

Webinar

Session 1. Introduction to Modern Methods for Analyzing Capture- Recapture Data: Closed Populations 1

by

Bryan F.J. Manly
Western Ecosystems Technology Inc.
Cheyenne, Wyoming
bmanly@west-inc.com

Summary

- ! What is in the full webinar
- ! The history of capture-recapture methods
- ! Maximum likelihood estimation.
- ! Model selection methods
- ! Introduction to closed population models
- ! Structure of the data
- ! Two sample situation
- ! Multi-sample situation
- ! The Otis *et al.* models
- ! Example of a likelihood calculation

Overview of Webinar

The webinar is based on the contents of *Handbook of Capture-Recapture Analysis*, (Princeton University Press, 2005) edited by Steve Amstrup, Trent McDonald and Bryan Manly. Individual chapters were written by the editors plus

Anne Chao
Richard Huggins
Kenneth Pollock
Russell Alpiraz-Jara
James Nichols
John Hoenig
William Hearn
Richard Barker
Carl Schwarz
and
Eric Regehr

! No attempt is made in the webinar to cover all of the topics in the book. There are ten chapters and the webinar is mainly based on the material in chapters 1 to 5.

! Material **not** covered includes

Tag-recovery models
Joint modeling of tag-recovery and live-resighting
data
and
Multistate models

See the book for more details about these topics.

Contents

Session 1	BM	What is in the webinar, history of capture-recapture studies, maximum likelihood methods, model selection methods, closed population models 1 (Handbook Chapters 1, 2 and 4).
Session 2	TMc	Closed population models 2 (Handbook Chapter 4).
Session 3	BM	Open populations 1 (Handbook Chapter 3).
Session 4	TMc	Open populations 2 (Handbook Chapter 5).

The material in the webinar was also covered in a one day workshop at the Wildlife Society meeting in Tucson, Arizona in 2007.

The Start of Capture-Recapture Studies

- ! In September 1802, **Pierre Simon Laplace** (1749-1827) used a capture-recapture type of approach to estimate the size of the human population of France.
- ! Live births were recorded for all of France on an annual basis and in the year before September 1802, Laplace estimated the number of such births to be approximately **one million**.
- ! These newly born individuals constituted a **marked** population.
- ! Laplace then obtained census and live birth data from several communities "with zealous and intelligent mayors" across all of France.
- ! He estimated that there were 71,866 births per year in those communities, with a total population of 2,037,615 individuals.

! So the proportion of "marked" individuals in the sampled communities was estimated as

$$p = 71866/2037615 = 0.0353.$$

! Assuming that this proportion of marked individuals applies for the whole of France with a population of N then gives

$$0.0353 \approx 1000000/N$$

so that

$$N \approx 1000000/0.0353 = 28328612.$$

! This estimation procedure is equivalent to the **Lincoln-Peterson estimator**.

! Laplace is commonly thought of as the first to use the capture-recapture idea, but he was preceded by almost 200 years by **John Graunt** in his attempts to use similar methods to estimate the effect of plague and the size of populations in England in the early 1600s.

! The theories and applications of capture-recapture have now moved far beyond the concepts of John Graunt and Pierre Laplace but still share the basic concept of ratios between known and unknown values that guided these pioneers.

Maximum Likelihood Methods

- ! Early methods for analyzing capture-recapture relied upon ad-hoc models for their justification but by the late 1960's the use of well-defined probability models with **maximum likelihood estimation** (MLE) of the unknown parameters had become the standard approach.
- ! MLE is known to produce estimates with good properties under a wide range of conditions. It consists of two steps.
- ! First, there is the construction of a model that states the probability of observing the data as a function of the unknown parameters that are of interest. This is called the **likelihood function**.
- ! Second, the estimates of the unknown parameters are chosen to be those values that make the likelihood function as large as possible, i.e., the values that **maximize the likelihood**.

- ! The simplest situation occurs when modeling the probability of observing the data from single independent animals and constructing the full likelihood as the product of the probabilities for all animals.

Example

- ! A four-sample experiment where n_1 animals are marked in the first sample, no more marking is done, and recapture data are obtained during samples 2, 3, and 4.
- ! Assume that the probability of an animal surviving from the time of the j th sample to the time of the next sample is s_j and the probability of a live animal being captured in the j th sample is p_j .
- ! Suppose that the capture-recapture pattern of an animal is

1011

where 1 indicates a capture and 0 no capture.

