# Mark-recapture with identification errors

Richard Vale

12/12/13

# 'Missing' badgers: call for answers

By Helen Briggs
BBC News

Badger numbers are estimated by hair trapping and counting setts

**Conservationists are calling for an investigation into plummeting badger numbers in the run up to the cull.**

The apparent 50% decline over a year before the cull started appears to be unprecedented, data from other badger populations suggests.

Government officials have blamed the cold winter, disease or lack of food for the dwindling numbers.

But a wildlife charity claims illegal killing of badgers may behind the fall in numbers and is calling for answers.


Badger cull

Q&A: The badger cull

Is a badger cull the only answer?

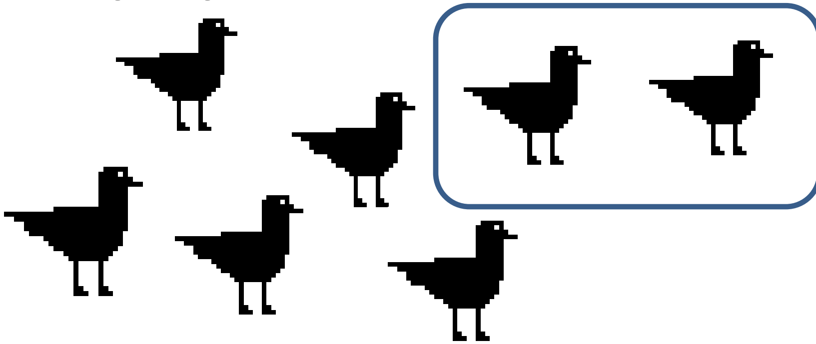
Badger cull v vaccines in TB fight

To cull or not to cull?

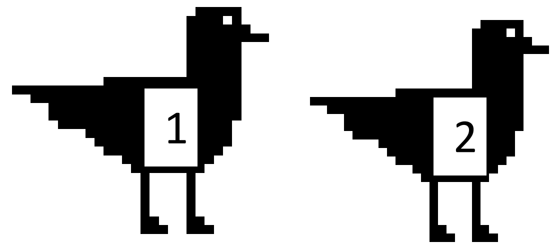

Hard to estimate the size of an animal population.

One popular method: mark-recapture sampling

# A population

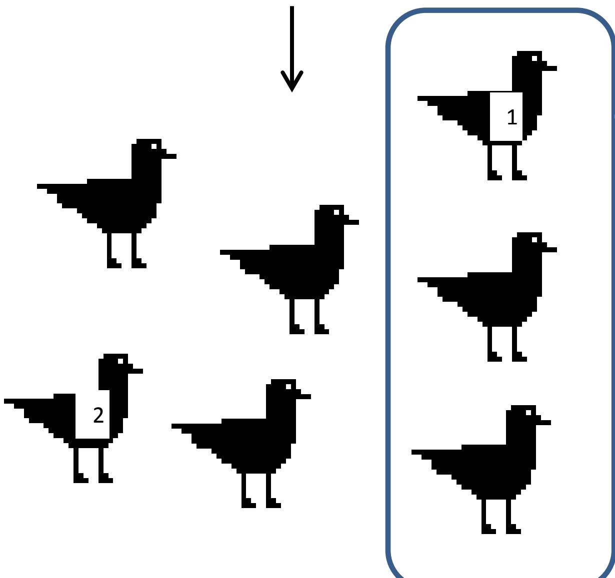


A sample, taken by a biologist




Sampled animals are identified with a mark.

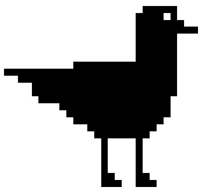
Sampled animals are returned to the population and mingle with the others.




Later, another sample is taken. Some of the marked animals may be recaptured, while others may not be.

