On the Statistical Analysis of Capture Experiments

R. M. Huggins

Biometrika, Vol. 76, No. 1. (Mar., 1989), pp. 133-140.

Stable URL:
http://links.jstor.org/sici?sici=0006-3444%28198903%2976%3A1%3C133%3AOTSAOC%3E2.0.CO%3B2-A

Biometrika is currently published by Biometrika Trust.

Biometrika (1989), **76**, 1, *pp.* 133–40
Printed in Great Britain

On the statistical analysis of capture experiments

By R. M. HUGGINS

Department of Statistics, La Trobe University, Bundoora, Victoria 3083, Australia

SUMMARY

A procedure is given for estimating the size of a closed population in the presence of heterogeneous capture probabilities using capture-recapture data when it is possible to model the capture probabilities of individuals in the population using covariates. The results include the estimation of the parameters associated with the model of the capture probabilities and the use of these estimated capture probabilities to estimate the population size. Confidence intervals for the population size using both the asymptotic normality of the estimator and a bootstrap procedure for small samples are given.

Some key words: Bootstrap; Capture experiment; Population size estimation; Variable capture probability.

1. INTRODUCTION

This work is concerned with the use of data from capture-recapture experiments to estimate the size of a closed population with heterogeneous capture probabilities modelled using covariates. Such an approach differs from those in the literature. Whilst Seber (1982) and Otis et al. (1978) give many techniques, all except the jackknife procedures of Burnham & Overton (1978), detailed by Otis et al. (1978), depend on homogeneous capture probabilities. The necessity of this homogeneity assumption when using methods based on the full likelihood is clear from its form.

Let a population consist of $i = 1, \ldots, \nu$ individuals which may be captured on occasions $j = 1, \ldots, t$ with probabilities Λ_{ij} that individual i is captured on trapping occasion j. It is crucial that the individuals behave independently. Then the full likelihood is

$$L^* = K \prod_{i=1}^{\nu} \prod_{j=1}^{t} \Lambda_{ij}^{\Delta_{ij}}(1-\Lambda_{ij})^{(1-\Delta_{ij})},$$

where $\Delta_{ij} = 1$ if individual i is captured on occasion j and 0 otherwise, and K may depend on ν but does not depend on any of the parameters which may be involved in the Λ_{ij}. Letting $i = 1, \ldots, n$ denote the captured individuals and $i = n+1, \ldots, \nu$ the individuals not captured during the course of the trapping experiment, we may write

$$L^* = K \prod_{i=1}^{n} \prod_{j=1}^{t} \Lambda_{ij}^{\Delta_{ij}}(1-\Lambda_{ij})^{(1-\Delta_{ij})} \prod_{i=n+1}^{\nu} \prod_{j=1}^{t} (1-\Lambda_{ij}).$$

It is clear from this partitioning of the likelihood that without some homogeneity assumption no expression for the contribution to the likelihood from the uncaptured individuals may be found and in general no maximum likelihood estimate of ν will be available.

To illustrate our methods we consider a linear logistic model for the capture probabilities,

$$\Lambda_{ij} = \frac{\exp(\beta_0 + \beta_1 z_i + \beta_2 x_j + \alpha z_{ij})}{1+\exp(\beta_0 + \beta_1 z_i + \beta_2 x_j + \alpha z_{ij})},$$

where Λ_{ij} is the probability that individual i is captured on occasion j given his past capture behaviour before j, z_i is a vector of individual covariates such as weight or age, x_j is a vector of environmental covariates such as minimum daily temperature or daily rainfall corresponding to trapping occasion j and z_{ij} is a vector of covariates such as the individual's past capture history which may change with time. For such a model one would typically collect data corresponding to the captured individuals of Table 1. We shall use the data on the 18 captured individuals to illustrate our technique and note that here as in the example below z_{ij} is taken to be an indicator of the capture of individual i prior to occasion j.

