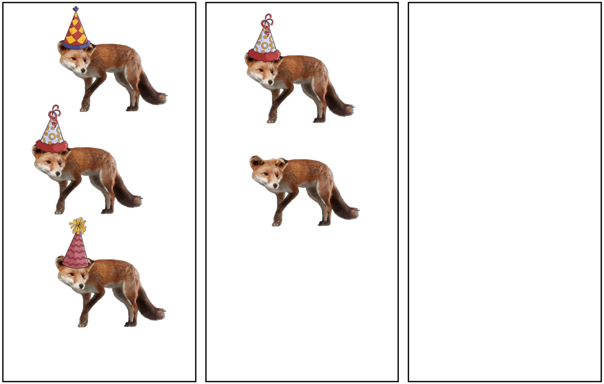
Suppose we have three time periods and capture the following foxes in each period:



[1] Image copyright: Museums Victoria

Mark-Recapture

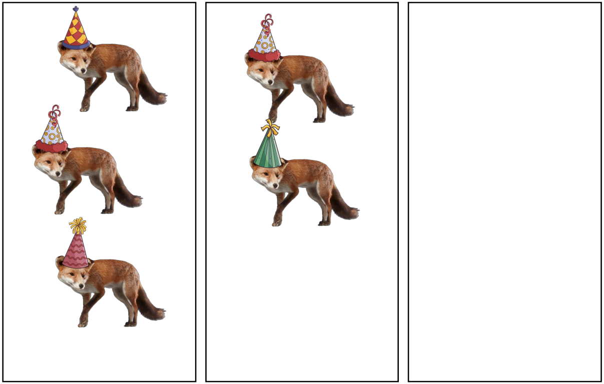
Suppose we have three time periods and capture the following foxes in each period:



[1] Image copyright: Museums Victoria

Mark-Recapture

Suppose we have three time periods and capture the following foxes in each period:



Mark-Recapture

Suppose we have three time periods and capture the following foxes in each period:

[1] Image copyright: Museums Victoria

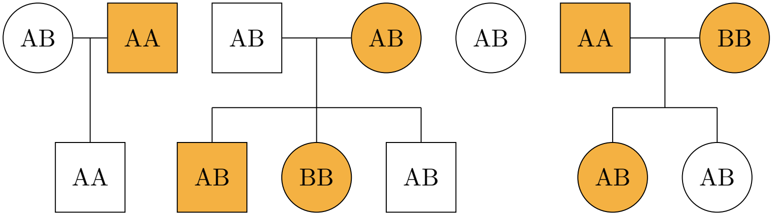
Close-Kin Mark-Recapture

- Bravington et al. (2016)[1] defined a pseudo-likelihood model, identifying the number of kinship pairs within the sample from the genetic data
- The model can be fit to dead recoveries without previous observation of the individual
- Rather than a physical tag a genetic sample is taken from each observed animal

[1] Bravington, M., Skaug, H. J., & Anderson, E. C. (2016). Close-kin mark-recapture. *Statistical Science*, *31*(2), 259–274. https://doi.org/http://doi.org/10.1214/16-STS552

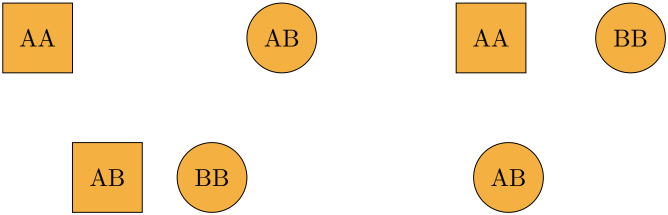
The population's pedigree

Consider two time periods, with a founder generation and one offspring period, and we observe the highlighted individuals.



The population's pedigree

Consider two time periods, with a founder generation and one offspring period, and we observe the highlighted individuals.



Model assumptions and restrictions

We assume:

1. The population is geographically closed
2. All individuals born before the study began are defined as founders and their parents are undefined
3. All individuals born during the study must have a female and male parent within the super population
4. Births are defined as individuals who were born within the period and survived until the start of the next time period

How would we simulate the population?

Let b_0 be the number of founders. We define

$$b_0 \sim \text{Pois}(\lambda_0) \quad (1)$$

$$s_i \sim \text{Bern}(\omega) \text{ for } i \in 1, ..., b_0 \quad (2)$$

where $s_i = 1$ if individual i is female.

