APPENDIX A

Notes on Estimation

Justification for the Use of Maximum Likelihood Estimation

Our philosophy throughout this work has been to present a variety of explicit mathematical models for capture data based upon fully specified assumptions. Given these models that incorporate various types of variability in capture probabilities (time, behavior, and heterogeneity), we are concerned with optimal estimation of population size, N, under each model. To achieve that optimality, we turn to the field of mathematical statistics.

A number of very general approaches to optimal statistical inference have been developed during the past 50 years. We feel that not all of them are suitable for use by biologists for the problem at hand (e.g., decision theoretic, or Bayesian approaches would require inputs from the biologist that we doubt they would be willing, or able, to supply). We have chosen to use the method of maximum likelihood (ML) estimation and inference; it is one of the best developed, omnibus tools of mathematical statistics.

The properties of ML estimation are well known (see Mood et al. 1974). For many practical models, ML estimators are optimal in many desirable ways. Because of its general applicability and good properties, ML estimation has been the basis for most modern methods of estimating parameters from animal marking experiments of all kinds (Seber 1973).

The only requirement for application of ML estimation is that one have a well-defined parametric model with fewer parameters than the dimension of the vector of minimal sufficient statistics. Thus, we have been able to use ML estimation for Models M_o, M_t, M_b, M_{bh}, and the removal models, but not for Model M_h. Only for Model M_h do we hesitate to claim there cannot be significant improvements in the estimator.

It is not necessary to understand the mechanics of how a ML estimate is computed; it suffices to understand this is not an ad hoc technique, but rather a well-developed inference method.

For any of the models dealt with here, there is a sampling probability distribution for the basic data (the $\{X_{ij}\}$) that can be represented as a mathematical function of appropriate parameters; symbolically it is

$$P\{X_{ij} | N,\mathbf{p}\},$$

for $N \geqslant M_{t+1}$, and with $\mathbf{p}$ representing a set of capture probability parameters (e.g., $\mathbf{p} = (p,c)$ under Model M_b). Given an actual sample, we can substitute these data for the X_{ij} variables and treat this formula as strictly a function of the parameters N and $\mathbf{p}$. This function (of N,$\mathbf{p}$) is called the likelihood function. The ML estimators $\hat{N}$ and $\hat{\mathbf{p}}$ are those values of N and $\mathbf{p}$ which maximize the function

$$L(N,\mathbf{p}) = P\{X_{ij} | N,\mathbf{p}\}$$

or, equivalently, which maximize the log of $L(N,\mathbf{p})$. In this way, the problem of deriving parameter estimators is reduced to the classic problem of maximizing a given function over a set of possible values of specified variables.

The theory of ML estimation goes much deeper than just deriving point estimators of parameters. It also gives methods for deriving approximate sampling variances for the estimator and provides justification for the usual procedure for confidence interval construction. Also, there is a strong tie between ML theory and the theory of minimal sufficient statistics.

A sufficient statistic is a condensation of the sample into a smaller number of statistics that still contain all the information there is about N and **p**. For example, M_{t+1} and n. are the minimal sufficient statistics for N and p under Model M_0 (representing quite a condensation of the entire sample of X_{ij} values). The ML estimator is always a function of the sufficient statistic, even if one has not explicitly determined the sufficient statistic. Partly because ML estimators are functions of the sufficient statistic, they have, for large samples, the smallest possible sampling variances in the class of consistent estimators. Stated differently, ML estimators are (asymptotically) the most efficient estimators possible under a sampling theory approach to inference. For all the above reasons, we have used likelihood theory as the main basis of estimation in this monograph.

Numerical Methods

Except for the special case of only 2 trapping occasions (t = 2), the ML estimators described in the preceding chapters do not have explicit mathematical expressions. Hence, numerical methods must be used to maximize the likelihood functions and thereby find the population estimates on a case by case basis. Conceptually, this turns out to be easy for models to which we can apply ML estimation. From above, the likelihood function is representable simply as L(N,**p**). In all applicable cases (exclude Model M_h and models where estimation is not possible), we find there is a closed form expression for the value of the capture probabilities which maximize the function L(N,**p**) for any *fixed* $N \geqslant M_{t+1}$. Represent this as $\hat{\mathbf{p}}(N)$. Thus we can write

$$\max_{N} \max_{\mathbf{p}} L(N,\mathbf{p}) = \max_{N} L(\hat{p}(N) \mid N)$$
$$= \max_{N} \lambda(N),$$

where $\lambda(N)$ is a function (analytically expressible) only of N. The maximization of $\lambda(N)$ over N must be done numerically, but this is a simple one-dimensional search.

An algorithm by Brent (1973) (also see Fletcher 1972) that does not require analytic derivatives is used by program CAPTURE to maximize $\lambda(N)$. The algorithm searches between 2 end points. Those end points are first taken as the number of animals captured, M_{t+1}, for the lower bound and a linear approximation to $\hat{N}$ plus M_{t+1} for the upper bound. If the upper bound is determined to be the maximum for the range specified, a new set of values lying next to the old values is selected and the search over N continues in this fashion until a true maximum is found. The algorithm searches along the real line although only integer values are appropriate. Because the search is one dimensional, the algorithm is very efficient.

The algorithm does not require the analytic first derivatives of $\lambda(N)$. This means it is not necessary to evaluate the derivative of the log-gamma function with respect to N(N! and $(N - M_{t+1})!$ appear in all the likelihood functions).

Integer $\hat{N}$ and Confidence Bounds

The domain of $\hat{N}$ includes only integer values. Therefore, the integer that produces a maximum in the likelihood function is taken as the estimate. The numerical search is conducted along the real number line; then the value returned is truncated to an integer, and a check is made to see that it is the estimate, and not $\hat{N} + 1$. A somewhat arbitrary rounding to the nearest whole integer is required for the jackknife estimator to maintain consistency with the other estimators.

For all estimators, the variance is computed on the basis of the value of $\hat{N}$ before it is rounded to an integer. We believe this procedure produces a slightly better estimator of $\mathrm{Var}(\hat{N})$, and at the same time avoids numerical problems that would occur when $\hat{N}$ equals M_{t+1} (this situation does occur).

Confidence intervals of 95 percent are computed as $\hat{N} \pm 1.96\ \hat{S}E(\hat{N})$. Confidence intervals (but not standard errors) are computed around the integer value of $\hat{N}$. The lower bound is then truncated to an integer and the upper bound rounded upward to an integer. Those interval end points are thus technically outside the true 95 percent confidence interval. When the lower bound is less than the number of animals captured, the value could be reset to M_{t+1}, a realistic procedure because we know that N is not less than M_{t+1}. The coverage of the interval is not changed by this procedure of using M_{t+1} for the lower bound when the calculated lower bound is less than M_{t+1}.

Admittedly, a confidence interval procedure that can yield a lower bound less than the number of distinct animals captured (M_{t+1}) is not desirable. We investigated alternatives that avoid this problem, but we judged them even less appropriate than the simple approach described above (see Appendix O).

Estimation of Sampling Variance

Maximum likelihood theory includes a general method for obtaining the large sample (i.e., asymptotic) formula for the true sampling variance of $\hat{N}$. We have denoted this simply as $\mathrm{Var}(\hat{N})$, thus suppressing the fact that it is an approximation to the "true" sampling variance of $\hat{N}$. The approximation is generally good for large samples; unfortunately, in the capture–recapture context we have no good measure of when a sample is sufficiently large. A further complication enters because the formula for $\mathrm{Var}(\hat{N})$ is almost always a function of the unknown parameters N and $\mathbf{p}$; symbolically it is generally of the form

$$\mathrm{Var}(\hat{N}) = Ng(\mathbf{p}),$$

with the form of the function g known. But because N and $\mathbf{p}$ are not known we must estimate $\mathrm{Var}(\hat{N})$ by

$$\hat{\mathrm{V}}\mathrm{ar}(\hat{N}) = \hat{N}g(\hat{\mathbf{p}}).$$

The properties of $\hat{\mathrm{V}}\mathrm{ar}(\hat{N})$ can be different (sometimes quite so) from those of $\mathrm{Var}(\hat{N})$, which itself is only an approximation to the appropriate finite sampling variance of $\hat{N}$. One particular problem that has been noted in the literature concerns the fact that $\hat{N}$ and $\hat{S}E(\hat{N}) = \sqrt{\hat{\mathrm{V}}\mathrm{ar}(\hat{N})} = \sqrt{\hat{N}g(\hat{\mathbf{p}})}$ can be substantially correlated. This is not surprising, because, of course, $\hat{N}$ and $\sqrt{\hat{N}}$ are highly correlated. The effect of this correlation is to cause an underestimate of $SE(\hat{N})$ when $\hat{N}$ is lower than the true N (which it will be much

of the time). This, in turn, causes the confidence intervals in such cases to be shorter than they should be. An area suitable for further research on capture–recapture statistics (both closed and open models) is the question of improved estimation of sampling variances *and* confidence intervals.

APPENDIX B

Estimation in Model M_o

In Model M_o, parameterized by the parameters N and p, the relevant part of the log-likelihood function is given by

$$\ln L(N,p|\mathbf{X}) = \ln\left(\frac{N!}{(N - M_{t+1})!}\right) + (n.)\ln(p) + (tN - n.)\ln(1 - p),$$

where $p\epsilon[0,1]$ and $N\epsilon\mathcal{N} = \{M_{t+1}, M_{t+1} + 1, M_{t+1} + 2, \ldots\}$. Given the value of N, the ML estimator $\hat{p}(N)$ of p is given as the solution to

$$\frac{\partial}{\partial p} \ln L(p|N,\mathbf{X}) = 0,$$

which reduces to

$$\frac{n.}{\hat{p}(N)} = \frac{tN - n.}{1 - \hat{p}(N)}.$$

This results in the solution

$$\hat{p}(N) = \frac{n.}{tN}.$$

Now, the ML estimator $\hat{N}_o$ of N satisfies

$$\begin{aligned}\ln L(\hat{N}_o,\hat{p}(\hat{N}_o)|\mathbf{X}) &= \max_{N\epsilon\mathcal{N}}\left[\max_{p\epsilon[0,1]} \ln L(p|N,\mathbf{X})\right] \\ &= \max_{N\epsilon\mathcal{N}}\left[\ln L(\hat{p}(N)|N,\mathbf{X})\right] \\ &= \max_{N\epsilon\mathcal{N}}\left[\ln\left(\frac{N!}{(N - M_{t+1})!}\right) + (n.)\ln(n.)\right. \\ &\qquad \left. + (tN - n.)\ln(tN - n.) - tN\ln(tN)\right].\end{aligned}$$

For a given data set, a search over $\mathcal{N}$ is performed to locate the ML estimate $\hat{N}_o$. This value is then used in the calculation of the ML estimate $\hat{p}(\hat{N}_o) = \hat{p}$ via

$$\hat{p} = \frac{n.}{t\hat{N}_o}.$$

The asymptotic variance of $\hat{N}_o$ is (cf. Darroch 1959)

$$\mathrm{Var}(\hat{N}_o) = N[(1 - p)^{-t} - t(1 - p)^{-1} + t - 1]^{-1}.$$

An estimate of this variance is

$$\hat{\mathrm{V}}\mathrm{ar}(\hat{N}_o) = \hat{N}_o[(1 - \hat{p})^{-t} - t(1 - \hat{p})^{-1} + t - 1]^{-1}.$$

APPENDIX C

Estimation in Model M_t

Model M_t is parameterized by the $t + 1$ parameters $N, p_1, p_2, \ldots, p_t$. The relevant log-likelihood function for estimation of the parameters is given by

$$\ln L(N,\mathbf{p}\,|\,\mathbf{X}) = \ln\left(\frac{N!}{(N - M_{t+1})!}\right) + \sum_{j=1}^{t} n_j \ln(p_j) + \sum_{j=1}^{t} (N - n_j)\ln(1 - p_j),$$

where $N \epsilon \mathcal{N} = \{M_{t+1}, M_{t+1} + 1, M_{t+1} + 2, \ldots\}$, $\mathbf{p} = \{p_1, p_2, \ldots, p_t\}$, $p_j \epsilon [0,1]$ for $j = 1, 2, \ldots, t$. Given the value of N, the ML estimators $\hat{p}_j(N)$ of p_j are given as the solutions to the system of equations

$$\frac{\partial}{\partial p_j} \ln L(\mathbf{p}\,|\,N,\mathbf{X}) = 0, \qquad j = 1,2,\ldots,t.$$

The jth one of these equations reduces to

$$\frac{n_j}{\hat{p}_j(N)} = \frac{N - n_j}{1 - \hat{p}_j(N)},$$

which results in the solution

$$\hat{p}_j(N) = \frac{n_j}{N}.$$

Now, the ML estimator $\hat{N}_t$ of N satisfies

$$\begin{aligned}
\ln L(\hat{N}_t, \hat{p}_1(\hat{N}_t), \ldots, \hat{p}_t(\hat{N}_t)|\mathbf{X}) &= \max_{N \epsilon \mathcal{N}} \left[\max_{p_j \epsilon [0,1]} \ln L(p_1, p_2, \ldots, p_t\,|\,N,\mathbf{X})\right] \\
&= \max_{N \epsilon \mathcal{N}} \left[\ln L(\hat{p}_1(N), \hat{p}_2(N), \ldots, \hat{p}_t(N)\,|\,N,\mathbf{X})\right] \\
&= \max_{N \epsilon \mathcal{N}} \left[\ln\left(\frac{N!}{(N - M_{t+1})!}\right) + \sum_{j=1}^{t} n_j \ln(n_j) \right. \\
&\qquad + \sum_{j=1}^{t} (N - n_j)\ln(N - n_j) \\
&\qquad \left. - tN\ln(N)\right].
\end{aligned}$$

A search over $\mathcal{N}$ is performed in order to locate the ML estimator $\hat{N}_t$. It is then possible to calculate the ML estimates $\hat{p}_j(\hat{N}_t) = \hat{p}_j$ of the p_j for $j = 1,2,\ldots,t$ via

$$\hat{p}_j = \frac{n_j}{\hat{N}_t}.$$

The asymptotic variance of $\hat{N}_t$ is given by Darroch (1958) as:

$$\mathrm{Var}(\hat{N}_t) = N\left[\frac{1}{\prod_{j=1}^{t}(1 - p_j)} + t - 1 - \sum_{j=1}^{t}(1 - p_j)^{-1}\right]^{-1}.$$

An estimate of this variance is

$$\hat{V}ar(\hat{N}_t) = \hat{N}_t\left[\frac{1}{\prod_{j=1}^{t}(1-\hat{p}_j)} + t - 1 - \sum_{j=1}^{t}(1-\hat{p}_j)^{-1}\right]^{-1}.$$