Table 1. *Simulated capture data for the linear logistic model depending on an occasion covariate x_j and an individual covariate z_i for* 10 *trapping occasions and* 20 *individuals*

(a) *Individual covariates*

6·3	4·0	4·7	4·8	6·6	3·9	5·9	5·0	5·6	4·6
5·9	4·4	4·6	4·4	5·6	3·6	3·7	4·4	3·6	4·4

(b) *Occasion covariates*

5·8	2·9	3·7	2·2	3·5	3·6	1·9	3·0	4·8	4·7

(c) *Trapping histories*

1:	0 1 0 1 1 0 0 1 0 0	11:	0 0 0 1 0 1 0 0 0 0
2:	0 1 1 1 1 0 0 1 0 0	12:	0 0 0 0 0 0 0 0 0 0
3:	0 1 0 0 1 1 1 1 0 0	13:	0 0 0 0 0 0 0 0 0 1
4:	0 0 1 1 1 0 1 1 0 0	14:	0 0 0 1 0 0 1 0 0 0
5:	0 0 0 0 1 0 1 0 1 1	15:	0 0 1 1 0 1 1 1 0 0
6:	0 0 0 0 0 0 0 0 0 0	16:	0 0 0 0 0 0 1 0 0 0
7:	0 1 1 1 0 1 1 0 0 0	17:	0 0 0 0 1 1 0 0 0 0
8:	0 0 0 1 0 1 0 0 1 0	18:	0 1 1 1 1 0 1 0 0 0
9:	0 0 0 0 0 1 1 1 0 0	19:	0 0 0 0 1 0 0 1 0 0
10:	0 0 0 1 0 1 0 0 0 0	20:	0 0 0 0 0 0 1 0 0 0

The second data used for illustration are those discussed by Otis et al. (1978, p. 62), available as one of the sets of data accompanying their program CAPTURE. The data were the result of a population ecology study on salt marsh rodents and concerns 10 capture occasions in both the mornings and evenings. The data recorded included the sex, the age and capture history of each captured individual. A total of 173 individuals was captured. The methods of Otis et al. (1978) lead to the selection of a model where the capture probabilities are heterogeneous and depend on time. For such a model there is no estimator available apart from that given below.

The model we consider here, partly for mathematical convenience, is again linear logistic with the covariates sex, age group, i.e. juvenile, semi-adult or adult, time of the trapping occasion, i.e. morning or evening, and prior capture history, yes or no. The particular form of the model has no essential role in the development of the theory below.

2. Main results

We now construct a likelihood conditionally on the captured individuals with the full parameter vector denoted by θ. We observe that information is obtained only on individuals captured at least once and we assume that the covariates z_i for such individuals are known for the entire experiment. Suppose a total of n individuals which we relabel $i=1, \ldots, n$ are captured. Let C_{ij} be the event individual i is captured in trapping period j, C_i be the event individual i is captured at least once, and let z_{ij} take the value 1 if individual i has been captured before occasion j and 0 otherwise. We denote by $\mathscr{F}_j$ the trapping histories of all the individuals in the population. Then, given $\mathscr{F}_{j-1}$, the probability that individual i is captured on the jth trapping occasion given he has been captured at least once during the trapping experiment is

$$\gamma_{ij} = \operatorname{pr}(C_{ij}|\mathscr{F}_{j-1})/\operatorname{pr}(C_i|\mathscr{F}_{j-1}) = \Lambda_{ij}\left\{1-(1-z_{ij})\prod_{l=j}^{t}(1-\Lambda_{il}^{*})\right\}^{-1},$$

where Λ_{il}^{*} is Λ_{il} evaluated if individual i has not been captured before l, z_{ij} is as above, and it can be shown that

$$L^{*} = K\prod_{i=1}^{n}\prod_{j=1}^{t}\gamma_{ij}^{\Delta_{ij}}(1-\gamma_{ij})^{(1-\Delta_{ij})}\prod_{i=1}^{n}\operatorname{pr}(C_i)\prod_{i=n+1}^{\nu}\operatorname{pr}(\bar{C}_i).$$

We base our inference for the parameters associated with the model on the conditional likelihood

$$L = \prod_{i=1}^{n}\prod_{j=1}^{t}\gamma_{ij}^{\Delta_{ij}}(1-\gamma_{ij})^{(1-\Delta_{ij})},$$

which involves only the captured individuals and maximum conditional likelihood estimates of the parameters associated with the model may be found by the usual methods.