How would we simulate the population?

Let b_0 be the number of founders. We define

$$b_0 \sim \text{Pois}(\lambda_0) \tag{1}$$
$$s_i \sim \text{Bern}(\omega) \text{ for } i \in 1, ..., b_0 \tag{2}$$

where $s_i = 1$ if individual i is female.
Using the properties of a Bernoulli marked Poisson we define b_{00} and b_{01} as the number of male and female founders respectively where

$$b_{00} \sim \text{Pois}(\lambda_0(1-\omega)) \tag{3}$$
$$b_{01} \sim \text{Pois}(\lambda_0\omega) \tag{4}$$

How would we simulate the population?

For time period 1 then:

1. Each female founder i mates with probability ζ and under random mating then

$$M_{1i} \sim \text{Cat}\left((1-\zeta, \frac{\zeta}{b_{00}}, ..., \frac{\zeta}{b_{00}})'\right) \tag{5}$$

   with sample space $\{0, 1, ..., b_{00}\}$.

How would we simulate the population?

For time period 1 then:

1. Each female founder i mates with probability ζ and under random mating then

$$M_{1i} \sim \text{Cat}\left((1-\zeta, \frac{\zeta}{b_{00}}, ..., \frac{\zeta}{b_{00}})' \right) \tag{5}$$

   with sample space $\{0, 1, ..., b_{00}\}$.

2. Let the number of offspring born to a mated female i, be independent Poisson random variables with parameter λ_1. Each of the offspring are female with probability ω then

$$b_{10} \sim \text{Pois}\left(\lambda_1 (1-\omega) \sum_{i=1}^{b_{01}} I(M_{1i} > 0) \right) \tag{6}$$

$$b_{11} \sim \text{Pois}\left(\lambda_1 \omega \sum_{i=1}^{b_{01}} I(M_{1i} > 0) \right) \tag{7}$$

How would we simulate the population?

For time period 1 then:

3. The mother of an offspring h is a Categorical random variable such that

$$r_h \sim \text{Cat}\left(I(\mathbf{M_1} > 0) \left(\frac{1}{\sum_{i=1}^{b_{01}} I(M_{1i} > 0)}, \ldots, \frac{1}{\sum_{i=1}^{b_{01}} I(M_{1i} > 0)} \right)' \right) \tag{8}$$

where $r_h = i$ if female i is the mother of h.

How would we simulate the population?

For time period 1 then:

3. The mother of an offspring h is a Categorical random variable such that

$$r_h \sim \text{Cat}\left(I(\mathbf{M_1} > 0) \left(\frac{1}{\sum_{i=1}^{b_{01}} I(M_{1i} > 0)}, \ldots, \frac{1}{\sum_{i=1}^{b_{01}} I(M_{1i} > 0)} \right)' \right) \tag{8}$$

   where $r_h = i$ if female i is the mother of h.

4. Each founder survives until the start of period 2 with probability ϕ_1.

How would we simulate the population?

For time period 1 then:

3. The mother of an offspring h is a Categorical random variable such that

$$r_h \sim \text{Cat}\left(I(\mathbf{M_1} > 0) \left(\frac{1}{\sum_{i=1}^{b_{01}} I(M_{1i} > 0)}, \ldots, \frac{1}{\sum_{i=1}^{b_{01}} I(M_{1i} > 0)} \right)' \right) \tag{8}$$

   where $r_h = i$ if female i is the mother of h.

4. Each founder survives until the start of period 2 with probability ϕ_1.
5. If a founder dies between the start of period 1 and 2 they are recovered with probability π_1.

How would we simulate the population?

We can generalise the model to K periods. For $j \in \{1, ..., K\}$:

- $M_{ji} \sim \text{Cat}\left(\left(1-\zeta, \frac{\zeta}{A_{j0.}}, ..., \frac{\zeta}{A_{j0.}}\right)'\right)$ if $A_{j1i} = 1$
- $b_{j0} \sim \text{Pois}\left(\lambda_j(1-\omega)\sum_{i=1}^{A_{.1}} I(M_{ji} > 0)\right)$
- $b_{j1} \sim \text{Pois}\left(\lambda_j \omega \sum_{i=1}^{A_{.1}} I(M_{ji} > 0)\right)$
- $r_h \sim \text{Cat}\left(I(\mathbf{M_j} > 0)\left(\frac{1}{\sum_{i=1}^{A_{.1}} I(M_{ji}>0)}, ..., \frac{1}{\sum_{i=1}^{A_{.1}} I(M_{ji}>0)}\right)'\right)$ for all h born in j.
- $d_i \sim \text{Cat}(\xi_t)$ where i is born in period t and ξ_t is a function of ϕ.
- $x_i \sim \text{Bern}(\pi_j)$ where i died in period j.