APPENDIX D

Estimation in Model M_b

Model M_b is parameterized by the parameters N,p, and c. The part of the log-likelihood necessary for estimation of the parameters is given by

$$\ln L(N,p,c|\mathbf{X}) = \ln\left(\frac{N!}{(N-M_{t+1})!}\right) + M_{t+1}\ln(p) + (tN - M. - M_{t+1})\ln(1-p)$$
$$+ m.\ln(c) + (M. - m.)\ln(1-c),$$

where $N \epsilon \mathcal{N} = \{M_{t+1}, M_{t+1}+1, M_{t+1}+2, \ldots\}$, $p\epsilon[0,1]$, $c\epsilon[0,1]$. The ML estimator $\hat{c}$ of c is produced by the equation

$$\frac{\partial}{\partial c}\ln L(N,p,c|\mathbf{X}) = 0,$$

which reduces to

$$\frac{m.}{\hat{c}} = \frac{M. - m.}{1-\hat{c}}.$$

Solving this equation gives $\hat{c} = m./M.$. Thus, we see that estimation of c is independent of the estimation of N and p. Now, the relevant part of the log-likelihood function for purposes of estimating N and p is

$$\ln L(N,p|\mathbf{X}) = \ln\left(\frac{N!}{(N-M_{t+1})!}\right) + M_{t+1}\ln(p) + (tN - M. - M_{t+1})\ln(1-p).$$

Given the value of N, the ML estimator $\hat{p}(N)$ of p is provided by the equation

$$\frac{\partial}{\partial p}\ln L(p|N,\mathbf{X}) = 0,$$

which reduces to

$$\frac{M_{t+1}}{\hat{p}(N)} = \frac{tN - M. - M_{t+1}}{1-\hat{p}(N)}.$$

The solution to this equation gives

$$\hat{p}(N) = \frac{M_{t+1}}{tN - M.}.$$

Now, the ML estimator $\hat{N}_b$ of N satisfies

$$\ln L(\hat{N}_b,\hat{p}(\hat{N}_b)|\mathbf{X}) = \max_{N\epsilon\mathcal{N}}\left[\max_{p\epsilon[0,1]}\ln L(p|N,\mathbf{X})\right]$$
$$= \max_{N\epsilon\mathcal{N}}\left[\ln L(\hat{p}(N)|N,\mathbf{X})\right]$$

$$= \max_{N \in \mathcal{N}} \left[\ln\left(\frac{N!}{(N - M_{t+1})!}\right) + M_{t+1}\ln(M_{t+1}) + (tN - M. - M_{t+1})\ln(tN - M. - M_{t+1}) - (tN - M.)\ln(tN - M.)\right].$$

A search over $\mathcal{N}$ is preformed to locate the ML estimate $\hat{N}_b$. It is then possible to calculate the ML estimate $\hat{p}(\hat{N}_b) = \hat{p}$ of p via

$$\hat{p} = \frac{M_{t+1}}{t\hat{N}_b - M.}.$$

The asymptotic variance of $\hat{N}_b$ is given by Zippin (1956) as:

$$\mathrm{Var}(\hat{N}_b) = \frac{N(1 - p)^t[1 - (1 - p)^t]}{[1 - (1 - p)^t]^2 - t^2p^2(1 - p)^{t-1}}.$$

An estimate of this variance is given by

$$\hat{\mathrm{V}}\mathrm{ar}(\hat{N}_b) = \frac{\hat{N}_b(1 - \hat{p})^t[1 - (1 - \hat{p})^t]}{[1 - (1 - \hat{p})^t]^2 - t^2\hat{p}^2(1 - \hat{p})^{t-1}}.$$

APPENDIX E

Estimation in Model M_h

The mathematical details of Model M_h are covered in detail in Burnham (unpublished dissertation) and Burnham and Overton (pers. comm.). The following gives a few of the basic results for this model and the jackknife estimator.

Under Model M_h we assume

$$p_{ij} = p_i \qquad i = 1, \ldots, N,$$

and we further assume that the $p_1, \ldots, p_N$ are a random sample from some (unknown) probability distribution F(p), $p \in [0,1]$. For any F(p), the MSS contains only the capture frequencies $f_1, \ldots, f_t$, and their distribution is multinomial:

$$P\{f_1, \ldots, f_t | F\} = \binom{N}{N - M_{t+1}, f_1, \ldots, f_t} (\pi_0)^{N-M_{t+1}} \prod_{j=1}^{t} (\pi_j)^{f_j},$$

where

$$\pi_j = \int_0^1 \binom{t}{j} p^j(1 - p)^{t-j} dF(p), \qquad j = 0, \ldots, t.$$

If one assumes a parameteric form for F(p), then standard approaches are possible (e.g., ML). Burnham (unpublished dissertation) investigated the case where F(p) is the class of beta distributions and found that approach basically unacceptable.

The "jackknife" estimator used here was developed by application of the generalized jackknife statistic (Gray and Schucany 1972) to the naive estimator M_{t+1}, assuming the bias of M_{t+1} as an estimator of N is expressible

TABLE E.1.—THE JACKKNIFE ESTIMATORS $\hat{N}_{hk}$ OF POPULATION SIZE, FOR K = 1 TO 5

$$\hat{N}_{h1} = M_{t+1} + \left(\frac{t-1}{t}\right)f_1$$

$$\hat{N}_{h2} = M_{t+1} - \left(\frac{2t-3}{t}\right)f_1 - \frac{(t-2)^2}{t(t-1)}f_2$$

$$\hat{N}_{h3} = M_{t+1} + \left(\frac{3t-6}{t}\right)f_1 - \left(\frac{3t^2-15t+19}{t(t-1)}\right)f_2 + \frac{(t-3)^3}{t(t-1)(t-2)}f_3$$

$$\hat{N}_{h4} = M_{t+1} + \left(\frac{4t-10}{t}\right)f_1 - \left(\frac{6t^2-36t+55}{t(t-1)}\right)f_2 + \left(\frac{4t^3-42t^2+148t-175}{t(t-1)(t-2)}\right)f_3 - \frac{(t-4)^4}{t(t-1)(t-2)(t-3)}f_4$$

$$\hat{N}_{h5} = M_{t+1} + \left(\frac{5t-15}{t}\right)f_1 - \left(\frac{10t^2-70t+125}{t(t-1)}\right)f_2 + \left(\frac{10t^3-120t^2+485t-660}{t(t-1)(t-2)}\right)f_3$$
$$- \left(\frac{(t-4)^5-(t-5)^5}{t(t-1)(t-2)(t-3)}\right)f_4 + \frac{(t-5)^5}{t(t-1)(t-2)(t-3)(t-4)}f_5$$

in a power series in 1/t. The end result of that application is that for the kth order jackknife the estimator is a linear function of the capture frequencies. For example, we have

$$\hat{N}_{h1} = M_{t+1} + \left(\frac{t-1}{t}\right)f_1 \equiv \left(1 + \frac{t-1}{t}\right)f_1 + \sum_{j=2}^{t} f_j.$$

These point estimators of N have been determined for up to the fifth order, and are given in Table E.1.

Because

$$M_{t+1} = \sum_{j=1}^{t} f_j,$$

any $\hat{N}_{hk}$ is expressible as a linear combination of the capture frequencies, say as

$$\hat{N}_{hk} = \sum_{j=1}^{t} a_{jk}f_j.$$

Using the fact that the f_j are multinomial random variables an approximate variance estimator of $\hat{N}_{hk}$ is

$$\hat{V}ar(\hat{N}_{hk}) = \sum_{j=1}^{t} (a_{jk})^2 f_j - \hat{N}_{hk},$$

and confidence intervals can be constructed on the basis of the asymptotic normality of $\hat{N}_{hk}$.

A procedure for selecting one of the estimators has been suggested by Burnham and Overton (pers. comm.). It involves testing whether $\hat{N}_{h,k+1}$ is significantly different from $\hat{N}_{hk}$ and stopping when no significant difference is found.

APPENDIX F

Discussion of Model M_{tb}

The most general model one might assume in the instance where both behavioral response to capture and variability among trapping occasions are found involves the following set of parameters:

N = population size
p_j = probability of capture on trapping occasion j of an animal not previously captured, $j = 1,2,\ldots,t$,
c^*_{kj} = probability of capture on trapping occasion j of an animal first captured on occasion k, $j = k + 1, k + 2,\ldots,t$; $k = 1,\ldots,t - 1$.

The corresponding probability distribution of $\{X_\omega\}$ is given by

$$P[\{X_\omega\}] = \frac{N!}{\left[\prod_\omega X_\omega!\right](N - M_{t+1})!}\left[\prod_{j=1}^{t} p_j^{u_j}(1 - p_j)^{N-M_{j+1}}\right]$$
$$\cdot\left[\prod_{k=1}^{t-1}\prod_{j=k+1}^{t} c^*_{kj}{}^{R_{kj}}(1 - c^*_{kj})^{u_k-R_{kj}}\right],$$

where

u_j = number of animals first caught on occasion j, $j = 1,2,\ldots,t$,
M_j = number of marked (previously captured) animals in the population at the time of the jth sample, $j = 2,3,\ldots,t$,
M_{t+1} = total number of *different* animals captured in the experiment

$$(\text{notice } M_{t+1} = \sum_{j=1}^{t} u_j),$$

R_{kj} = number of animals caught on occasion j that were first captured on occasion k, $j = k + 1, k + 2, \ldots, t$; $k = 1, \ldots, t - 1$.

A MSS for this distribution is $\{u_1, u_2, \ldots, u_t, R_{12}, R_{13}, \ldots, R_{1t}, R_{23}, \ldots, R_{2t}, \ldots, R_{t+1,t}\}$, that has dimension $t(t + 1)/2$. There are $t(t + 1)/2 + 1$ parameters involved and straightforward ML estimation of N is not possible.

If we restrict the model so that $c^*_{kj} = c_j$, i.e., the probability of recapture on occasion j does not depend on the occasion of first capture, then Model M_{tb} (as previously defined) results. It is easily verified that, when $c^*_{kj} = c_j$, $P[\{X_\omega\}]$ given above reduces to the probability distribution given in the section concerning Model M_{tb} where a discussion of the nonidentifiability of N is presented. Thus, for purposes of estimating N, the assumption that $c^*_{kj} = c_j$ has no utility.

ML estimation of N is theoretically possible if one is willing to further restrict the model so that there is some relationship between c^*_j and p_j. There is no unique restriction, and as an example we consider $c^*_j = \theta p_j$, $j = 2, 3, \ldots, t$, where $0 \leq p_j \leq 1$, $j = 1, 2, \ldots, t$ and $0 \leq; \theta p_j \leq 1$, $j = 2, 3, \ldots, t$. With this assumption, the probability distribution of $\{X_\omega\}$ becomes

$$P[\{X_\omega\}] = \frac{N!}{\prod_\omega X_\omega!(N - M_{t+1})!}\left[\prod_{j=1}^{t} p_j^{u_j}(1 - p_j)^{n-M_{j+1}}\right]$$
$$\cdot\left[\prod_{j=2}^{t} (\theta p_j)^{m_j}(1 - \theta p_j)^{M_j-m_j}\right]$$

$$= \frac{N!}{\prod_{\omega} X_{\omega}!(N - M_{t+1})!} p_1^{u_1}(1 - p_1)^{N-u_1}(\theta^{m_\cdot})$$

$$\cdot \prod_{j=2}^{t} (1 - p_j)^{N-M_{j+1}}(1 - \theta p_j)^{M_j - m_j} p_j^{n_j}.$$

A sufficient statistic for this distribution is $\{u_1, u_2, \ldots, u_t, m_2, m_3, \ldots, m_t\}$, that has dimension $2t - 1$. Since there are $t + 2$ parameters involved, the model allows all parameters, in particular N, to be identified if $t \geqslant 3$. Estimation under this model is not considered here, however, due to the questionable assumption that recapture probabilities bear a constant relationship to initial capture probabilities over all trapping occasions, for all animals. Thus, at present we are not able to present an estimator of N that is appropriate under the assumptions of Model M_{tb}.

APPENDIX G

Discussion of Model M_{th}

In the section concerned with Model M_h, the random variable X_{ij} was stated to be distributed according to a Bernoulli distribution with parameter p_{ij}, that represents the probability that the ith animal is captured on the jth occasion. Here, we assume that the elements of the set $\{X_{ij}\}$ constitute a mutually independent collection of random variables, and that $p_{ij} = p_i p_j$, $i = 1,2,\ldots,N$, $j = 1,2,\ldots,t$. Furthermore, we let the p_i, $i = 1,2,\ldots,N$ be a random sample of size N from some probability distribution function $F(p;\boldsymbol{\theta})$ defined on [0, 1], and we restrict the p_j, $j = 1, 2, \ldots, t$ so that $0 \leqslant p_i p_j \leqslant 1$ for all i,j. Under those conditions, the probability distribution of the observed sample $\{X_{ij}\}$, where $i = 1, 2, \ldots, M_{t+1}$, $j = 1, 2, \ldots, t$ can be written as

$$P[\{X_{ij}\}] = P[\{X_{ij}\} \mid M_{t+1}]P[M_{t+1}],$$

where

$$P[\{X_{ij}\} \mid M_{t+1}] = \left(\prod_{j=1}^{t} p_j^{n_j}\right)\left(\prod_{i=1}^{M_{t+1}} \int_0^1 p^{y_i}\left[\prod_{j=1}^{t} (1 - pp_j)^{1-X_{ij}}\right] dF(p;\boldsymbol{\theta})\right),$$

y_i = the number of times the ith animal is captured,

$P[M_{t+1}]$ = the probability distribution of the number of different animals captured in the experiment. This distribution will involve the parameters N, $p_1, p_2, \ldots, p_t$, and the distribution $F(p;\boldsymbol{\theta})$.

Obviously, if the form of $F(p;\boldsymbol{\theta})$ is left unspecified, ML estimation of N will not be possible. If the form of $F(p;\boldsymbol{\theta})$ is specified, but the vector of parameters $\boldsymbol{\theta}$ is not, likelihood estimation will be possible if the dimension of the minimal sufficient statistic is sufficiently large to ensure identifiability of N. As mentioned in the section on Model M_{th}, estimation of N will be possible if $F(p;\boldsymbol{\theta})$ is completely specified. In that case, a MSS of dimension $t + 1$ is $\{n_1, n_2, \ldots, n_t, M_{t+1}\}$, and the number of parameters to be estimated is also $t + 1$, i.e., $p_1, p_2, \ldots, p_t, N$. Therefore, unless the experimenter is willing to make the doubtful assumption that at least the form of $F(p;\boldsymbol{\theta})$ is known, we can present no satisfactory estimation procedure appropriate under the assumptions of Model M_{th}.