In § 3 we justify the use of L as an ordinary likelihood function for the purpose of asymptotic theory.

Simulations reveal that the distribution of the parameter estimators is not normal in small samples and we consider a conditional bootstrap procedure. The captured individuals and their covariates were treated as fixed and the capture experiment was repeatedly resimulated only for these individuals according to the estimated parameters.

For the data of Table 1 the maximum likelihood estimates, with estimates of their asymptotic standard deviations in parentheses, are $\hat{\beta}_0=-0{\cdot}680$ (1·09), $\hat{\beta}_1=0{\cdot}395$ (0·133), $\hat{\beta}_2=-0{\cdot}835$ (0·248) and $\hat{\alpha}=1{\cdot}068$ (0·575). The actual values were 0·5, 0·3, −1·0, 1·0. The bootstrap estimates of the standard deviations, from a small bootstrap sample of 100, were 2·042, 0·223, 0·185 and 1·068 respectively. The corresponding values from the simulation study were 1·369, 0·227, 0·206 and 0·523.

For the data of Coulombe the corresponding estimates and standard deviations were $\hat{\beta}_0=-0{\cdot}717$ (0·173), $\hat{\beta}_{\text{male}}=-0{\cdot}217$ (0·117), $\hat{\beta}_{\text{ju}}=-0{\cdot}230$ (0·258), $\hat{\beta}_{\text{sa}}=-0{\cdot}494$ (0·125), $\hat{\beta}_{\text{morning}}=0{\cdot}668$ (0·109), and $\hat{\alpha}=-0{\cdot}031$ (0·151). This agrees with Otis et al. (1978, pp. 62–4) but our model explains the heterogeneous capture probabilities via observable characteristics of the individuals and the time dependence via an observable characteristic of the trapping occasions.

To estimate the population size, usually our main concern rather than the capture probabilities, we use a method of moments.

First suppose θ is known. Let the probability that an individual is captured at least once during the course of the trapping experiment be denoted by

$$p_i(\theta) = \operatorname{pr}(C_i) = 1-\prod_{j=1}^{t}(1-\Lambda_{ij}^{*}).$$

Then if we set

$$\hat{\nu}(\theta)=\sum_{i=1}^{n} p_i^{-1}(\theta),$$

where as usual we denote the captured individuals by $i=1,\ldots,n$, we have

$$E\{\hat{\nu}(\theta)\}=E\left\{\sum_{i=1}^{n} p_i^{-1}(\theta)\right\}=E\left\{\sum_{i=1}^{\nu} I(C_i)p_i^{-1}(\theta)\right\}=\nu$$

because $E\{I(C_i)\}=p_i(\theta)$. Thus $\hat{\nu}(\theta)$ is an unbiased estimator of ν.