- ! The animal is seen in the first, third and fourth sample but not in the second sample.

- ! Assuming independence of capture and survival events the probability of the observed data (conditional on the first capture) is then

$$P(1011) = n_1 (1 - p_2)^{n_1 - 1} p_3^{n_1 - 1} p_4^{n_1 - 1}.$$

- ! The full likelihood for the data would be obtained by working out a probability like this for each of the n_1 animals seen in sample 1 and multiplying them together.
- ! There are often restrictions on what parameters can be estimated, so that in the example it is actually only possible to estimate the product $p_3 p_4$ rather than the individual parameters.
- ! Section 1.3 of the book discusses in more detail how to derive likelihood functions.
- ! In some cases once the likelihood function is defined it is possible to determine explicit values for the parameters that maximize the function (the **ML estimators**).

- ! Often this is not possible so that the likelihood function has to be maximized using a **numerical procedure**. Section 1.3 of the book gives a brief explanation of how this can be done, with an example.
- ! See also the spreadsheet M0-LIK.xls for an example of maximizing a likelihood function in a simple case where there are only two parameters to be estimated (the **probability of capture** and the **population size**).
- ! Usually it is the **log-likelihood function** that is maximized, but this is equivalent to maximizing the likelihood function.

Model Selection

- ! With capture-recapture data there are often many alternative models that might be appropriate so that model selection procedures become important.
- ! Section 1.4 of the book discusses two of these procedures.
- ! One approach is to see whether the goodness of fit of a model in terms of the log-likelihood is significantly improved by adding extra parameters into the model. This works with **nested models**.

- ! The other approach uses the **Akaike information criterion (AIC)**. This attempts to find the best compromise between the goodness of fit of the model (as measured by the log-likelihood) and the number of parameters to be estimated.
- ! In general increasing the number of estimated parameters gives an improvement in the fit of a model but may result in poor estimates of those parameters.

Notation

- ! A large amount of specialized notation is required to describe models for capture-recapture data. Chapter 1 of the book concludes with a list of the symbols used most often.

Introduction to Closed Population Models

- ! This part of the webinar concentrates on early methods for the estimation of **closed animal populations**, i.e. where the size is not changing because of individuals moving in and out of the population.
- ! As noted before, the idea of the two-occasion capture-recapture method can be traced to Pierre Laplace, who used it to estimate the human population size of France in 1802, and even earlier to John Graunt who used the idea to estimate the effect of plague and the size of the population of England in the 1600s.
- ! Other early applications of this method in ecology included Petersen's and Dahl's work on sampling fish populations in 1896 and 1917, respectively, while Lincoln was the first to apply the method to wildlife in 1930 when he used returned leg bands from hunters to estimate duck numbers.

Structure of Capture-Recapture Data

! The data from closed capture-recapture experiments are the capture records of all the individuals observed in a study - can be arranged in a **capture history matrix**.

Example

- ! Deer mice (*Peromyscus Sp.*) data collected by V. Reid.
- ! Arose from a live-trapping experiment that was conducted for six consecutive nights (columns) with a total of 38 mice (rows) captured over the six capture occasions.
- ! Time period relatively short so it is reasonable to assume that the population was closed.
- ! Capture history of each mouse is expressed as a series of 0's (non-captures) and 1's (captures).

Individual capture history of 38 deer mice with six capture occasions

Occasion 1	Occasion 2	Occasion 3	Occasion 4	Occasion 5	Occasion 6
1	1	1	1	1	1
1	0	0	1	1	1
1	1	0	0	1	1
1	1	0	1	1	1
1	1	1	1	1	1
1	1	0	1	1	1
1	1	1	1	1	0
1	1	1	0	0	1
1	1	1	1	1	1
1	1	0	1	1	1
1	1	0	1	1	1
1	1	1	0	1	1
1	1	1	1	1	1
1	0	1	1	1	0
1	0	0	1	0	0
0	1	0	0	1	0
0	1	1	0	0	1
0	1	0	0	0	1
0	1	0	1	0	1
0	1	1	0	1	0
0	1	0	1	0	1
0	1	0	0	0	1
0	1	0	0	1	1
0	0	1	0	0	0
0	0	1	1	1	1
0	0	1	0	1	1
0	0	1	1	1	1
0	0	1	0	1	0
0	0	1	0	0	0
0	0	0	1	0	0
0	0	0	1	1	1
0	0	0	1	1	0
0	0	0	0	1	0
0	0	0	0	1	0
0	0	0	0	1	0
0	0	0	0	0	1
0	0	0	0	0	1
0	0	0	0	0	1