APPENDIX H

Estimation in Model M_{bh}

Pollock (unpublished dissertation) considered Model M_{bh} and found the distribution of the set of possible capture histories $\{X_\omega\}$ under the assumption that (p_i,c_i), which represents the pair of initial and recapture probabilities, respectively, of the ith animal, is the ith member of a random sample of size N from some bivariate distribution function $G(p,c;\boldsymbol{\theta})$. Let us consider rewriting the distribution $P[\{X_\omega\}|G(p,c;\boldsymbol{\theta})]$ as $P[\{u_1,u_2,\ldots,u_t\}|G(p,c;\boldsymbol{\theta})]\cdot P[\{X_\omega\}|\{u_1,u_2,\ldots,u_t\}, G(p,c;\boldsymbol{\theta})]$. The distribution of the removals $P[\{u_1, u_2,\ldots,u_t\}|G(p,c;\boldsymbol{\theta})]$ will be multinomial, with parameter N and cell probabilities that are functions of the moments of $G(p,c;\boldsymbol{\theta})$. The conditional distribution $P[\{X_\omega\}|\{u_1, u_2, \ldots, u_t\}, G(p, c; \boldsymbol{\theta})]$ will also depend on the moments of $G(p,c;\boldsymbol{\theta})$ but not on the parameter N. Therefore, the multinomial distribution

$$P[\{u_1,u_2,\ldots,u_t\}|G(p,c;\boldsymbol{\theta})] = \frac{N!}{\left[\prod_{j=1}^{t} u_j!\right](N - M_{t+1})!}\left(\prod_{j=1}^{t}\pi_j^{u_j}\right)(\pi_{t+1})^{N-M_{t+1}},$$

where π_j is a function of the moments of $G(p,c;\boldsymbol{\theta})$ and

$$\pi_{t+1} = 1 - \sum_{j=1}^{t}\pi_j,$$

is the relevant distribution for purposes of estimating N. Pollock (unpublished dissertation) showed that if $G(p,c;\boldsymbol{\theta}) = G_1(p;\boldsymbol{\theta}_1)\cdot G_2(c;\boldsymbol{\theta}_2)$, then $\pi_j = E[p(1-p)^{j-1}]$ and the conditional distribution $P[\{X_\omega\}|\{u_1,u_2,\ldots,u_t\}, G(p,c;\boldsymbol{\theta})]$ depends only upon $G(c;\boldsymbol{\theta}_2)$. Regardless of whether p and c are independent, once we agree to base estimation of N solely on the first capture "removal" type of data, then the problem can be reformulated. The behavioral response is then irrelevant and all we need to consider is the probability distribution of first captures. That is, we assume that $p_1,\ldots,p_N$ are a random sample from some distribution $G_1(p;\boldsymbol{\theta})$.

Consider transforming the parameters $\pi_1,\pi_2,\ldots,\pi_t$ into the set of parameters $\bar{p}_1, \bar{p}_2, \ldots, \bar{p}_t$ by using the relationship $\pi_j = (1-\bar{p}_1)(1-\bar{p}_2)\ldots(1-\bar{p}_{j-1})\bar{p}_j$, $j = 1, 2, \ldots, t$. Thus, $\bar{p}_j$ is a conditional probability that represents the average first capture probability of those members of the population that have not yet been captured at the time of the jth trapping occasion. Given this interpretation, it is not unreasonable to assume that $\bar{p}_1 > \bar{p}_2 > \cdots > \bar{p}_t$. Furthermore, we assume that $(\bar{p}_1 - \bar{p}_2) > (\bar{p}_2 - \bar{p}_3) > \cdots > (\bar{p}_{t-1} - \bar{p}_t)$, so that larger differences in average first capture probability occur initially. (Note: if $G_1(p;\boldsymbol{\theta})$ is the class of beta distributions both assumptions are easily shown to be true.) The assumptions and the distribution $P[\{u_1,u_2,\ldots,u_t\}|G(p,c;\boldsymbol{\theta})]$, which we abbreviate $P[u_1,u_2,\ldots,u_t]$, form the basis of the generalized removal method outlined below.

For $k = 1,2,\ldots,t-2$,

(i) Assume $\bar{p}_k = \bar{p}_{k+1} = \cdots = \bar{p}_t = \bar{p}$ and that $\bar{p}_1,\bar{p}_2,\ldots,\bar{p}_{k-1}$ differ. This reduces the number of parameters involved in the estimation of N to $k + 1$.

(ii) Estimate N by the ML method. This task is greatly simplified by rewriting

$P[u_1,u_2,\ldots,u_t]$ as $\prod_{j=1}^{t} P[u_j \mid u_1,u_2,\ldots,u_{j-1}]$, where $P[u_j \mid u_1,u_2,\ldots,u_{j-1}]$ is the conditional distribution of the jth removal, given the values of the previous removals.

Then

(iii) Choose the smallest value of k that produces a sufficient fit to the data, and take as the estimate of N the estimate associated with this value of k. The fit of the data $u_1, u_2, \ldots, u_t$ can be measured by the usual size α chi-square goodness of fit test. We have used $\alpha = 0.20$ because of the seriousness of Type II errors. If significance levels of all the tests (one for each value of k) are less than 0.20 we have chosen the value of k corresponding to the largest achieved significance level.

For a given value of k, the asymptotic sampling variance of $\hat{N}_{bh}$ is

$$\mathrm{Var}(\hat{N}_{bh}) = N\left\{\frac{\sum_{j=1}^{t}\pi_j}{1-\sum_{j=1}^{t}\pi_j} - \sum_{j=1}^{k-1}\frac{\bar{p}_j^{\,2}}{\pi_j(1-\bar{p}_j)} - \frac{(t-k+1)^2\bar{p}^2}{(1-\bar{p})\left[\prod_{j=1}^{k-1}(1-\bar{p}_j)+\sum_{j=1}^{t}\pi_j-1\right]}\right\}^{-1}.$$

An estimate $\hat{\mathrm{V}}\mathrm{ar}(\hat{N}_{bh})$ of $\mathrm{Var}(\hat{N}_{bh})$ is obtained by replacing $N,\bar{p},\bar{p}_1,\bar{p}_2,\ldots,\bar{p}_{k-1}$, $\pi_1,\pi_2,\ldots,\pi_t$ by their respective ML estimates.

Finally, we mention that, for a given value of k, it is possible that the experiment "fails," i.e., valid estimation of N is not possible. Recall that Seber and Whale (1970) provided a failure criterion for the estimator associated with Model M_b. (This model corresponds to the case $k = 1$.) Following their method of proof, it is easily shown that the failure criterion for any value of k is

$$\sum_{j=k}^{t}(t+k-2j)u_j \leq 0.$$

If the experiment fails for a value of k, the corresponding model is clearly excluded from those eligible for selection as the appropriate model for estimating N.

APPENDIX I

Discussion of Model M_{tbh}

In Model M_{tbh}, it is assumed that the ith animal in the population has, on the jth trapping occasion, both a unique probability of first capture p_{ij} and a unique probability of recapture c_{ij}. The model therefore requires tN parameters concerning first capture, $(t-1)N$ parameters concerning recapture (because $c_{11} = c_{21} = \ldots = c_{N1} = 0$), and the parameter N for its complete specification. This totals $(2t-1)N + 1$ parameters and obviously all parameters are not identifiable for estimation purposes.

The assumption can be made that the $2t-1$ dimensional vectors $(p_{11}, p_{12},\ldots,p_{1t},c_{12},c_{13},\ldots,c_{1t})$, $(p_{21},p_{22},\ldots,p_{2t},c_{22},c_{23},\ldots,c_{2t}),\ldots,$ $(p_{N1},p_{N2},\ldots,p_{Nt},c_{N2},\ldots,c_{Nt})$ are a random sample of size N from some probability distribution function $F(p_1,p_2,\ldots,p_t,c_2,c_3,\ldots,c_t;\boldsymbol{\theta})$ parameterized by the vector $\boldsymbol{\theta}$ and de-

fined on $[0,1]^{2t-1}$. The probability distribution function of $\{X_\omega\}$ can then be written as

$$P[\{X_\omega]\} = \frac{N!}{\left[\prod_\omega X_\omega!\right](N - M_{t+1})!} \prod_\omega \left[E[\pi_\omega]\right]^{X_\omega},$$

where π_ω = a scalar random variable that is a function of the $2t - 1$ dimensional random variable $(p_1,p_2,\ldots,p_t,c_2,c_3,\ldots,c_t)$ corresponding to the capture history ω, and

$$E[\pi_\omega] = \int_0^1 \cdots \int_0^1 \pi_\omega \, dF(p_1,\ldots,p_t,c_2,\ldots,c_t;\boldsymbol{\theta}).$$

For instance, if $t = 4$ and $\omega = \{1,1,0,1\}$, then $\pi_\omega = p_1c_2(1 - c_3)c_4$. If one is willing to make certain assumptions concerning the family of distributions to which $F(\cdot;\boldsymbol{\theta})$ belongs, the dependence structure among the variables $\{p_1,p_2,\ldots,p_t,c_2,c_3,\ldots,c_t\}$ and the vector of parameters $\boldsymbol{\theta}$, then ML estimation of N will theoretically be possible. However, it is not unreasonable to suspect that the amount of numerical computation and the number of assumptions required will prohibit such estimates from being useful in practice.

APPENDIX J

Estimation in Removal Models

For the removal experiment, it is assumed there may be heterogeneity among the capture probabilities of the N members of the population subject to removal. Moreover, the N capture (removal) probabilities $p_i, i = 1,2,\ldots,N$, are a random sample from some probability distribution function $G(p;\boldsymbol{\theta})$ defined on $[0,1]$ and parameterized by the vector $\boldsymbol{\theta}$. Under those assumptions, the distribution of the vector of removals $\{u_1,u_2,\ldots,u_t\}$ is given by

$$P[u_1, u_2, \ldots, U_t] = \frac{N!}{\left[\prod_{j=1}^{t} u_j!\right](N - M_{t+1})!} \left(\prod_{j=1}^{t} \pi_j^{u_j}\right) \pi_{t+1}^{N-M_{t+1}},$$

where

$$\pi_j = E[p(1 - p)^{j-1}] = \int_0^1 p(1 - p)^{j-1} \, dG(p;\boldsymbol{\theta}), \qquad j = 1,2,\ldots,t,$$

$$\pi_{t+1} = E[(1 - p)^t] = \int_0^1 (1 - p)^t \, dG(p;\boldsymbol{\theta})$$

$$= 1 - \sum_{j=1}^{t} E[p(1 - p)^{j-1}].$$

Notice that this distribution of $\{u_1,u_2,\ldots,u_t\}$ is of the form, and the parameters of the same nature as the distribution of $\{u_1,u_2,\ldots,u_t\}$ in Model M_{bh}. Therefore, the generalized removal method developed for Model M_{bh} is also applicable to removal experiments where estimation of N is desired. Although the experimental situations associated with those 2 models are quite different, the fact that the removals are the only statistics relevant for purposes of estimating N, and the nature of the parameters π_j combine to make the generalized removal method appropriate in both cases.

APPENDIX K

Tests of Model Assumptions

The details of the statistical tests described in the section entitled TESTS OF MODEL ASSUMPTIONS are given here. Tests have been numerically identified and correspond to the identification numbers used in program CAPTURE.

Test 1

Since Model M_o is a special case of Model M_h, a likelihood ratio test of H_O: $p_i = p$, $i = 1, \ldots, N$ versus H_A: all p_i are not equal, seems plausible. However, due to the nonidentifiability of parameters in Model M_h a valid likelihood ratio procedure is not possible.

An alternative approach is taken by examining the fit of the observed frequencies $f_1, \ldots, f_t$ (recall these are the elements of the MSS for M_h) to their expected values under Model M_o. The resulting test statistic is intended to be sensitive to departures from Model M_o in the direction of Model M_h. If H_O is true, we would expect the test statistic T_1 to be approximately distributed as a chi-square random variable with $t - 2$ degrees of freedom, where

$$T_1 = \sum_{j=1}^{t} \frac{(f_j - \hat{f}_j)^2}{\hat{f}_j},$$

$$\hat{f}_j = \hat{N}_o \binom{t}{j} \hat{p}^j (1 - \hat{p})^{t-j},$$

and $\hat{N}_o$ and $\hat{p}$ are the ML estimates of N and p under Model M_o.

Test 2

Testing the null hypothesis of Model M_o versus the alternative of Model M_b is equivalent to testing the null hypothesis H_O: $p = c$ versus the alternative H_A: $p \neq c$. If it is assumed that the bivariate random variable $\{\hat{p}\ \hat{c}\}$ is distributed as a bivariate normal with mean vector $\{p\ c\}$ and covariance matrix

$$\begin{bmatrix} \mathrm{Var}(\hat{p}) & 0 \\ 0 & \mathrm{Var}(\hat{c}) \end{bmatrix},$$

then, under H_O, the quantity $T_2' = (\hat{p} - \hat{c})^2/[\mathrm{Var}(\hat{p}) + \mathrm{Var}(\hat{c})]$ is distributed as a chi-square random variable with 1 degree of freedom (cf. Theorem 4.4.5 in Graybill 1976). Here, $\hat{p}$ and $\hat{c}$ are the ML estimates of p and c under Model M_b (cf. Appendix D), and we use $\mathrm{Var}(\hat{p}) = p^2q^2(1 - q^t)/N[q(1 - q^t)^2 - p^2t^2q^t]$, where $q = 1 - p$ (cf. Seber 1973:312). Furthermore, we approximate $\mathrm{Var}(\hat{c})$ by treating $\hat{c}$ as a binomial variable with M. fixed so that $\mathrm{Var}(\hat{c}) = c(1 - c)/M$.. Obviously, both $\mathrm{Var}(\hat{p})$ and $\mathrm{Var}(\hat{c})$ will have to be estimated by substituting the ML estimates of N, p, and c under Model M_b, so that the actual test statistic becomes

$$T_2 = \frac{(\hat{p} - \hat{c})^2}{\hat{\mathrm{Var}}(\hat{p}) + \hat{\mathrm{Var}}(\hat{c})}.$$

It follows that T_2 has an approximate chi-square distribution with 1 degree of freedom under H_O.

If H_0 is false, T_2 has an approximate noncentral chi-square distribution with 1 degree of freedom and noncentrality parameter $\lambda = (p - c)^2/2[\mathrm{Var}(\hat{p}) + \mathrm{Var}(\hat{c})]$ (cf. Theorem 4.4.5 in Graybill 1976). Thus, theoretical approximations to the power of the above test can be found for fixed values of N, p, and c. These approximations, obtained by the use of the noncentral chi-square tables of Haynam et al. (1970), are given for the alternatives involved in Table 8, Appendix N. Finally, it should be mentioned that, due to the conditional nature of $\mathrm{Var}(\hat{c})$, it was necessary to substitute

$$E[M.] = Np \sum_{k=1}^{t} \sum_{\ell=0}^{k-2} q^{\ell}$$

for M. to enable evaluation of the noncentrality parameter λ.