Similarly the variance of $\hat{\nu}(\theta)$ under our independence assumption is

$$\operatorname{var}\{\hat{\nu}(\theta)\}=\sum_{i=1}^{\nu}\{1-p_i(\theta)\}p_i^{-1}(\theta).$$

An unbiased estimator of $\operatorname{var}(\hat{\nu})$ is then easily seen to be

$$s^2=\sum_{i=1}^{n} p_i^{-2}(\theta)\{1-p_i(\theta)\}.$$

For the data of Table 1 the resulting estimators were $\hat{\nu}(\hat{\theta})=20{\cdot}86$, $s=1{\cdot}87$, whilst for the data of Coulombe we have $\hat{\nu}(\hat{\theta})=176{\cdot}9$ and $s=2{\cdot}01$. Note that these s values underestimate the standard error of $\hat{\nu}(\hat{\theta})$ as they assume θ is known. The estimated standard error for the data of Coulombe is comparable to those stated for the various methods of Otis et al. (1978, p. 64); however their estimates also do not appear to take into account the variance of their estimator of the capture probabilities. In § 3 below we see that the asymptotic variance of $\hat{\nu}(\hat{\theta})$ is of the form

$$\operatorname{var}\{\hat{\nu}(\hat{\theta})\}=s^2+\hat{D}_\theta^{\mathrm{T}}\hat{I}_\theta^{-1}\hat{D}_\theta$$

and for the data of Table 1 this gives a standard error of $\hat{\nu}(\hat{\theta})$ of $4{\cdot}51$. The conditional bootstrap estimator of the standard error based on 100 samples is $5{\cdot}89$. For the Coulombe data the estimated asymptotic standard error of $\hat{\nu}(\hat{\theta})$ is $3{\cdot}59$.

3. Asymptotic formulation

To obtain the asymptotic properties of our estimators we consider the following setting.

Suppose we have an increasing sequence of similarly composed populations. That is we take the x's to be the same for all populations and for each population we take the covariates z_i to be independent observations, for each individual, on some underlying distribution which is common to all populations. The parameters θ associated with the capture probabilities remain constant from population to population. Intuitively the situation may be thought of as corresponding to increasing the catchable population by increasing the size of the trapping region whilst keeping the trapping density constant. A common use of population size estimates is to compute an estimate of population density which is then used to estimate the population size over a much larger region. Thus if a preliminary estimate of population density is available the asymptotic results given here may help in designing the trapping experiment to give a required degree of precision. The later results show that if the population size is large enough then the distribution of $\hat{\nu}(\hat{\theta})$, the estimate of the population size evaluated at the estimated values of the parameters associated with the capture probabilities, is approximately normal. However for small populations such as that given in Table 1 simulations show that the distribution of $\hat{\nu}(\hat{\theta})$ is skewed to the right. Thus the use of the asymptotic variance of $\hat{\nu}(\hat{\theta})$ to give confidence intervals can be misleading.

To overcome problems caused by the nonnormality of the estimators for small samples the above conditional bootstrap method is proposed. This treats the captured individuals and their covariates as fixed and repeatedly resimulates the experiment using the estimated parameters. Note that we are conditioning on the observed number of captures and the observed covariates. In view of our conditional likelihood approach and the implicit assumption that the covariates are ancillary for θ this conditional bootstrap is quite acceptable for making inferences about θ. Note that the covariates need not be ancillary for ν. An obvious such case is where one is capturing fish in a small pond. Here if the population were large the sizes of the fish would be small and vice versa so the covariates corresponding to the sizes of the fish would contain some information about ν. However in this case it is not clear that it is possible to recover this information or what the unconditional distribution of the covariates corresponding to the sizes of the fish should be and the lesser of many evils appears to be to condition on the observed covariates.

Now n gives a ready lower bound for ν and we are conditioning on n in the bootstrap procedure so that for small populations it seems preferable to construct a $(1-\alpha)$ confidence interval for ν by deleting the upper α of the observations rather than the lower and upper $\frac{1}{2}\alpha$. Further in small samples the bootstrap distribution of the estimator has a short left tail and little is to be gained by deleting the lower $\frac{1}{2}\alpha$ of observations as rounding to the nearest integer will usually still give the lower bound as n.