Summary Statistics

For many estimation procedures, the data can be summarized by:

$k =$ the number of capture occasions,

$n_j =$ the number of animals captured on the j th capture occasion,

$u_j =$ the number of unmarked animals captured on the j th capture occasion,

$m_j =$ the number of marked animals captured on the j th capture occasion, with $m_1 = 0$,

$M_j =$ the number of distinct animals captured before the j th capture occasion, (the number of marked animals in the population just before the j th capture occasion), with $M_1 = 0$, and

$f_j =$ the number of animals captured exactly j times, $j = 1, \dots, k$.

Data summary for deer mice

j	n_j	m_j	u_j	M_j	f_j
1	15	0	15	0	9
2	20	12	8	15	6
3	16	10	6	23	7
4	19	16	3	29	6
5	25	22	3	32	6
6	25	22	3	35	4
Total	120	82	38		38

Assumptions Common for All Closed Population Models

- ! The population remains constant over the study period (no births, deaths, immigration or emigration) = a closed population.
- ! Animals do not lose their marks or tags.
- ! All marks or tags are correctly recorded.
- ! Animals act independently.

Two-Occasion Models (the Petersen-Lincoln Estimator)

- ! The origin of capture-recapture methodology.
- ! Assume that on the first capture occasion, a sample of n_1 animals is captured, marked and released, so that the marked proportion of the population is

$$n_1/N.$$

- ! On the second capture occasion, n_2 animals are caught, m_2 of which are marked, so that the proportion of marked animals in the second sample is m_2/n_2 .
- ! Assuming that the marked proportion in the sample is equal to the marked proportion in the population suggests that

$$m_2/n_2 \approx n_1/N,$$

so that

$$N \approx (n_1 n_2)/m_2.$$

- ! This is the **Petersen-Lincoln estimator**, which is also known as the **Lincoln-Petersen estimator**.

Example

- ! First two samples for deer mouse data.
- ! $n_1 = 15$, $n_2 = 20$ and $m_2 = 12$, with a total of 23 distinct mice being captured on the these first two occasions.
- ! The Petersen-Lincoln estimator is then
$$15 \times 20 / 12 = 25.$$
- ! A total of 38 distinct mice were actually seen at the end of the six samples.

Properties of the Estimator

- ! Various sampling models have been proposed to justify the Petersen-Lincoln estimator and obtain estimates of the associated standard error.
- ! If all animals have the same capture probability on each occasion then the number of captures on each occasion is a random variable. There are three observable capture histories (10, 01 and 11) and three parameters (N , p_1 , p_2), where p_1 and p_2 denote the capture probabilities for the first and second capture occasions.
- ! The Petersen-Lincoln estimator turns out to be an approximate **ML estimator** of N .
- ! Some other models are also reasonable.

- ! The Petersen-Lincoln estimator is biased but modifications to remove or reduce the bias are available, such as the Chapman (1951) estimator

$$\hat{N} = (n_1 + 1)(n_2 + 1)/(m_2 + 1) - 1.$$

- ! With or without a bias correction the estimators have the same approximate equation for the variance, given by

$$\text{Var} \approx \frac{(n_1+1)(n_2+1)(n_1-m_2)(n_2-m_2)}{[(m_2+1)^2 (m_2+2)].}$$

- ! Approximate confidence intervals for true population sizes are calculated in the usual way, e.g. as the estimate +/- 1.96 estimated standard errors.

Multiple-Recapture Models (The Schnabel Census)

- ! Two-occasion models were extended to multiple occasions by Schnabel (1938) and Darroch (1958).
- ! Like the Petersen-Lincoln model, these early models assume that on each sampling occasion all animals have the same probability of capture, although this probability can be allowed to vary among sampling occasions.
- ! A special case of the equal catchability model is where the capture probability p is constant over the capture occasions. This is usually referred to as model M_0 , the subscript 0 referring to no variation.
- ! There are only the two parameters, N and p , and the maximum likelihood estimates can be obtained using numerical methods.
- ! A large-sample variance equation is available for the estimator of N .
- ! This model does not fit the deer mouse data (see the Handbook).