Test 3

Testing the null hypothesis of Model M_0 versus the alternative of Model M_t is equivalent to testing H_0: $p_j = p$, $j = 1,2,\ldots,t$, against H_A: Not all the p_j are equal. We assume that the t-variate random variable $(p_1,p_2,\ldots,p_t)$ is distributed as a t-variate normal distribution with mean vector $(p_1,p_2,\ldots,p_t)$ and covariance matrix

$$\begin{bmatrix} \mathrm{Var}(\hat{p}_1) & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathrm{Var}(\hat{p}_t) \end{bmatrix}, \text{ where}$$

$$\hat{p}_j = n_j/\hat{N}_t = \text{ML estimator of } p_j \text{ under Model } M_t, \text{ and}$$

$$\mathrm{Var}(\hat{p}_j) = p_j q_j/N, \qquad q_j = 1 - p_j, \qquad j = 1,2,\ldots,t.$$

Now, using Theorem 4.4.5 in Graybill (1976), the random variable

$$T_3' = \sum_{j=1}^{t} \hat{p}_j^2/\mathrm{Var}(\hat{p}_j) - \left(\sum_{j=1}^{t} \hat{p}_j/\mathrm{Var}(\hat{p}_j)\right)^2 \Big/ \sum_{j=1}^{t} \mathrm{Var}(\hat{p}_j)^{-1}$$

has a chi-square distribution with $t - 1$ degrees of freedom under H_0. As in test 2, the quantities $\mathrm{Var}(\hat{p}_j)$ will have to be estimated using the ML estimates $\hat{p}_j$ and $\hat{N}_t$. Upon making these substitutions the test statistic T_3' reduces to

$$T_3 = \sum_{j=1}^{t} n_j/\hat{q}_j - \left(\sum_{j=1}^{t} n_j/\hat{p}_j\hat{q}_j\right)^2 \Big/ \sum_{j=1} n_j/\hat{p}_j^2\hat{q}_j.$$

Under H_0, T_3 has an approximate chi-square distribution with $t - 1$ degrees of freedom.

An approximation to the power of this test given any alternative is provided by the knowledge that, under H_A, T_3 has an approximate noncentral chi-square distribution with $t - 1$ degrees of freedom and noncentrality parameter

$$\lambda = (N/2)\left[\sum_{j=1}^{t} p_j/q_j - \left(\sum_{j=1}^{t} 1/q_j\right)^2 \Big/ \sum_{j=1}^{t} (p_j q_j)^{-1}\right].$$

Utilizing the tables of Haynam et al. (1970), these approximations were calculated for the alternatives involved in Table 9 of Appendix N, and the results are contained therein.

Test 4

An overall goodness of fit test of Model M_h can be thought of as equivalent to testing H_0: $p_{ij} = p_i$ versus H_A: not all $p_{ij} = p_i$, $i = 1,2,\ldots,M_{t+1}$, $j = 1,\ldots,t$. Thus, the null hypothesis assumes heterogeneous individual capture probabilities that do not change over time, and the alternative states that, given heterogeneity, capture probabilities also change over time. Notice that H_A does not state how probabilities change over time, i.e., whether the change is due to behavioral response, variation in trapping occasions, or a combination of both. Thus, if H_0 is rejected, it is not rejected in favor of an alternative Model; rather it is the goodness of fit of Model M_h that is being rejected.

Burnham (unpublished dissertation) has shown that under H_0 the test statistic

$$T_4 = \frac{\sum_{j=1}^{t} (n_j - n./t)^2}{\sum_{j=1}^{t} f_j\left(\frac{j}{t}\right)\left(1 - \frac{j}{t}\right)} \cdot \frac{t-1}{t}$$

has an approximate chi-square distribution with $t - 1$ degrees of freedom and is appropriate for testing H_0 vs. H_A. The test is conditional on the frequency of capture statistics $f_1, f_2, \ldots, f_t$. Burnham also recommended that if f_k is large enough, a statistic appropriate for testing H_0: $p_{ij} = p_i$ versus H_A: not all $p_{ij} = p_i$, for all i such that $y_i = k$, is given by

$$T_{4a} = \frac{\sum_{j=1}^{t} \left(z_{kj} - \frac{k}{t} f_k\right)^2}{\frac{k}{t} f_k} \cdot \frac{t-1}{t-k},$$

where z_{kj} = number of animals caught on day j that were captured exactly k times, and
y_i = number of times the ith animal was captured.

Under H_0, T_{4a} has an approximate chi-square distribution with $t - 1$ degrees of freedom, conditional on the value of f_k. Notice that a test statistic of the form of T_{4a} can be constructed for any $k = 1,2,\ldots,t-1$ as long as f_k is large enough. We have used the criterion that f_k must be larger than t before the test is performed.

Test 5

An overall goodness of fit test of Model M_b can be constructed by combining the results of 2 independent tests. The first of these tests was introduced by Zippin (1956) for testing H_0: $p_j = p$ versus H_A: Not all $p_j = p$, $j = 1,2,\ldots,t$. In the context of Model M_b, p_j represents the probability of first capture on the jth trapping occasion, and thus Zippin's test attempts to determine the constancy of first capture probability over time. The test statistic, which has an approximate chi-square distribution with $t - 2$ degrees of freedom when H_0 is true, is given as

$$T_{5a} = \sum_{j=1}^{t} \frac{(u_j - \hat{N}_b\hat{p}\hat{q}^{j-1})^2}{\hat{N}_b\hat{p}\hat{q}^{j-1}} + \frac{(\hat{N}_b - M_{t+1} - \hat{N}_b\hat{q}^t)^2}{\hat{N}_b\hat{q}^t},$$

where $\hat{N}_b$ and $\hat{p}$ are the ML estimates of N and p under Model M_b, and $\hat{q} = 1 - \hat{p}$. (Note: we have used the following "pooling" strategy for this chi-square test. If $\hat{N}_b\hat{p}\hat{q}^{r-1} < 2$ and $\hat{N}_b\hat{p}\hat{q}^{s-1} \geqslant 2$ for $s = 1,2,\ldots,r-1$, then the "cells" corresponding to $r, r+1, \ldots, t$ are pooled into 1 cell. This will reduce the degrees of freedom associated with T_{5a} to $r - 2$.)

A test (independent of the above procedure) for the constancy of recapture probability over time can be constructed using the so-called variance test for homogeneity of binomial proportions (cf. Snedecor and Cochran 1967: 240). That is, a test statistic appropriate for testing H_0: $c_j = c$ versus H_A: not all $c_j = c$, $j = 2,3,\ldots,t$, is given by

$$T_{5b} = \sum_{j=2}^{t} M_j(\hat{c}_j - \hat{c})^2/\hat{c}(1 - \hat{c}),$$

where

$\hat{c}_j = m_j/M_j$,
$\hat{c}$ = ML estimate of c under $M_b = m./M.$.

The statistic T_{5a} has an approximate chi-square distribution with $t - 2$ degrees of freedom when H_0 is true.

Because T_{5a} and T_{5b} are independent and have chi-square distributions when Model M_b is true, an overall test statistic for the goodness of fit of Model M_b is given by $T_5 = T_{5a} + T_{5b}$. This test statistic has an approximate chi-square distribution with $2t - 4$ degrees of freedom under Model M_b.

Test 6

Testing the goodness of fit of Model M_t can be interpreted as being equivalent to testing H_0: $p_{ij} = p_j$ versus H_A: not all $p_{ij} = p_j$, $i = 1,2,\ldots,M_{t+1}$ and $j = 1,2,\ldots,t$. Thus, the null hypothesis assumes variation in capture probabilities among trapping occasions while asserting that on a given occasion all animals have the same probability of capture. The alternative H_A allows capture probabilities to differ among members of the population on a given occasion, as well as allowing trapping occasions to affect capture probabilities. Notice that H_A does not specify why capture probabilities differ among animals on a given occasion. That is, this difference may be due to behavioral response, individual heterogeneity, or a combination of both. Thus, when H_0 is rejected, it is not in favor of a specific alternative model; rather it is the goodness of fit of M_t that is being rejected. We have chosen to adopt the test procedure proposed by Leslie (1958) for the purpose of testing H_0 versus H_A. The proper test statistic is given by

$$T_6 = \sum_{j=1}^{t-3}\left[\sum_{k=1}^{t-j+1} \frac{f_k^{(j)}(k - 1 - \hat{\mu}_j)^2}{\hat{\mu}_j - \sum_{\ell=j+1}^{t} R^2_{j\ell}/u_j^2}\right] I_j,$$

where

$f_k^{(j)}$ = number of animals captured exactly k times that were first captured on the jth occasion,
u_j = number of animals first captured on the jth occasion,
$R_{j\ell}$ = number of animals recaptured on the ℓth occasion that were first caught on the jth occasion,

$$\hat{\mu}_j = \sum_{k=1}^{t-j+1} f_k^{(j)} (k-1)/u_j, \text{ and}$$

$$I_j = \begin{cases} 1 \text{ if } u_j \geqslant 20 \\ 0 \text{ otherwise.} \end{cases}$$

Notice that we have followed Leslie's recommendations that a "cohort" of newly identified animals first captured on the jth occasion not be included in the overall test unless it consists of more than 20 individuals, i.e., u_j must be greater than 20. Furthermore, a new cohort is not included unless it is subject to at least 3 subsequent trapping occasions, i.e., $j = 1,2,\ldots,t-3$. Under the null hypothesis of "equicatchability," T_6 has an approximate chi-square distribution with

$$\sum_{j=1}^{t-3} (u_j - 1) I_j$$

degrees of freedom.

It should be mentioned that Carothers (1971) proposed an improved version of Leslie's test. However, practical use of the procedure requires some arbitrary trimming of the data and therefore is difficult to use for simulation purposes.

Test 7

Pollock (unpublished dissertation) developed a procedure designed to test H_0: Model M_h fits the data versus H_A: Model M_{bh} should be used. The test statistic depends on the 2 vectors of statistics

$\mathbf{f}^* = \{f_1^{(1)}, f_1^{(2)}, \ldots, f_1^{(t)}, \ldots, f_{t-1}^{(1)}, f_{t-1}^{(2)}, f_t^{(1)}\}$ where $f_k^{(j)}$ is the number of animals captured exactly k times that were first captured on the jth occasion,

$\mathbf{f} = \{f_1, f_2, \ldots, f_t\}$, where f_j is the number of animals captured exactly j times.

This is a chi-square goodness of fit test formed by pooling $t-1$ independent chi-square tests. The kth of these tests is conditional on the value of f_k and has $t-k$ degrees of freedom. The overall test statistic is given by

$$T_7 = \sum_{k=1}^{t-1} \sum_{j=1}^{t-k+1} \frac{\left\{ f_k^{(j)} - \left[\binom{t-j}{t-k-j+1} \Big/ \binom{t}{k} \right] f_k \right\}^2}{\left[\binom{t-j}{t-k-j+1} \Big/ \binom{t}{k} \right] f_k}.$$

Under H_0, T_7 has an approximate chi-square distribution with

$$\sum_{k=1}^{t-1} (t-k) = t(t-1)/2$$

degrees of freedom. (Note: for each of the $t-1$ distributions we have used the same pooling strategy described for the test involving T_{5a}. In this case, the quantity checked for sufficiently large expectation is

$$\left[\binom{t-j}{t-k-j+1} \Big/ \binom{t}{k} \right] f_k.$$

Therefore, it may be that in a given case degrees of freedom will not be $t(t-1)/2$, but rather this quantity less the appropriate number of degrees of freedom lost by pooling.)

A Test for Closure

An approach to a closure test suggested by Burnham and Overton (pers. comm.) can be conceptualized by first considering the null hypothesis H_0: $p_{ij} = p_i$, $j = 1, 2, \ldots, t$ for only those animals captured 2 or more times. Essentially, H_0 merely asserts that individual capture probabilites are invariant over time. Now consider, however, an alternative hypothesis H_A that states that for at least some i, i.e., for some animals that were captured at least twice, $p_{i1} = p_{i2} = \cdots = p_{ir} = 0$ and/or $p_{is} = \cdots = p_{it} = 0$, where $1 \leq r < s \leq t$. H_A states that some members of the population were not present in the population for some initial or terminal part of the study, or both. If that alternative is true, one might expect that the time between first and last capture for animals is, on the average, shorter than one could expect under H_0. That conjecture is the rationale for the test procedure given below. Note that the test is designed to detect birth–death or immigration–emigration phenomena, or both, that occur only during the initial and latter stages of the study, not phenomena occurring toward the middle of the study. For example, the test would not be appropriate when some animals are present during the initial stages of the study, leave the study area for a time, and then return prior to the termination of the study.

Given that the ith animal was captured exactly y_i times, and that $y_i \geq 2$, let $Q_i = W_i - V_i$, where W_i is the occasion of last capture and V_i is the occasion of first capture. Thus, Q_i is merely the time between first and last capture for the ith animal. Conditional upon the value of y_i, the expectation and variance of Q_i are:

$$E(Q_i | y_i = k) = (k-1)(t+1)/(k+1),$$
$$\mathrm{Var}(Q_i | y_i = k) = 2(t-k)(k-1)(t+1)/(k+2)(k+1)^2.$$

Under H_0, the statistic

$$\hat{E}(Q|k) = \frac{1}{f_k} \sum_{i=1}^{M_{t+1}} Q_i I_i$$

has the conditional expectation given above

where $I_i = \begin{cases} 1 \text{ if the } i^{th} \text{ animal was captured exactly k times} \\ 0 \text{ otherwise.} \end{cases}$

If f_k is large, the test statistic,

$$C_k = \frac{\hat{E}(Q|k) - (k-1)(t+1)/(k+1)}{\left[\dfrac{2(t-k)(k-1)(t+1)}{(k+2)(k+1)^2 f_k}\right]^{1/2}}, \qquad k = 2, \ldots, t-1,$$

can be assumed to be approximately distributed as a standard normal. (We have required that $f_k \geq 10$.) An overall test statistic appropriate for testing H_0 versus H_A is given as

$$C = \frac{\sum_{k=2}^{t-1} [\hat{E}(Q|k) - (k-1)(t+1)/(k+1)]}{\left[\sum_{k=2}^{t-1} \frac{2(t-k)(k-1)(t+1)}{(k+2)(k+1)^2 f_k}\right]^{1/2}}.$$

The test statistic C is also approximately distributed as a standard normal if H_0 is true and the number of animals captured at least twice is large (greater than 10). Tests of H_0 versus H_A calculated by using either C_k or C should be one-sided tests since the alternative specifies that the $\hat{E}(Q|k)$ should be smaller than $E(Q_i|y_i = k)$. Thus, closure is rejected only if the test statistic is small. Finally, we emphasize that the test involving C_k is conditional on the value of f_k, and the test using C is conditional on the values of $f_2, f_3, \ldots, f_{t-1}$.