Our procedure when applied to the captured individuals of Table 1 resulted in a 95% confidence interval, based on 100 bootstrap samples, of 18 to 30·9 which is reasonable, if somewhat conservative, compared with the simulated distribution of $\hat{\nu}$ when the number of captures is 18 which gives a 95% confidence interval of 18 to 24. The conservative nature of the bootstrap confidence interval for the data of Table 1 is to be expected as the estimated β_0 of $-0{\cdot}68$ is quite different from the actual β_0 of 0·5 whilst the other estimates are similar to the actual values. The effect of this is that the bootstrap distribution of $\hat{\beta}_0$ tends to be negative giving smaller values of $p_i(\hat{\theta})$ and hence larger $\hat{\nu}(\hat{\theta})$. The net effect is to increase the spread of the simulated distribution of $\hat{\nu}$. Note that the common procedure of giving a 95% confidence interval as being from the number of animals captured to $\hat{\nu}+2\surd\mathrm{var}(\hat{\nu})$ would give approximately the same interval as the conditional bootstrap. For moderately large data sets the bootstrap procedure may be prohibitively expensive and confidence intervals based on the estimated asymptotic standard deviation may be preferred.

The long-tailed nature of the distribution of $\hat{\nu}(\hat{\theta})$ in small samples and some identifiability problems with the parameters is the price one pays for using a conditional likelihood. To see this consider the model

$$\Lambda_{ij}=\frac{\exp(\beta+\alpha z_{ij})}{1+\exp(\beta+\alpha z_{ij})},$$

with 10 trapping occasions and suppose we capture only one individual with trapping history 0 0 0 0 1 1 0 1 1 0. The contribution to the likelihood from the first five occasions can be shown to be at most 0·1 which cannot be attained for $\beta>-\infty$. However, subject to the tolerances of a computer algorithm, an estimate will usually be found, say $\hat{\beta}=-6{\cdot}9$ which gives the contribution to the likelihood from these first five occasions to be 0·099991. The conditional likelihood can now be shown to be maximized, within a tolerance, for this $\hat{\beta}$ and $\hat{\alpha}=7{\cdot}305$. Calculation of $\hat{\nu}$ based on this one observation gives an estimate of the population size to be approximately 100. Our example also helps to explain the large bootstrap estimates of the standard errors of the corresponding parameters for the data of Table 1. This example clearly illustrates that the effect of no captures in the early trapping occasions is to give unrealistic parameter estimates and an excessively large estimate of the population size and to overcome this in practice it may be necessary to disregard all trapping occasions before a capture is recorded. In practice this lack of early captures would often indicate the individuals were not given time to adjust to the presence of the traps in their environment.

Our theoretical results depend on the assumption that the individuals behave independently of one another and this is essentially a design problem.

Finally there exist many unbiased estimators of ν of the form,

$$\nu^* = \sum_{i=1}^{n} \sum_{j=1}^{t} \Lambda_{ij}^{-1} \Delta_{ij} b_{ij}$$

where b_{ij} are functions of $(\Delta_{i1}, \ldots, \Delta_{i,j-1})$ and the covariates, with $E(b_{i1} + \ldots + b_{ij}) = 1$. These estimators perform poorly compared to $\hat{\nu}$ for they are more sensitive to the extreme parameter estimates resulting from the conditional likelihood than $\hat{\nu}$ and, as illustrated in the following example, perform poorly even when the Λ_{ij} are known.

Suppose trapping is conducted on three occasions and on each occasion there is a known constant capture probability of $\Lambda_{ij} = 0{\cdot}25$. Only three individuals were captured and their capture histories are: 1 0 1, 0 0 1, and 0 1 0. It can be shown that setting $b_{ij} = t^{-1}$ gives the minimum variance of ν^* in this case and an unbiased estimate of the variance of ν^* is

$$s^{*2} = t^{-2} \sum_{i=1}^{n} \sum_{j=1}^{t} \Delta_{ij}(1-p)/p^2.$$

The estimates of ν are then $\hat{\nu} = 5{\cdot}19$, $s^2 = 3{\cdot}79$ and $\nu^* = 5{\cdot}33$, $s^{*2} = 5{\cdot}33$. Such examples lead us to prefer $\hat{\nu}$ to ν^*.