- ! If the model is changed to let the capture probabilities vary with time then this gives the model M_t , where the subscript t refers to time variation.
- ! The likelihood function is still quite simple but there are now k parameters p_1, p_2, \dots, p_k to describe the capture probabilities. plus the population size parameter N . Again maximum likelihood estimates can be obtained using a numerical procedure.
- ! This model still fails to provide a good fit to the deer mouse data (again see the Handbook).

Failure of Assumptions

- ! Marked animals that have a higher probability of being captured (e.g., because they are attracted by the bait) on subsequent occasions would be called **trap happy**.
- ! Marked animals may become **trap shy** because of the unpleasant experience of being captured. The recapture probability on subsequent capture occasions thus becomes lower than that for un-captured animals.
- ! The capture probability may be a property of the animal and thus the individual capture probabilities may be **heterogeneous** across the individuals in a population, with variation due to age, gender, body weight, activity, the number of traps near the home range or other unobservable individual characteristics.
- ! Violation of the equal catchability assumption leads to biases for the usual estimators.

The Models of Otis *et al.* (1978)

- ! Otis *et al.* considered the following three sources of variations in capture probabilities:

Time effects, where capture probabilities may vary due to environmental variables (e.g., temperature, time of day, rainfall or humidity) or sampling effort but covariates to explain the variation are not available.

Behavioral responses to capture (e.g., trap shyness).

Individual heterogeneity due to observable factors (e.g., gender, age or body weight) or unobservable inherent characteristics.

- ! Based on these three sources of variation, Otis *et al.* (1978) and White *et al.* (1982) formulated eight models including the starting null model M_0 with a constant capture probability..
- ! Also developed a computer program called **CAPTURE** to calculate estimators for each model.

The Eight Models

- ! Let P_{ij} = probability the i th individual is captured on the j th occasion.
- ! Also, the subscripts t , b , h on M denote time-variation, behavioral response, and heterogeneity respectively.
- ! Use the Reid deer mice data as an example.

1. Model M_0 : $P_{ij} = p$

- ! This simplest model with a **constant capture probability**.
- ! Only two parameters, N and p , are in the model.
- ! The likelihood equation is in the Handbook. For the mouse data minus twice the maximized log-likelihood $-2\ln(L_0) = 109.5$ (measuring goodness of fit). Akaike's information criterion (AIC) is 113.5 (with small values for good models).
- ! These two statistics are used to compare M_0 with the other models.

2. Model M_t : $P_{ij} = p_j$

- ! **Time-varying capture probabilities.**
- ! There are $k+1$ parameters (N, p_1, \dots, p_k). and the likelihood is given in the Handbook. For the mouse data the maximized log-likelihood function gives $-2\ln(L_t) = 99.7$ and $AIC = 113.7$.
- ! This model is not as good as M_0 according to AIC.

3. Model M_b : $P_{ij} = p$ until the first capture and $P_{ij} = c$ for any recaptures

- ! There is evidence that mice, voles, and other small mammals, as well as other species exhibit **trap response** to capture, especially the first capture.
- ! This model requires that all animals have the same behavioral response to initial capture.
- ! On any occasion, the capture probability for animals not previously caught is p whereas the probability is c ($c \neq p$) for any previously captured animal. If $p > c$, then this gives the **trap-shy** case while if $p < c$ then it gives the **trap-happy** case.
- ! All animals are assumed to behave in the same fashion with regard to their trap response.
- ! There are three parameters (N , p , c) for the model.

- ! For the deer data $\hat{N} = 41$, $\hat{p} = 0.34$, and $\hat{c} = 0.61$, suggesting that the animals became trap-happy on their first capture.
- ! For model M_b , $-2\ln(L_b) = 98.0$ and $AIC = 104.0$.
- ! Compared with model M_0 , model M_b provides a significantly better fit as twice the difference of the two log-likelihood functions is $2\ln(L_b) - 2\ln(L_0) = 11.5$, which is very significantly large for a chi-squared distribution with 1 df ($p < 0.001$).
- ! A similar comparison between models M_0 and M_t shows that these two models are not significantly different.