APPENDIX L

Density Estimation Based on Subgrids

The density estimation procedure using nested subgrids is based on Dice's (1938) boundary strip idea. The fundamental aspect of this approach was proposed by MacLulich (1951). Let the trapping grid have area A_g and a perimeter length of P. Then for any convex grid (that includes all rectangular grids) the effective trapping area A(W) is

$$A(W) = A_g + PW/c + \pi W^2/c,$$

where c is a conversion factor to express PW and W^2 in the units of A. From that equation, we derive the expected population size at risk of capture as

$$E(N) = DA(W),$$

where D is the true density of animals. Dividing through by the known area A_g we derive

$$\frac{E(N)}{A_g} = D[1 + aW + bW^2],$$

where

$$a = \frac{P}{A_g(c)}, \qquad b = \frac{\pi}{A_g(c)}.$$

Note that the unknown parameters are D and W and that E(N) is estimable from the trapping study.

Given at least 2 grids of different sizes, we can estimate the parameters D and W. Assume there are k different grids (these may be subgrids of 1 overall study). The relevant equations can be written as

$$\frac{N_i}{A_g} = D[1 + a_i W + b_i W^2] + \epsilon_i, \qquad i = 1, \ldots, k.$$

By assumption, $E(\boldsymbol{\epsilon}) = \mathbf{0}$. Let the variance covariance matrix of $\boldsymbol{\epsilon}$ be Σ. We are assuming the same density (D) and strip width (W) apply to all grids. This seems especially reasonable when the grids are nested (see section on DENSITY ESTIMATION).

Given estimates $\hat{N}_i$, the above equations put estimation of D and W in the framework of generalized nonlinear regression. All we need to carry out the estimation is a knowledge of the variance–covariance matrix Σ. Letting $Y_i = \hat{N}_i/A_g$, we have

$$\mathrm{Var}(Y_i) = \mathrm{Var}(\hat{N}_i)/A_g^2,$$

which is estimable. Also we know that

$$\mathrm{Cov}(Y_i,Y_j) = r_{ij}\mathrm{SE}(\hat{Y}_i)\mathrm{SE}(\hat{Y}_j),$$

where r_{ij} is the correlation of $\hat{N}_i$ and $\hat{N}_j$. This correlation is not known. Burnham and Cushwa (pers. comm.) suggested the following way of approximating r_{ij}.

The correlation of $\hat{N}_i$ and $\hat{N}_j$ will depend in large part on the overlap of the 2 populations of size N_i and N_j that in turn depends on the overlap of the areas $A_i(W)$ and $A_j(W)$. Thus let

$$r_{ij} = \mathrm{corr}(\hat{N}_i,\hat{N}_j) = \frac{A_i(W) \cap A_j(W)}{A_i(W) \cup A_j(W)},$$

that is, r_{ij} is the area of the intersection of $A_i(W)$ and $A_j(W)$ divided by the area of their union.

Using this formula, we can arrive at an estimator of Σ that seems reasonable; hence, from the live trapping data we can compute $\hat{Y}_1, \ldots, \hat{Y}_k$ and $\hat{\Sigma}$.

We are now in a position to obtain weighted nonlinear least squares estimators of D and W as $\hat{D}$, $\hat{W}$ satisfying

$$\min_{D,\,W} (\hat{\mathbf{Y}} - \mathbf{f})'\hat{\Sigma}^{-1}(\hat{\mathbf{Y}} - \mathbf{f}), \qquad \text{where}$$

$$f_i = D[1 + a_iW + b_iW^2], \qquad i = 1,\ldots,k.$$

An approximate variance–covariance matrix for the estimators $\hat{D}$ and $\hat{W}$ is given by the 2×2 matrix $V'\hat{\Sigma}^{-1}V$ where V is the Jacobian matrix

$$V = \left[\frac{\partial \mathbf{f}}{\partial D}, \frac{\partial \mathbf{f}}{\partial W}\right].$$

The elements of the columns of the matrix V are, respectively,

$$\frac{\partial f_i}{\partial D} = 1 + a_iW + b_iW^2,$$

$$\frac{\partial f_i}{\partial W} = D[a_i + b_i2W], \qquad i = 1,\ldots,t.$$

Because $\hat{\Sigma}$ depends upon W (but not D), an iterative procedure is needed wherein an initial value of $\hat{W}_0$ is chosen, $\hat{\Sigma}$ is computed based on it, and the new $\hat{D}_1$, $\hat{W}_1$ obtained. Iteration is continued until stable estimates of D and W are obtained.

APPENDIX M

General Simulation Methods

The various estimators and tests described in this monograph were simulated to study their small sample properties and operating characteristics. The asymptotic properties of the tests and estimators are known in most cases, but not the finite sample properties. Monte Carlo simulations provide a method by which our estimation procedures may be evaluated under the exact model from which they were derived. Because we would not expect real data to fit any model exactly, by simulating data, we can study the properties of the procedures without interference due to the data not fitting the model.

Pseudo-random uniform (0,1) variables were generated using the Burroughs 6700 FORTRAN intrinsic function RANDOM. The mixed congruential method is used by this function (Anonymous 1971).

Beta variables for the simulation of heterogeneous probabilities of first capture were generated using subroutine GGBTA from the IMSL (1976) package, with the required uniform variables furnished by RANDOM. That routine uses a rejection method (Fishman 1973).

In the most general simulations, the trapping process is simulated for each animal in the population for each trapping occasion. The probability of capture for a particular animal on a particular day is compared against a uniform (0,1) variable. If the value of the random variable is less than the probability of capture, then the animal is assumed captured, and that element of the **X** matrix is set to unity. Otherwise, the value in the **X** matrix is set to zero, indicating no capture. When the **X** matrix is completely filled, the necessary MSS are computed and the tests and estimation completed.

APPENDIX N

Simulation Results

Simulation Results Regarding Estimation Procedures

In developing this material, we used 6 different models to generate simulated data: all capture–recapture models except M_{tb} and M_{tbh}. For each of the data sets, various estimation procedures considered here were applied to gain insight into the operating characteristics of such procedures. The results of this simulation study are presented here in tabular form.

For each model there are 2 corresponding tables. Obviously, for any given model one may choose any number of sets of parameter values needed to completely specify the model. Thus, the first table for each model lists all the different sets of parameter values (each of which is called a Trial) used in the simulation study. For example, Trial 1 in Table N.1.a indicates that there were 400 animals in the population and that every animal had a 0.30 probability of capture on each trapping occasion. That information is all that is required to specify an example of Model M_o.

The second table for each model presents the simulation results for each estimation procedure used on data generated from the Trials of that model. Column headings in the tables are:

Estimator—identifies the estimation procedure used; e.g., "$\hat{N}_b$" indicates that the estimation procedure described in Appendix C for Model M_b was applied to the data.
R—the number of replications (data sets).
t—the number of trapping occasions.
Ave($\hat{N}$)—the average value of $\hat{N}$ taken over all replications, i.e.,

$$\text{Ave}(\hat{N}) = \frac{1}{R}\sum_{k=1}^{R}\hat{N}_k.$$

$\sigma(\hat{N})$—the "sample" standard error of $\hat{N}$ taken over all replications, i.e.,

$$\sigma(\hat{N}) = \sqrt{\sum_{k=1}^{R}(\hat{N}_k - \text{Ave}(\hat{N}))^2/(R-1)}.$$

Ave$\sqrt{\hat{\text{Var}}(\hat{N})}$—the average value of $\sqrt{\hat{\text{Var}}(\hat{N})}$ taken over all replications, i.e.,

$$\text{Ave}\sqrt{\hat{\text{Var}}(\hat{N})} = \frac{1}{R}\sum_{k=1}^{R}\sqrt{\hat{\text{Var}}(\hat{N}_k)}.$$

Coverage—the proportion of replications in which the constructed 95% confidence interval contained the true population size N.
Trial—identifies which population (set of parameter values) generated the data.

Two points concerning $\sigma(\hat{N})$ and Ave$\sqrt{\hat{\text{Var}}(\hat{N})}$ should be mentioned. First the value of $\sigma(\hat{N})$ estimates the true standard error of the estimator $\hat{N}$ in any given Trial and Model so that a measure of the relative bias of the estimator $\sqrt{\hat{\text{Var}}(\hat{N})}$ is given by $|\text{Ave}\sqrt{\hat{\text{Var}}(\hat{N})} - \sigma(\hat{N})|/\sigma(\hat{N})$. Second, the average confidence interval width in a given Trial and Model is given by $2\cdot(1.96)$ Ave$\sqrt{\hat{\text{Var}}(\hat{N})}$.

As an example, consider the first line of Table N.1.b. We see that 200 data sets, each consisting of 5 trapping occasions were generated according to Model M_0, Trial 1 and that the estimation procedure associated with Model M_0 was applied to each set. For those 200 data sets, $\hat{N}_0$ was essentially unbiased (Ave($\hat{N}_0$) − N = −0.7), as was the estimator of the standard error of $\hat{N}_0$ (Ave$\sqrt{\hat{\text{Var}}(\hat{N}_0)} - \sigma(\hat{N}_0) = 0.05$). Moreover, 95 percent of the constructed confidence intervals covered the true value of N. This achieved confidence coefficient of 0.95 corresponds exactly with the stated level of the interval, namely 0.95. Such excellent performance of the confidence interval procedure is due in large part to the fact that $\hat{N}_0$ and $\sqrt{\hat{\text{Var}}(\hat{N}_0)}$ appear to be essentially unbiased. Finally, note that the average width of the 200 constructed confidence intervals is given by 2(1.96)(11.96) = 46.88.

Simulation Results Regarding the Size and Power of Testing Procedures

Because the testing of assumptions plays a vital role in the analysis of capture–recapture data, it is important to have some insight into the operating characteristics of such tests. Simulation studies of the tests T_1 through T_7 defined in Appendix K were carried out on data simulated from all 8 models to provide such insight, and the results are presented in Tables N.7 through N.13. In addition, Tables N.8 and N.9 give results of an approximation to the power of tests T_2 and T_3 based on theoretical results in Appendix K.

The column headings of the tables are:

α—indicates the nominal size of the test.
Data Model—indicates the Trial and Model that were used to generate the data.
R—indicates number of replications (data sets).
t—indicates number of trapping occasions.

The reader will note that in some instances the number of replications R appears to be strange. The explanation is that a few of the estimators and tests simulated are subject to "failure," i.e., a particular data set may have characteristics that do not allow calculation of the desired statistics. In those cases, the data sets are excluded from the reported simulation results.

Finally, in Table N.14 we give a description of the Trials of Models M_{tb} and M_{tbh}. Those Trials were involved in the simulation of some of the tests of model assumptions.

TABLE N.1.a.—DESCRIPTION ON THE TRIALS OF MODEL M_o

N	p	Trial
400	0.30	1
400	0.10	2
400	0.05	3
200	0.25	4
100	0.20	5
100	0.15	6

TABLE N.1.b.—SIMULATION RESULTS OF ESTIMATION PROCEDURES USED ON DATA GENERATED ACCORDING TO MODEL M_o

Estimator	N	Ave($\hat{N}$)	$\sigma(\hat{N})$	Ave$\sqrt{\hat{V}ar(\hat{N})}$	Coverage	R	t	Trial
$\hat{N}_o$	400	399.30	11.91	11.96	0.950	200	5	1
$\hat{N}_o$	400	456.93	219.31	160.25	0.922	500	5	3
$\hat{N}_o$	400	405.97	52.84	55.54	0.930	200	5	2
$\hat{N}_o$	100	101.70	19.43	–	–	200	5	6
$\hat{N}_o$	400	397.34	46.56	–	–	200	5	2
$\hat{N}_o$	400	453.82	203.75	–	–	372	5	3
$\hat{N}_o$	400	405.73	52.54	–	–	200	5	2
$\hat{N}_o$	200	199.60	6.58	–	–	100	7	4
$\hat{N}_h$	100	113.46	17.25	13.74	0.830	200	5	6
$\hat{N}_h$	400	386.89	27.65	27.93	0.920	200	5	2
$\hat{N}_h$	400	389.12	29.56	28.02	0.898	400	5	2
$\hat{N}_h$	200	226.38	10.77	10.07	0.260	100	7	4
$\hat{N}_b$	400	319.25	204.51	–	–	372	5	3
$\hat{N}_b$	400	485.90	387.42	–	–	198	5	2
$\hat{N}_b$	200	198.59	11.66	10.47	0.880	100	7	4
$\hat{N}_t$	400	407.71	61.81	57.45	0.940	400	5	2
$\hat{N}_t$	400	439.96	161.42	148.13	0.940	500	5	3
$\hat{N}_t$	200	199.37	6.51	6.85	0.950	100	7	4

TABLE N.2.a.—DESCRIPTION OF THE TRIALS OF MODEL M_t

N	$(p_1, p_2, \ldots, p_t)$	Trial
400	(0.55, 0.55, 0.50, 0.45, 0.45)	1
400	(0.05, 0.10, 0.15, 0.15, 0.05)	2
400	(0.10, 0.10, 0.10, 0.10, 0.01)	3
400	(0.01, 0.01, 0.02, 0.03, 0.03)	4
400	(0.04, 0.05, 0.03, 0.07, 0.06)	5
400	(0.60, 0.40, 0.20, 0.10, 0.10, 0.10, 0.10, 0.10, 0.10, 0.10)	6
400	(0.50, 0.20, 0.10, 0.10, 0.10, 0.10, 0.10)	7
400	(0.60, 0.40, 0.20, 0.10, 0.10)	8
400	(0.20, 0.40, 0.30, 0.10, 0.20, 0.30, 0.20)	9
200	(0.30, 0.40, 0.10, 0.40, 0.30)	10
800	(0.02, 0.01, 0.03, 0.03, 0.01)	11
100	(0.05, 0.05, 0.10, 0.15, 0.15)	12