The process of capturing animals, marking them and releasing them is repeated several times …

Here is how the data might look. Notice that some individuals were never captured.



| | | | |
|---|---|---|---|
| 1 | 1 | 1 | 0 |
| 2 | 1 | 0 | 1 |
| 3 | 0 | 1 | 0 |
| 4 | 0 | 1 | 0 |
| 5 | 0 | 0 | 1 |
| 6 | 0 | 0 | 1 |

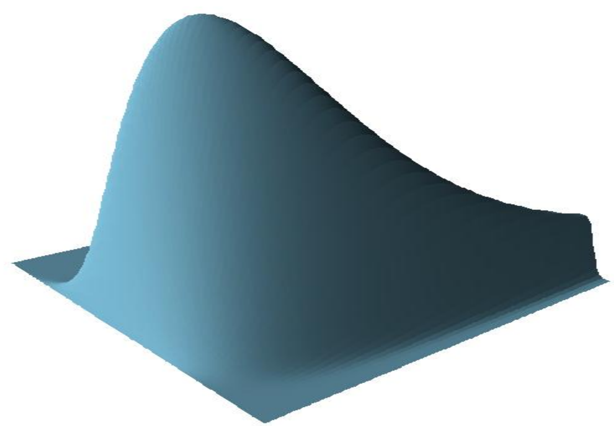
Individual (mark)

Indicators: 1 if the individual was captured on that particular occasion and 0 otherwise.

Aim: estimate the unknown population size $N$.

If captures are assumed to be independent Bernoulli trials with constant capture probability $p$, then we can write down the likelihood function and maximise it.

Maximum at $N = 8.0, \;\; p = 0.33$



A complication: it is difficult to label some kinds of animals.

Alternatives:

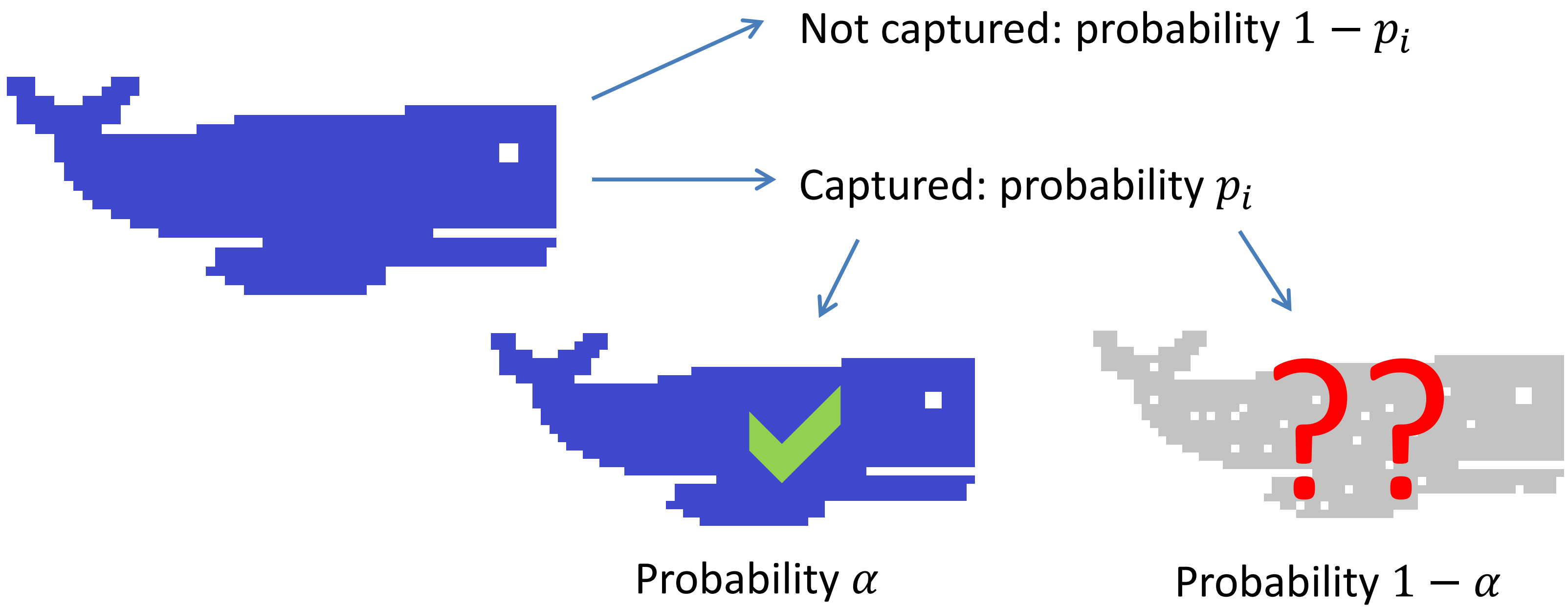
- Genetic samples (from hair, faeces etc.)
- Human observers
- Photographs

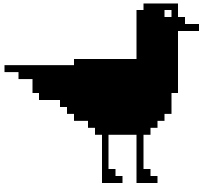
These methods might lead to misidentification errors.