4. Model M_{tb} : $P_{ij} = p_j$ until the first capture and $P_{ij} = c_j$ for any recaptures

- ! A natural extension of model M_t to incorporate **behavioral response**.
- ! A total of $2k$ parameters ($N, p_1, p_2, \dots, p_k, c_2, c_3, \dots, c_k$) but it can be shown that the model is non-identifiable (with more parameters than data values) unless some restrictions or extra assumptions are made.
- ! Research on setting reasonable restrictions to remove the non-identification, is considered later in the webinar.

5. Model $M_h: P_{ij} = p_i$

- ! This model assumes that the i th animal has its own **unique capture probability**, p_i , which is constant over all trapping occasions.
- ! Studies have shown that the usual estimators based on the classical equal catchability assumption are negatively biased by heterogeneity of capture probabilities.
- ! Main difficulty with the model is the large numbers of parameters (p_1, p_2, \dots, p_N).
- ! One approach for estimation is to assume a parametric distribution for p_i such as the beta distribution, so that all the p_i values are replaced by the two parameters of the beta distribution but it has been found that estimation has poor properties.

- ! Alternatively, under the assumption that (p_1, p_2, \dots, p_N) are a random sample from any unknown distribution, Burnham and Overton (1978) showed that the capture frequencies (f_1, f_2, \dots, f_k) are all that is needed for estimating the population size, where f_j denotes the number of animals captured exactly j times on occasions 1 to k .
- ! Burnham and Overton proposed the use of **jackknife estimators** up to the fifth order for estimating population size. See the Handbook for more details.

6. Model M_{bh} : $P_{ij} = p_i$ until first capture and $P_{ij} = c_i$ for any recapture

- ! If model M_h is generalized to allow the individual capture probability to depend on previous capture history, it gives model M_{bh} .
- ! For this model, Otis *et al.* proposed a **generalized removal method**, based on the idea that animals with larger capture probabilities tend to be removed in earlier occasions than animals with small capture probabilities (thinking of animals being "removed" on their first capture).
- ! Conditional on those animals not previously removed, the average removal probabilities of the k occasions are expected to be decreasing. For a homogeneous population, these conditional removal probabilities are constant.

- ! The generalized removal method first considers fitting a simple model with all average removal probabilities equal. If the simple model fits the data well, then the model is reduced to model M_b . In this case, the inferences are identical to those of model M_b .
- ! If the model does not fit, then a model is considered where the removal probability on the first occasion is larger than the others, so that it is assumed that after u_1 animals have been removed, the remaining $N-u_1$ animals have identical probabilities of being removed and can be fitted to model M_b .
- ! The generalized removal method fits successively more general models until an acceptable fit is found. Then the population size is obtained based on the final selected model.
- ! Simulation studies have shown that the generalized removal estimator typically has a negative bias, which can be large if the removal probabilities are very heterogeneous.

- ! An **improved estimator** due to Pollock and Otto is described in the Handbook.
- ! For the deer mouse data the model assuming that all removal probabilities are equal fits well, giving the same estimates as model M_b .

7. Model M_{th} : $P_{ij} = p_i e_j$

- ! This model combines **time and heterogeneity effects**, with a multiplicative form of the effects being assumed.
- ! No estimator of population size had been suggested at the time that the paper by Otis *et al.* (1978) was published.
- ! Recent developments of this model are considered later in the webinar.

8. Model M_{tbh} : $P_{ij} = p_{ij}$ until first capture and $P_{ij} = c_{ij}$ for any recapture

- ! Three sources of variations in capture probabilities are considered.
- ! This most general model has in the past been considered conceptually useful but too complicated to be applied to practical situations if no restrictions are made.
- ! New methods for handling this model that usually require extensive numerical computations have been developed and are discussed later in the webinar.

Example of a Likelihood Calculation

! Seen in the spreadsheet M0-LIK.xls.

Summary

- ! What is in the full webinar
- ! The history of capture-recapture methods
- ! Maximum likelihood estimation.
- ! Model selection methods
- ! Introduction to closed population models
- ! Structure of the data
- ! Two sample situation
- ! Multi-sample situation
- ! The Otis *et al.* models
- ! Example of a likelihood calculation