TABLE N.2.b.—SIMULATION RESULTS OF ESTIMATION PROCEDURES USED ON DATA GENERATED ACCORDING TO MODEL M_t

Estimator	N	Ave($\hat{N}$)	$\sigma(\hat{N})$	Ave $\sqrt{\hat{V}ar(\hat{N})}$	Coverage	R	t	Trial
$\hat{N}_t$	400	399.30	202.39	348.63	0.839	155	5	4
$\hat{N}_t$	400	408.57	73.91	74.19	0.949	198	5	3
$\hat{N}_t$	100	108.44	37.18	35.77	0.942	989	5	12
$\hat{N}_t$	800	1,015.62	630.46	748.23	0.902	194	5	11
$\hat{N}_t$	400	400.33	54.50	56.00	0.955	200	5	2
$\hat{N}_t$	400	399.28	3.55	3.86	0.960	200	5	1
$\hat{N}_t$	400	442.66	187.00	151.93	0.930	200	5	5
$\hat{N}_t$	400	397.14	206.97	341.61	0.880	292	5	4
$\hat{N}_t$	400	398.59	15.64	16.45	0.940	100	7	7
$\hat{N}_t$	400	398.81	8.03	7.57	0.950	100	10	6
$\hat{N}_t$	200	198.70	8.98	7.98	0.880	100	5	10
$\hat{N}_t$	400	399.41	10.16	9.96	0.930	100	7	9
$\hat{N}_t$	400	399.45	12.80	11.94	0.940	100	5	8
$\hat{N}_h$	400	349.17	27.27	26.54	0.515	200	5	3
$\hat{N}_h$	400	100.83	27.37	12.93	0.000	200	5	4
$\hat{N}_h$	100	98.70	14.43	13.77	0.924	1,000	5	12
$\hat{N}_h$	800	219.60	27.94	20.79	0.000	200	5	11
$\hat{N}_h$	400	244.46	24.04	22.17	0.000	200	5	5
$\hat{N}_h$	400	98.56	30.27	12.43	0.000	400	5	4
$\hat{N}_h$	400	506.72	27.76	22.00	0.000	100	5	8
$\hat{N}_h$	400	516.22	31.59	26.91	0.000	100	7	7
$\hat{N}_h$	200	456.68	50.28	15.91	0.000	100	10	6
$\hat{N}_h$	400	235.94	15.78	12.54	0.170	100	5	10
$\hat{N}_h$	200	459.60	18.32	15.30	0.010	100	7	9
$\hat{N}_{bh}$[1]	400	199.70	21.08	18.08	0.810	100	5	10
$\hat{N}_{bh}$[1]	400	410.30	54.16	70.10	0.860	100	7	9
$\hat{N}_{bh}$[1]	400	387.07	21.14	27.63	0.540	100	10	6
$\hat{N}_{bh}$[1]	400	364.06	55.82	65.24	0.510	100	7	7
$\hat{N}_{bh}$[1]	400	342.56	9.48	5.72	0.040	100	5	7

[1] Note, we use $\hat{N}_{bh}$ and $\hat{N}_R$ as equivalent notation.

TABLE N.3.a.—DESCRIPTION OF THE TRIALS OF MODEL M_b

N	p	c	Trial
400	0.30	0.10	1
400	0.20	0.50	11
400	0.20	0.05	2
400	0.10	0.30	3
400	0.10	0.15	4
200	0.30	0.50	5
200	0.25	0.25	6
200	0.20	0.10	7
100	0.20	0.05	8
100	0.20	0.50	9
100	0.10	0.30	10
100	0.40	0.20	12

TABLE N.3.b.—SIMULATION RESULTS OF ESTIMATION PROCEDURES USED ON DATA GENERATED ACCORDING TO MODEL M_b

Estimator	N	Ave($\hat{N}$)	$\sigma(\hat{N})$	Ave $\sqrt{\hat{V}ar(\hat{N})}$	Coverage	R	t	Trial
$\hat{N}_b$	100	82.00	36.73	64.76	0.708	161	5	10
$\hat{N}_b$	200	208.09	38.53	40.27	0.930	100	5	7
$\hat{N}_b$	200	198.36	13.32	13.13	0.900	100	5	5
$\hat{N}_b$	400	401.23	25.43	19.91	0.920	100	5	1
$\hat{N}_b$	100	109.83	48.05	–	–	199	5	8
$\hat{N}_b$	400	400.22	22.38	–	–	200	5	1
$\hat{N}_b$	400	511.04	408.42	–	–	193	5	3
$\hat{N}_b$	400	403.56	58.70	–	–	200	5	2
$\hat{N}_b$	200	197.88	13.32	–	–	200	5	5
$\hat{N}_b$	400	461.48	269.25	–	–	196	5	4
$\hat{N}_b$	200	206.99	37.11	–	–	200	5	7
$\hat{N}_b$	200	198.59	11.66	10.47	0.880	100	7	6
$\hat{N}_h$	100	82.39	6.75	5.09	0.140	100	5	9
$\hat{N}_h$	100	176.49	14.40	18.82	0.000	100	5	8
$\hat{N}_h$	400	347.67	29.55	26.16	0.495	200	5	4
$\hat{N}_t$	100	71.75	4.80	2.43	0.000	100	5	9
$\hat{N}_t$	100	299.25	193.37	115.92	0.630	100	5	8
$\hat{N}_o$	100	51.38	7.06	–	–	161	5	10
$\hat{N}_o$	100	293.88	109.76	–	–	199	5	8
$\hat{N}_o$	400	720.39	57.05	–	–	200	5	1
$\hat{N}_o$	400	212.48	13.56	–	–	193	5	3
$\hat{N}_o$	400	1,070.94	176.01	–	–	200	5	2
$\hat{N}_o$	200	175.18	5.80	–	–	200	5	5
$\hat{N}_o$	400	305.73	33.32	–	–	196	5	4
$\hat{N}_o$	200	311.67	48.50	–	–	200	5	7

TABLE N.4.a.—DESCRIPTION OF THE TRIALS OF MODEL M_h

N	p_i, i = 1,2,...,N	Trial
400	$p_i = 0.05$, i = 1,200; $p_i = 0.15$, i = 201,300; $p_i = 0.50$, i = 301,400.	1
400	$p_i = 0.01$, i = 1,100; $p_i = 0.05$, i = 101,200; $p_i = 0.10$, i = 201,300; $p_i = 0.20$, i = 301,400.	2
400	$p_i = 0.10$, i = 1,100; $p_i = 0.20$, i = 101,200; $p_i = 0.25$, i = 201,300; $p_i = 0.30$, i = 301,400.	3
400	$p_i = 0.01$, i = 1,50; $p_i = 0.15$, i = 51,200; $p_i = 0.25$, i = 201,300; $p_i = 0.30$, i = 301,400.	4
400	$p_i = 0.20$, i = 1,100; $p_i = 0.30$, i = 101,200; $p_i = 0.40$, i = 201,300; $p_i = 0.50$, i = 301,400.	5
200	$p_i = 0.05$, i = 1,50; $p_i = 0.15$, i = 51,150; $p_i = 0.25$, i = 151,200.	6
200	$p_i = 0.15$, i = 1,50; $p_i = 0.20$, i = 51,100; $p_i = 0.25$, i = 101,150; $p_i = 0.30$, i = 151,200.	7
100	$p_i = 0.05$, i = 1,40; $p_i = 0.10$, i = 41,80; $p_i = 0.30$, i = 81,100.	8
100	$p_i \sim \beta(3,22)$, i = 1,100.[1]	9
100	$p_i \sim \beta(1,22/3)$, i = 1,100.[1]	10
100	$p_i \sim \beta(3/22,1)$, i = 1,100.[1]	11
100	$p_i = 0.10$, i = 1,40; $p_i = 0.20$, i = 41,80; $p_i = 0.30$, i = 81,100.	12
400	$p_i = 0.05$, i = 1,50; $p_i = 0.10$, i = 51,200; $p_i = 0.15$, i = 201,300; $p_i = 0.25$, i = 301,400.	13
200	$p_i = 0.05$, i = 1,50; $p_i = 0.10$, i = 51,150; $p_i = 0.25$, i = 151,200.	14

[1] Indicates that the probabilities p_i were a random sample of size 100 from a beta probability distribution with the indicated parameter values. E[p] = 0.12 in those cases.

TABLE N.4.b.—SIMULATION RESULTS OF ESTIMATION PROCEDURES USED ON DATA GENERATED ACCORDING TO MODEL M_h

Estimator	N	Ave($\hat{N}$)	$\sigma(\hat{N})$	Ave $\sqrt{\hat{V}ar(\hat{N})}$	Coverage	R	t	Trial
$\hat{N}_h$	100	35.41	9.36	4.38	0.000	200	5	11
$\hat{N}_h$	100	79.46	15.15	10.91	0.545	200	5	10
$\hat{N}_h$	100	84.66	14.32	11.56	0.675	200	5	8
$\hat{N}_h$	400	331.06	23.99	22.69	0.180	200	5	1
$\hat{N}_h$	400	298.06	24.52	24.22	0.000	200	5	2
$\hat{N}_h$	100	100.17	21.15	10.54	0.675	200	10	8
$\hat{N}_h$	400	460.10	32.31	26.77	0.395	200	5	3
$\hat{N}_h$	200	226.78	19.82	11.94	0.440	100	7	7
$\hat{N}_h$	400	443.95	19.71	15.20	0.170	100	5	5
$\hat{N}_h$	400	461.22	32.64	26.92	0.410	100	5	3
$\hat{N}_h$	400	417.08	30.91	24.92	0.850	100	5	4
$\hat{N}_h$	100	107.57	16.80	12.01	0.810	100	5	12
$\hat{N}_h$	200	207.34	16.14	11.59	0.870	100	10	6
$\hat{N}_o$	100	26.83	4.79	–	–	200	5	11
$\hat{N}_o$	100	62.68	12.39	–	–	200	5	10
$\hat{N}_o$	100	68.90	12.43	–	–	200	5	8
$\hat{N}_o$	400	227.83	10.97	–	–	200	5	1
$\hat{N}_o$	400	260.29	30.93	–	–	200	5	2
$\hat{N}_o$	100	73.72	7.28	–	–	200	10	8
$\hat{N}_o$	400	365.12	17.15	–	–	200	5	3
$\hat{N}_t$	200	193.61	9.70	7.69	0.770	100	7	7
$\hat{N}_t$	400	376.18	10.00	7.67	0.240	100	5	5
$\hat{N}_t$	400	364.31	16.44	17.01	0.460	100	5	3
$\hat{N}_t$	400	333.17	17.92	16.19	0.060	100	5	4
$\hat{N}_t$	100	87.87	12.07	10.19	0.690	100	5	12
$\hat{N}_t$	200	173.64	8.68	6.83	0.120	100	10	6
$\hat{N}_t$	400	228.02	10.73	7.40	0.000	100	5	1

TABLE N.5.a.—DESCRIPTION OF THE TRIALS OF MODEL M_{th}. THE PROBABILITY THAT THE ith ANIMAL IS CAUGHT ON THE jth TRAPPING OCCASION IS EQUAL TO $p_i p_j$, $j = 1, \ldots, t$ AND $i = 1, \ldots, N$

N	p_i, $i = 1,2,\ldots,N$	p_j, $j = 1,2,\ldots,t$	Trial
400	$p_i = 0.15$, $i = 1,100$; $p_i = 0.33$, $i = 101,200$; $p_i = 0.67$, $i = 201,300$; $p_i = 1.00$, $i = 301,400$.	$p_1 = 0.30$, $p_2 = 0.60$, $p_3 = 0.10$, $p_4 = 0.30$, $p_5 = 0.60$.	1
400	$p_i = 0.40$, $i = 1,100$; $p_i = 0.60$, $i = 101,200$; $p_i = 0.80$, $i = 201,300$; $p_i = 1.00$, $i = 301,400$.	$p_1 = p_2 = 0.25$, $p_3 = 0.50$, $p_4 = 0.15$, $p_5 = 0.25$.	2
400	$p_i = 0.30$, $i = 1,150$; $p_i = 0.40$, $i = 151,250$; $p_i = 1.00$, $p_i = 251,400$.	$p_1 = 0.15$, $p_2 = 0.25$, $p_3 = 0.05$, $p_4 = 0.10$, $p_5 = 0.30$.	3
400	$p_i = 0.03$, $i = 1,100$; $p_i = 0.35$, $i = 101,200$; $p_i = 0.55$, $i = 201,300$; $p_i = 1.00$, $i = 301,400$.	$p_1 = p_2 = p_3 = 0.35$, $p_4 = 0.40$, $p_5 = 0.60$.	4
200	$p_i = 0.40$, $i = 1,50$; $p_i = 0.60$, $i = 51,100$; $p_i = 0.80$, $i = 101,150$; $p_i = 1.00$, $i = 151,200$.	$p_1 = 0.50$, $p_2 = p_3 = 0.30$, $p_4 = 0.60$, $p_5 = 0.20$, $p_6 = 0.40$, $p_7 = 0.30$.	5
200	$p_i = 0.25$, $i = 1,50$; $p_i = 0.50$, $i = 51,150$; $p_i = 1.00$, $i = 151,200$.	$p_1 = 0.15$, $p_2 = 0.25$, $p_3 = 0.05$, $p_4 = 0.10$, $p_5 = 0.30$.	6
400	$p_i = 0.45$, $i = 1,100$; $p_i = 0.55$, $i = 101,200$; $p_i = 0.65$, $i = 201,300$; $p_i = 0.75$, $i = 301,400$.	$p_1 = 0.65$, $p_2 = 0.75$, $p_3 = 0.55$, $p_4 = 0.35$, $p_5 = 0.60$, $p_6 = 0.50$, $p_7 = 0.80$.	7
100	$p_i = 0.35$, $i = 1,25$; $p_i = 0.45$, $i = 26,50$; $p_i = 0.55$; $i = 51,75$; $p_i = 0.65$, $i = 76,100$.	$p_1 = 0.65$, $p_2 = 0.75$, $p_3 = 0.55$, $p_4 = 0.35$, $p_5 = 0.60$.	8

TABLE N.5.b.—SIMULATION RESULTS OF ESTIMATION PROCEDURES APPLIED TO DATA GENERATED ACCORDING TO MODEL M_{th}

Estimator	N	Ave($\hat{N}$)	$\sigma(\hat{N})$	Ave $\sqrt{\hat{V}ar(\hat{N})}$	Coverage	R	t	Trial
$\hat{N}_t$	400	303.60	14.13	11.74	0.000	100	5	1
$\hat{N}_t$	400	369.90	21.69	19.69	0.600	100	5	2
$\hat{N}_t$	400	309.04	37.46	35.66	0.340	100	5	3
$\hat{N}_t$	400	272.77	12.53	10.32	0.000	100	5	4
$\hat{N}_t$	200	187.93	5.68	5.29	0.360	100	7	5
$\hat{N}_t$	200	164.17	34.05	29.11	0.590	100	5	6
$\hat{N}_h$	400	402.88	31.42	23.46	0.850	100	5	1
$\hat{N}_h$	400	478.73	33.55	29.20	0.220	100	5	2
$\hat{N}_h$	400	353.90	25.02	26.62	0.590	100	5	3
$\hat{N}_h$	400	352.78	29.38	20.68	0.430	100	5	4
$\hat{N}_h$	200	217.72	8.42	9.46	0.560	100	7	5
$\hat{N}_h$	200	178.21	21.29	18.67	0.760	100	5	6
$\hat{N}_{bh}$	400	Failed in all 100 replications				100	5	1
$\hat{N}_{bh}$	400	272.56	11.65	6.27	0.000	100	5	2
$\hat{N}_{bh}$	400	Failed in all 100 replications				100	5	3
$\hat{N}_{bh}$	400	489.06	275.72	319.02	0.840	100	5	4
$\hat{N}_{bh}$	200	179.41	9.27	9.16	0.340	100	7	5
$\hat{N}_{bh}$	200	235.91	156.90	337.63	0.793	92	5	6