# Model $M_{t,\alpha}$

This version of the mark-recapture model was invented by Lukacs/Burnham (2005) and Yoshizaki et al. (2011).

- probability $p_i$ of an animal being captured at time $i$.
- Captured animals are correctly identified with a fixed probability $\alpha$.
- A misidentified animal produces a ghost record which is seen only once.





| | | | |
|---|---|---|---|
| 1 | 1 | 1 | 0 |
| 2 | 1 | 0 | 1 |
| 3 | 0 | 1 | 0 |
| 4 | 0 | 1 | 0 |
| 5 | 0 | 0 | 1 |
| 6 | 0 | 0 | 1 |

Possible ghost records

Under model $M_{t,\alpha}$ , there could be as few as three individuals in the population.

Inference for model $M_{t,\alpha}$ is complicated by the fact that we don't know whether an animal which was seen only once is a real animal, or a ghost produced by the misidentification of some other animal.

However, Vale & Fewster (2012/13) were able to write down an expression for the likelihood function and found a way to compute it.

- Example (from previous slide) $\widehat{N} = 3, \widehat{\alpha} = 0.6$

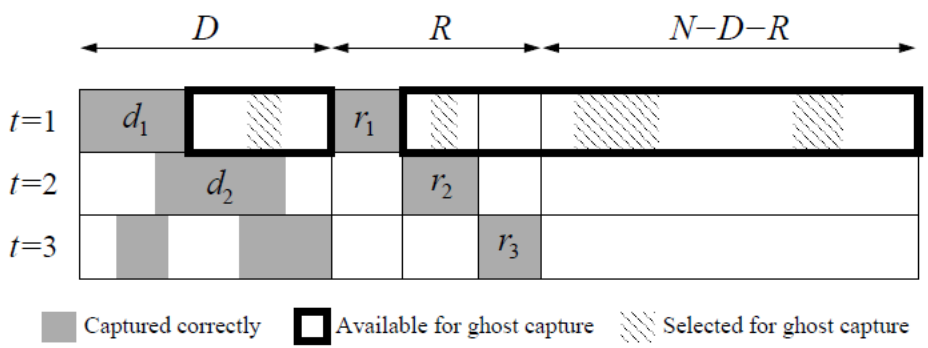



**Figure 1.** Diagram showing partition of animals into: $D$ possessing histories with duplicates, where duplicated records are known to be correct; $R$ assigned to have correct histories at exactly one timepoint, of which $r_t$ $(t = 1, \ldots, T)$ have their sole correct capture at time $t$; and the remaining $N - D - R$ which are either never caught or never caught correctly. At time $t = 1$, for a particular selection of $r_1$, there remain $N - (d_1 + r_1)$ animals available for ghost captures, of which $u_1 - r_1$ must be selected as ghosts.

Some combinatorics is involved

Direct combinatorial calculation gives the following likelihood, which is further explained in Figure 1.

$$\mathcal{L}(N, p_1, \ldots, p_T, \alpha\,;\, \boldsymbol{f}) = \left\{ \alpha^C \prod_{t=1}^{T} p_t^{n_t} (1 - p_t)^{N - n_t} \right\} \times \left\{ \sum_{r \in \mathcal{R}(\boldsymbol{f}, N)} \frac{N!\, \alpha^R (1 - \alpha)^{U - R}}{\left(\prod_{k\,:\,|\omega_k| \geqslant 2} f_k!\right) r_1! \ldots r_T!\, (N - D - R)!} \prod_{t=1}^{T} \binom{N - d_t - r_t}{u_t - r_t} \right\}. \qquad (2)$$

The model can already be fitted by Bayesian methods (Link et al. 2010), so what is the advantage of having a new expression for the likelihood? The main advantage is that the model can be fitted much more quickly, and simulation studies can be carried out on a large scale.

Example: whale capture study of Carroll et al. Nine sampling occasions, $\widehat{N} = 144$ whales.

95% confidence interval for $\widehat{N}$: $(49, 419)$. Conclusion: there are some whales.

Our simulation studies show that parameter estimates tend to be biased unless the sample sizes and capture probabilities are unrealistically large. Other authors have also had trouble applying the model to real data (or neglected real data altogether). It seems that without strongly informative priors, $M_{t,\alpha}$ gives very large error estimates.

Research directions:

- Pin down why the model fails by analysing simpler models in more detail
- Develop better approaches for photographic studies