How would we simulate the population?

We model a founder's genotype at L independent loci by

$$G_{il} \sim \mathsf{Cat}(\boldsymbol{\gamma}_l),\ l \in \{1, ..., L\} \tag{9}$$

and a non-founder i born in j with mother r_i and father M_{jr_i} by

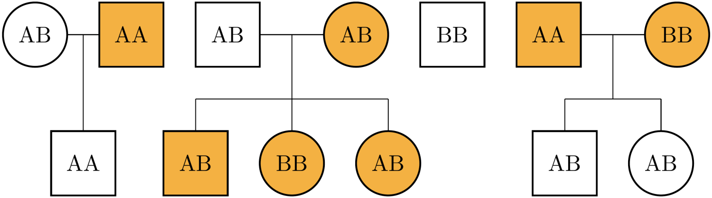
$$G_{il} \sim \mathsf{Cat}\left(\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)'\right) \tag{10}$$

with sample space

$$\begin{aligned}\mathcal{G}_{il} = \{&(G_{r_i 1}, G_{M_{jr_i} 1}), (G_{r_i 2}, G_{M_{jr_i} 1}), \\ &(G_{r_i 1}, G_{M_{jr_i} 2}), (G_{r_i 2}, G_{M_{jr_i} 2})\}\end{aligned} \tag{11}$$

Implementation

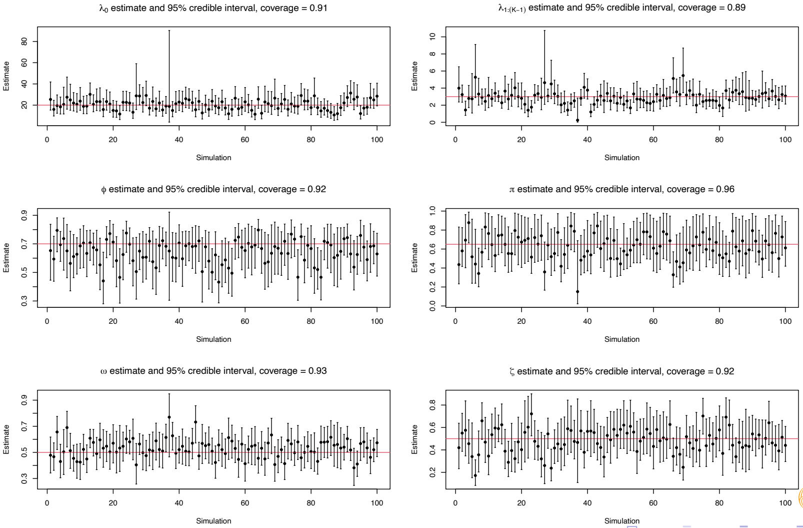
1. Update B_{js} for $j \in \{0, 1, ..., K-2\}$ and $s \in \{0, 1\}$ using a Metropolis-Hastings split-merge reversible jump
2. Update $B_{K-1,s}$ for $s \in \{0, 1\}$ using a Metropolis-Hastings reversible jump
3. Update all other variables using either Metropolis-Hastings reversible jump moves or Gibbs sampling



Simulation results

We simulated 100 pedigrees with true parameters values:

$$\theta = (K, \lambda_0, \lambda_{1:(K-1)}, \phi, \pi, \omega, \zeta) = (4, 20, 3, 0.7, 0.65, 0.5, 0.5) \quad (12)$$



Continuing work

- Test on real data
- Speed up algorithm:
  - Marginalising over unobserved branches of the pedigree
  - Joint updates for genotype
- Explore adjustments to the model required to allow for inbreeding

Thanks!

Thank-you to the NZSA for the NZSA Tidy International Travel Scholarship to attend this conference.