TABLE N.6.a.—DESCRIPTION OF THE TRIALS OF MODEL M_{bh}

N	p_i, c_i; i = 1, 2, . . ., N	Trial
400	$p_i = 0.05$, i = 1,200; $p_i = 0.15$, i = 201,300; $p_i = 0.50$, i = 301,400.	1[1]
400	$p_i = 0.01$, i = 1,50; $p_i = 0.15$, i = 51,200; $p_i = 0.25$, i = 201,300; $p_i = 0.30$, i = 301,400.	2[1]
400	$p_i = 0.10$, i = 1,100; $p_i = 0.20$, i = 101,200; $p_i = 0.25$, i = 201,300; $p_i = 0.30$, i = 301,400.	3[1]
400	$p_i = 0.20$, i = 1,100; $p_i = 0.30$, i = 101,200; $p_i = 0.40$, i = 201,300; $p_i = 0.50$, i = 301,400.	4[1]
200	$p_i = 0.05$, i = 1,50; $p_i = 0.15$, i = 51,150; $p_i = 0.25$, i = 151,200.	5[1]
200	$p_i = 0.15$, i = 1,50; $p_i = 0.20$, i = 51,100; $p_i = 0.25$, i = 101,150; $p_i = 0.30$, i = 151,200.	6[1]
100	$p_i = 0.10$, i = 1,40; $p_i = 0.20$, i = 41,80; $p_i = 0.30$, i = 81,100.	7[1]
400	$p_i = 0.10$, i = 1,50; $p_i = 0.25$, i = 51,200; $p_i = 0.35$, i = 201,300; $p_i = 0.45$, i = 301,400; $c_i = \max[(p_i - \text{ran}(i)/4),0]$.	8[2]
400	$p_i = 0.05$, i = 1,50; $p_i = 0.10$, i = 51,200; $p_i = 0.15$, i = 201,300; $p_i = 0.25$, i = 301,400; $c_i = \min[(p_i + \text{ran}(i)/4),1]$.	9[2]
400	$p_i = 0.05$, i = 1,50; $p_i = 0.10$, i = 51,200; $p_i = 0.15$, i = 201,300; $p_i = 0.25$, i = 301,400; $c_i = \max[(p_i - \text{ran}(i)/4),0]$.	10[2]
100	$p_i \sim \beta(1,22/3)$, i = 1,100[3]; $c_i = \min[(p_i + \text{ran}(i)/2,1)]$	11[2]
100	$p_i \sim \beta(1,22/3)$, i = 1,100[3]; $c_i = \max[(p_i - \text{ran}(i)/2),0]$	12[2]
100	$p_i \sim \beta(1,22/3)$, i = 1,100[3]; $c_i = \min\{\max[p_i + (\text{ran}(i) - 0.5)/2,0],1\}$	13[2]
400	$p_i = 0.15$, i = 1,100, $p_i = 0.25$, i = 101,300; $p_i = 0.35$, i = 301,400. $c_i = 0.5\, p_i$	14
100	$p_i = 0.15$, i = 1,33; $p_i = 0.25$, i = 34,66; $p_i = 0.35$, i = 67,100. $c_i = 0.5\, p_i$	15

[1] Indicates the trial was used for estimation purposes only. Because the performance of the estimator associated with Model M_{bh} depends only upon N and the probabilities of first capture p_i, recapture probabilities; c_i need not be specified.
[2] The function ran(i) produces a random value of a variable distributed uniformly over the interval [0,1].
[3] $p_i \sim \beta(1,22/3)$ indicates that the random variable p_i has a beta distribution with parameters 1 and 22/3.

TABLE N.6.b.—SIMULATION RESULTS OF ESTIMATION PROCEDURES USED ON DATA GENERATED ACCORDING TO MODEL M_{bh}

Estimator	N	Ave($\hat{N}$)	$\sigma(\hat{N})$	Ave$\sqrt{\hat{V}ar(\hat{N})}$	Coverage	R	t	Trial
$\hat{N}_{bh}$[1]	400	246.59	38.04	32.26	0.120	100	5	1
$\hat{N}_{bh}$[1]	400	340.83	57.57	47.90	0.360	100	5	2
$\hat{N}_{bh}$[1]	400	366.43	41.87	35.63	0.600	100	5	3
$\hat{N}_{bh}$[1]	400	383.00	21.30	15.89	0.560	100	5	4
$\hat{N}_{bh}$[1]	200	175.51	16.26	13.52	0.380	100	10	5
$\hat{N}_{bh}$[1]	200	193.72	13.90	13.72	0.780	100	7	6
$\hat{N}_{bh}$[1]	100	94.14	43.26	41.48	0.620	100	5	7

[1] Note, we use $\hat{N}_{bh}$ and $\hat{N}_R$ as equivalent notations.

TABLE N.7.—SIMULATION RESULTS CONCERNING THE SIZE AND POWER OF THE TEST T_1 OF MODEL M_o VS. MODEL M_h

α					
0.01	0.05	0.10	Data model	R	t
0.0100	0.0300	0.0550	M_o, Trial 2	200	5
0.0100	0.0150	0.0450	M_o, Trial 6	200	5
0.9900	0.9900	0.9900	M_h, Trial 1	200	5
0.0900	0.1150	0.1400	M_h, Trial 2	200	5
0.0850	0.1350	0.1900	M_h, Trial 3	200	5
0.1350	0.1650	0.1950	M_h, Trial 8	200	5
0.3950	0.5650	0.6250	M_h, Trial 8	200	10
0.1700	0.1750	0.2500	M_h, Trial 10	200	5
0.7450	0.7750	0.8400	M_h, Trial 11	200	5

TABLE N.8.—SIMULATED AND THEORETICAL RESULTS CONCERNING THE SIZE AND POWER OF THE TEST T_2 OF MODEL M_o VS. MODEL M_b

	α					
Method	0.01	0.05	0.10	Data model	R	t
Simulation	0.0100	0.0550	0.0860	M_o, Trial 2	198	5
Theoretical	0.0100	0.0500	0.1000	M_o, Trial 2	198	5
Simulation	0.0280	0.0830	0.1240	M_o, Trial 3	372	5
Theoretical	0.0100	0.0500	0.1000	M_o, Trial 3	372	5
Simulation	0.0600	0.1300	0.1400	M_o, Trial 4	100	7
Theoretical	0.0100	0.0500	0.1000	M_o, Trial 4	100	7
Simulation	1.0000	1.0000	1.0000	M_b, Trial 1	200	5
Theoretical	0.9990	0.9990	0.9940	M_b, Trial 1	200	5
Simulation	0.9150	0.9700	0.9800	M_b, Trial 2	200	5
Theoretical	0.9480	0.9880	0.9947	M_b, Trial 2	200	5
Simulation	0.8800	0.9640	0.9790	M_b, Trial 3	193	5
Theoretical	0.8480	0.9500	0.9750	M_b, Trial 3	193	5
Simulation	0.0050	0.0660	0.1890	M_b, Trial 4	200	5
Theoretical	0.0510	0.1550	0.2450	M_b, Trial 4	200	5
Simulation	0.9700	0.9900	1.0000	M_b, Trial 5	200	5
Theoretical	0.9460	0.9870	0.9980	M_b, Trial 5	200	5
Simulation	0.2800	0.5100	0.6350	M_b, Trial 7	200	5
Theoretical	0.2610	0.4900	0.6140	M_b, Trial 7	200	5
Simulation	0.4220	0.6280	0.7140	M_b, Trial 8	199	5
Theoretical	0.4570	0.6940	0.7950	M_b, Trial 8	199	5
Simulation	0.0190	0.1990	0.3230	M_b, Trial 10	161	5
Theoretical	0.2200	0.4370	0.5620	M_b, Trial 10	161	5

TABLE N.9.—SIMULATED AND THEORETICAL RESULTS CONCERNING THE SIZE AND POWER OF THE TEST T_3 OF MODEL M_o VS. MODEL M_t

Method	α 0.01	α 0.05	α 0.10	Data model	R	t
Simulation	0.0150	0.0525	0.1100	M_o, Trial 2	400	5
Theoretical	0.0100	0.0500	0.1000	M_o, Trial 2	400	5
Simulation	0.0120	0.0500	0.1060	M_o, Trial 3	500	5
Theoretical	0.0100	0.0500	0.1000	M_o, Trial 3	500	5
Simulation	1.0000	1.0000	1.0000	M_t, Trial 2	200	5
Theoretical	0.9998	0.9999	0.9999	M_t, Trial 2	200	5
Simulation	0.3800	0.6550	0.7350	M_t, Trial 5	200	5
Theoretical	0.4269	0.6635	0.7703	M_t, Trial 5	200	5
Simulation	0.6138	0.8028	0.7984	M_t, Trial 12	989	5
Theoretical	0.5925	0.7984	0.8751	M_t, Trial 12	989	5

TABLE N.10.—SIMULATION RESULTS CONCERNING THE SIZE AND POWER OF THE GOODNESS OF FIT TEST T_4 OF MODEL M_h

α 0.01	α 0.05	α 0.10	Data model	R	t
0.0200	0.0500	0.0900	M_o, Trial 2	100	5
0.0000	0.0400	0.0900	M_h, Trial 3	100	5
0.0000	0.0500	0.1000	M_h, Trial 8	100	10
0.0000	0.0800	0.1300	M_h, Trial 10	100	5
1.0000	1.0000	1.0000	M_{bh}, Trial 14	100	7
0.0900	0.2300	0.3700	M_{bh}, Trial 15	100	5
1.0000	1.0000	1.0000	M_{th}, Trial 7	100	7
0.5400	0.7900	0.8800	M_{th}, Trial 8	100	5
1.0000	1.0000	1.0000	M_{tbh}, Trial 1	100	7
0.1600	0.3300	0.4200	M_{tbh}, Trial 2	100	5

TABLE N.11.—SIMULATION RESULTS CONCERNING THE SIZE AND POWER OF THE GOODNESS OF FIT TEST T_5 OF MODEL M_b

α 0.01	α 0.05	α 0.10	Data model	R	t
0.0000	0.0200	0.0200	M_b, Trial 11	100	7
0.0000	0.0500	0.1200	M_b, Trial 12	100	5
0.0000	0.0600	0.1000	M_{bh}, Trial 14	100	7
0.0000	0.1000	0.1600	M_{bh}, Trial 15	100	5
1.0000	1.0000	1.0000	M_{tb}, Trial 1	100	7
0.3300	0.5700	0.7100	M_{tb}, Trial 2	100	5
1.0000	1.0000	1.0000	M_{tbh}, Trial 1	100	7
0.0600	0.1800	0.3000	M_{tbh}, Trial 2	100	5

TABLE N.12.—SIMULATION RESULTS CONCERNING THE SIZE AND POWER OF THE GOODNESS OF FIT TEST T_6 OF MODEL M_t

α 0.01	α 0.05	α 0.10	Data model	R	t
0.0000	0.0200	0.1100	M_t, Trial 13	100	7
0.0000	0.0300	0.0600	M_t, Trial 14	100	5
0.0000	0.0100	0.0900	M_{tb}, Trial 1	100	7
0.0100	0.0700	0.1200	M_{tb}, Trial 2	100	5
0.0600	0.2800	0.3600	M_{th}, Trial 7	100	7
0.0300	0.0700	0.2000	M_{th}, Trial 8	100	5
0.9300	0.9800	0.9900	M_{tbh}, Trial 1	100	7
0.0200	0.0600	0.1000	M_{tbh}, Trial 2	100	5

TABLE N.13.—SIMULATION RESULTS CONCERNING THE SIZE AND POWER OF THE TEST T_7 OF MODEL M_h VS. MODEL M_{bh}

α 0.01	α 0.05	α 0.10	Data model	R	t
0.0100	0.0800	0.1500	M_h, Trial 10	100	5
0.0100	0.0300	0.0800	M_h, Trial 13	100	5
0.0600	0.0700	0.0800	M_h, Trial 14	100	10
0.4000	0.6700	0.7600	M_{bh}, Trial 8	100	5
0.1500	0.4000	0.5100	M_{bh}, Trial 9	100	5
0.0900	0.2300	0.3300	M_{bh}, Trial 10	100	5
0.2800	0.5000	0.6100	M_{bh}, Trial 11	100	5
0.0200	0.1400	0.1800	M_{bh}, Trial 12	100	5
0.0100	0.0500	0.1100	M_{bh}, Trial 13	100	5

TABLE N.14.—DESCRIPTION OF THE TRIALS OF MODELS M_{tb} AND M_{tbh}

M_{tb}[1]			
N	P_j, j = 1,2,...,t	c	Trial
400	$p_1 = 0.05$, $p_2 = 0.35$, $p_3 = 0.25$, $p_4 = 0.15$, $p_5 = 0.20$, $p_6 = 0.10$, $p_7 = 0.30$.	2.5	1
100	$p_1 = 0.35$, $p_2 = 0.25$, $p_3 = 0.15$, $p_4 = 0.20$, $p_5 = 0.30$.	0.5	2

M_{tbh}[2]				
N	P_i, i = 1,2,...,N	P_j, j = 1,2,...,t	c	Trial
400	$p_i = 0.15$, i = 1,100; $p_i = 0.25$, i = 101,200; $p_i = 0.35$, i = 201,300; $p_i = 0.45$, i = 301,400.	$p_1 = 0.65$, $p_2 = 0.75$, $p_3 = 0.55$, $p_4 = 0.45$, $p_5 = 0.60$, $p_6 = 0.50$, $p_7 = 0.70$.	2.5	1
100	$p_i = 0.15$, i = 1,25; $p_i = 0.25$, i = 26,50; $p_i = 0.35$, i = 51,75; $p_i = 0.45$, i = 76,100.	$p_1 = 0.65$, $p_2 = 0.75$, $p_3 = 0.55$, $p_4 = 0.45$, $p_5 = 0.60$.	0.75	2

[1] The probability of any animal being captured on the jth occasion, j = 1,2,...,t, is p_j if the animal has not previously been caught and cp_j if the animal is being recaptured.

[2] The probability of the ith animal being captured on the jth occasion is p_ip_j if the animal has not previously been caught and p_ip_jc if the animal is being recaptured.

APPENDIX O

Interval Estimation

Use of the Central Limit Theorem (cf. Mood et al. 1974:195) in setting so-called "normal theory" confidence intervals for parameters is widespread. The theorem usually is cited as justification for asserting that, for "large samples," a 95 percent confidence interval for the parameter of interest θ is given by $P\{\hat{\theta} - 1.96\hat{\sigma}(\hat{\theta}) \leq \theta \leq \hat{\theta} + 1.96\hat{\sigma}(\hat{\theta})\} = 0.95$, where $\hat{\sigma}(\hat{\theta})$ represents an estimate of the standard deviation of $\hat{\theta}$ (also see Seber 1973:134 for an example of an indirectly constructed confidence interval based on asymptotic normality). Further impetus is given to this argument when ML estimators are involved by the knowledge that, under certain regularity conditions, those estimators are BAN (Best Asymptotically Normal). Furthermore, large sample (normal theory) confidence intervals based on ML estimators are known to have smaller expected width than intervals constructed using any other estimator (cf. Mood et al. 1974:393).

Unfortunately, it is generally true that assertions concerning the operating characteristics of large sample intervals cannot be made if sample sizes are small or regularity conditions are not met or both. Even more unfortunate is the fact that small or moderate sample sizes are more often the rule than the exception in capture–recapture experiments. Therefore, it was not unexpected that initial simulation of capture–recapture experiments revealed that "normal theory" confidence intervals often exhibit undesirable properties. For example, the lower limit of a given interval for population size N may be less than the number of different animals captured in the experiment. Also, coverage of the confidence interval, i.e., the percentage of simulated intervals that contain the true value N, is often significantly less than the nominal level of 0.95. Because of such problems, 2 alternative interval estimation procedures were investigated in the hope that a more satisfactory procedure could be suggested for practical use.

The first procedure is based upon the supposition that the distribution of the estimator $\hat{N}^{-1}$ is more symmetric (hence closer to normality) than the distribution of $\hat{N}$ (cf. Cormack 1968). Thus, the following procedure was proposed. First, calculate the ML estimates $\hat{N}$ and $\hat{V}ar(\hat{N})$. Construct a 95 percent confidence interval for $\hat{N}^{-1}$ of the form $P\{\hat{N}^{-1} - 1.96\hat{N}^{-2}\sqrt{\hat{V}ar(\hat{N})} \leq N^{-1} \leq \hat{N}^{-1} + 1.96\hat{N}^{-2}\sqrt{\hat{V}ar(\hat{N})}\}$. Finally, invert the interval in the obvious manner to arrive at a confidence interval for N. Simulation results involving Model M_t showed that, although the distribution of $\hat{N}^{-1}$ was nearly normal, the above procedure did not represent a significant improvement over the usual large sample interval. Coverage of the 2 procedures was roughly the same, but the "reciprocal" procedure had, on the average, greater width than the usual large sample procedure. Moreover, lower confidence limits for $\hat{N}^{-1}$ sometimes were less than zero and thus, upon inversion, the upper limit for $\hat{N}$ was negative.

The second alternative method for interval estimation involved the use of only the likelihood function of the sample, and is based largely upon the likelihood principle (cf. Kendall and Stuart 1973:226). That principle asserts that the likelihood function provides all the information necessary for making statistical inferences concerning the data and has been subjected to serious theoretical questioning (Stein 1962, Birnbaum 1968, Kendall and Stuart 1973). Nevertheless, we felt that it would be beneficial to simulate the operating characteristics of these "likelihood intervals" in capture–recapture experiments to evaluate their practical utility.

Briefly, the mechanics of constructing a likelihood interval are as follows. (The reader is referred to Hudson [1971] for a thorough explanation.) For a given data set **X**, form the log-likelihood function $\ln L(\theta|\mathbf{X})$. (Assume θ is a scalar for simplicity of presentation.) Under the assumption that $\hat{\theta}$, the ML estimator of θ, is unique, and that the likelihood function is unimodal, the likelihood interval $I(\theta)$ is defined as $I(\theta) = \{\theta: \ln L(\theta|\mathbf{X}) \geq \ln L(\hat{\theta}|\mathbf{X}) - 2\}$. $I(\theta)$ consists of all those θ for which $\ln L(\theta|\mathbf{X})$ (the log-likelihood function evaluated at θ) is no more than 2 units away from the maximum value of the likelihood function $\ln L(\hat{\theta}|\mathbf{X})$. The assumption is that these values of θ are "plausible" values of θ, in view of the data observed. That is, these values of θ produce values of the likelihood function that are not "far" from its maximum, and thus they cannot be discounted. The use of the value 2 may seem arbitrary, but Hudson (1971) argued that this value leads to asymptotic 95 percent confidence intervals.

Likelihood intervals for population size N were constructed from data simulated according to the 2-sample removal experiment treated by Seber and Whale (1970). Varying the values of N and p (probability of removal) did not appear to significantly affect coverage probability of the intervals, although it is theoretically true that the coverage probability varies at least slightly with the true values of the parameters. Moreover, lower limits of the intervals did not extend below the number of animals seen. Average coverage of the likelihood intervals (95.5%) was approximately the same as that of the "normal theory" confidence intervals (92.2%) constructed from the same data. However, average interval width for the likelihood intervals was consistently greater than that of the normal confidence intervals; on the average they were approximately 10 percent greater. In view of these somewhat mixed results, a second simulation study was performed on data from Model M_t. That study revealed similar results, in that coverage of the 2

procedures was on the average approximately the same (96.0% for likelihood vs. 93.2% for "normal" intervals), and the average width for the likelihood intervals was always greater than that of the normal confidence intervals. In one case, the average width of the likelihood interval was more than 3 times the average width of the normal confidence interval. Finally, in virtually every simulation involving likelihood intervals, we noted that approximately half of those intervals that did not contain N were too low (i.e., the upper limit of the interval was less than N) and half were too high (i.e., the lower limit of the interval was greater than N). This is in sharp contrast to the results of simulating normal confidence intervals, where we have found that the large majority of intervals that do not contain N are too low.

On the basis of the results described above, we recommend continued use of the usual large sample confidence interval procedure rather than either of the 2 alternative procedures discussed. All 3 procedures possess both attractive and unattractive operating characteristics, and the choice therefore cannot be clear-cut. The decision to continue the use of large sample confidence intervals was made for 2 basic reasons. First, biologists in general tend to be more familiar with the computation and use of that procedure. Second, and more important, is the fact that more is known about the theoretical properties of the large sample procedure than is known about the 2 alternatives, and, therefore, it may be possible to assess theoretically the small sample behavior of the procedure in certain capture–recapture situations. Finally, we wish to reemphasize that large sample "normal" confidence intervals should be used with great caution in many capture–recapture experiments, and that much more theoretical work appears necessary before more adequate procedures are available.

Copies of the following WILDLIFE MONOGRAPHS are available from the Executive Director of The Wildlife Society, Suite 611, 7101 Wisconsin Avenue, N. W., Washington, D.C. 20014. All others are out of print. Maryland residents add 5% sales tax. For a complete list of all WILDLIFE MONOGRAPHS published through 1975, see MONOGRAPH No. 48.

No. 8. Ecology of the scaled quail in the Oklahoma panhandle. Sanford D. Schemnitz. December 1961. 47 pages. Price $2.70.

No. 11. The ecology of a spring stream, Doe Run, Meade County, Kentucky. W. L. Minckley. September 1963. 124 pages. Price $3.40. Reprint.

No. 13. Social behavior in confined populations of the cottontail and the swamp rabbit. Halsey M. Marsden and Nicholas R. Holler. April 1964. 39 pages. Price $2.60. Reprint.

No. 16. The status and ecology of the Roosevelt elk in California. James A. Harper, Joseph H. Harn, Wallace W. Bentley, and Charles F. Yocom. August 1967. 49 pages. Price $2.70.

No. 18. The status, population dynamics, and harvest of the dusky Canada goose. Joseph A. Chapman, Charles J. Henny, and Howard M. Wight. November 1969. 48 pages. Price $2.70.

No. 22. The shortgrass prairie Canada goose population. Jack R. Grieb. May 1970. 49 pages. Price $2.70.

No. 23. The Sun River Elk Herd. Richard R. Knight. October 1970. 66 pages. Price $2.85.

No. 28. A Columbia River Canada goose population. W. C. Hanson and L. L. Eberhardt. December 1971. 61 pages. Price $2.85.

No. 30. Ecology of blesbok with special reference to productivity. S. S. du Plessis. August 1972. 70 pages. Price $3.00.

No. 31. Adaptation of a free-ranging rhesus monkey group to division and transplantation. John A. Morrison and Emil W. Menzel, Jr. November 1972. 78 pages. Price $3.00.

No. 33. Home ranges and activity patterns of nonmigratory elk of the Madison Drainage Herd as determined by biotelemetry. John J. Craighead, Frank C. Craighead, Jr., Robert L. Ruff, and Bart W. O'Gara. August 1973. 50 pages. Price $2.70.

No. 34. Effects of hunting and some other environmental factors on scaled quail in New Mexico. Howard Campbell, Donald K. Martin, Paul E. Ferkovich, and Bruce K. Harris. August 1973. 49 pages. Price $2.70.

No. 35. Mountain lion social organization in the Idaho Primitive Area. John C. Seidensticker IV, Maurice G. Hornocker, Wilbur V. Wiles, and John P. Messick. December 1973. 60 pages. Price $2.85.

No. 36. Systematic status of the cottontail complex in western Maryland and nearby West Virginia. Joseph A. Chapman and Raymond P. Morgan II. December 1973. 54 pages. Price $2.70.

No. 37. Ecology and management of the Atitlán grebe, Lake Atitlán, Guatemala. Anne LaBastille. August 1974. 66 pages. Price $2.85.

No. 38. Social behavior and ecology of the pronghorn. David W. Kitchen. August 1974. 96 pages. Price $3.15.

No. 39. Growth and morphometry of the carcass, selected bones, organs, and glands of mule deer. Allen E. Anderson, Dean E. Medin and David C. Bowden. October 1974. 122 pages. Price $3.40.

No. 40. Demographic analysis of a northern Utah jackrabbit population. Jack E. Gross, L. Charles Stoddart, and Frederic H. Wagner. October 1974. 68 pages. Price $2.85.

No. 41. Nesting ecology of the bobwhite in southern Illinois. W. D. Klimstra and John L. Roseberry. February 1975. 37 pages. Price $2.60.

Additional monographs listed on the inside back cover.

No. 42. On the behavior and socialization of pronghorn fawns. Robert E. Autenrieth and Edson Fichter. March 1975. 111 pages. Price $3.30.

No. 43. Ecology of the timber wolf in northeastern Minnesota. Victor Van Ballenberghe, Albert W. Erickson, and David Byman. April 1975. 44 pages. Price $2.60.

No. 44. Ungulate–habitat relationships in a South African woodland/savanna ecosystem. Stanley M. Hirst. November 1975. 60 pages. Price $2.85.

No. 45. Breeding biology of cackling geese and associated species on the Yukon–Kuskokwim Delta, Alaska. P. G. Mickelson. November 1975. 35 pages. Price $2.60.

No. 46. Experimental duck hunting seasons, San Luis Valley, Colorado, 1963–1970. Richard M. Hopper, Aelred D. Geis, Jack R. Grieb, and Lewis Nelson, Jr. November 1975. 68 pages. Price $2.85.

No. 47. A field study of the short-eared owl *Asio flammeus* (Pontoppidan) in North America. Richard J. Clark. November 1975. 67 pages. Price $2.85.

No. 48. Moose habitat selection and relationships to forest management in northeastern Minnesota. James M. Peek, David L. Urich, and Richard J. Mackie. April 1976. 65 pages. Price $2.85.

No. 49. Morphology, reproduction, dispersal, and mortality of midwestern red fox populations. Gerald L. Storm, Ronald D. Andrews, Robert L. Phillips, Richard A. Bishop, Donald B. Siniff, and John R. Tester. April 1976. 82 pages. Price $3.00.

No. 50. The potential and realized influences of temperature on the distribution of fishes in the New River, Glen Lyn, Virginia. Jay R. Stauffer, Jr., Kenneth L. Dickson, John Cairns, Jr., and Donald S. Cherry. October 1976. 40 pages. Price $2.60.

No. 51. Crippling effects of lead, steel, and copper shot on experimental mallards. Robert L. Cochrane. Effects of lead and steel shot on shooting of flighted mallards. Ronald H. Nicklaus. November 1976. 29 pages. Price $2.60.

No. 52. The biology of the mountain duck on Rottnest Island, Western Australia. Thomas L. Riggert. January 1977. 67 pages. Price $2.85.

No. 53. Social behavior of white-tailed deer in relation to habitat. David H. Hirth. April 1977. 55 pages. Price $2.70.

No. 54. Ecology and factors limiting roan and sable antelope populations in South Africa. David E. Wilson and Stanley M. Hirst. April 1977. 111 pages. Price $3.30.

No. 55. Effect of logging on songbird populations in a northern hardwood forest. William L. Webb, Donald F. Behrend, and Boonruang Saisorn. July 1977. 35 pages. Price $2.60.

No. 56. Reproductive and physiological cycles in the cottontail complex in western Maryland and nearby West Virginia. Joseph A. Chapman, Amy L. Hartman, and David E. Samuel. July 1977. 73 pages. $2.85.

No. 57. Ecology of the emperor goose in Alaska. David I. Eisenhauer and Charles M. Kirkpatrick. October 1977. 62 pages. Price $2.85.

No. 58. A demographic analysis of the snowshoe hare cycle. Lloyd B. Keith and Lamar A. Windberg. April 1978. 70 pages. Price $2.85.

No. 59. New principles for the conservation of wild living resources. Sidney J. Holt and Lee M. Talbot. April 1978. 33 pages. $2.60.

No. 60. Disease and physiologic characteristics of two cottontail populations in Virginia. Harry A. Jacobson, Roy L. Kirkpatrick, and Burd S. McGinnes. July 1978. 53 pages. $2.70.

No. 61. The history and breeding biology of the Canada geese of Marshy Point, Manitoba. James A. Cooper. July 1978. 87 pages. Price $3.15.

No. 62. Statistical inference from capture data on closed animal populations. David L. Otis, Kenneth P. Burnham, Gary C. White, and David R. Anderson. October 1978. 135 pages. Price $3.50.