4. Asymptotic properties of the estimators

The asymptotic properties of our estimators are determined in the following setting.

Consider a sequence of populations of individuals satisfying our earlier assumptions with the rth population being denoted by $i = 1, \ldots, \nu_r$. We suppose that $\nu_r \to \infty$ as $r \to \infty$ and for each r associate the covariates $z_i^{(r)}$ with the ith individual in the rth population. Further we assume that the trapping occasions $j = 1, \ldots, t$ remain the same for each population and that the environmental characteristics associated with the jth trapping occasion for the rth population, the $x_j^{(r)}$'s, are nonrandom and are the same for each population. It is supposed that for each fixed r the $z_i^{(r)}$ represent independent observations on some common underlying distribution so that we have a sequence of probabilistically similarly constituted populations. There is nothing in our assumptions to prevent the unconditional results below applying to a nested sequence of populations as the independence only needs hold for each fixed r and the same individuals may appear in different populations; however this is not necessary. If one considers a nested sequence of populations a modification of our argument shows that with probability 1 we have asymptotic normality conditional on the z_i's. For clarity we often omit reference to r in what follows but n_r shall always denote the number of individuals captured in the rth population. The z_{ij} remain the indicators of past capture.

We denote the derivative of the logarithm of the conditional likelihood by

$$S_r(\theta)=\frac{d\log L_r(\theta)}{d\theta}$$
$$=\sum_{i=1}^{n_r}\left[\sum_{j=1}^{t-1}\left(\frac{d\gamma_{ij}}{d\theta}\right)\left\{\frac{\Delta_{ij}-\gamma_{ij}}{\gamma_{ij}(1-\gamma_{ij})}\right\}+\left(\frac{d\Lambda_{it}}{d\theta}\right)\left\{\frac{\Delta_{ij}-\Lambda_{it}}{\Lambda_{it}(1-\Lambda_{it})}\right\}I(z_{it}=1)\right]$$
$$=\sum_{i=1}^{\nu_r}U_iI(C_i).$$

Note that, as

$$E\{(\Delta_{ij}-\gamma_{ij})I(C_i)|\mathscr{F}_{i-1}\}=E(\Delta_{ij}|\mathscr{F}_{i-1})-\Lambda_{ij}\operatorname{pr}(C_i|\mathscr{F}_{i-1})/\operatorname{pr}(C_i|\mathscr{F}_{i-1})=0, \qquad (4\cdot1)$$

we have that $E\{U_iI(C_i)\}=0$. Thus $S_r(\theta)$ is the sum of ν_r independently and identically distributed random vectors with zero means and variance-covariance matrix

$$i_\theta=E\left[\sum_{j=1}^{t-1}\left(\frac{d\gamma_{ij}}{d\theta}\right)^{\otimes2}\left\{\frac{\Lambda_{ij}}{\gamma_{ij}^2(1-\gamma_{ij})}\right\}+\left(\frac{d\Lambda_{it}}{d\theta}\right)^{\otimes2}\left\{\frac{\operatorname{pr}(z_{it}=1|\mathscr{F}_{t-1})}{\Lambda_{it}(1-\Lambda_{it})}\right\}\right],$$

where for a vector a, $a^{\otimes2}$ denotes aa^{T}.

In view of (4·1) we have

$$E(S_r^{\otimes2})=dS_r(\theta)/d\theta,$$

so that in particular the usual arguments imply the consistency of $\hat\theta_r$ and that asymptotically the distribution of $\nu_r^{1/2}(\hat\theta_r-\theta)$ is multivariate normal with covariance matrix $i(\theta)^{-1}$.

Next note that an unbiased estimator of $I_\theta=\nu_r i_\theta$ is given by

$$\hat I_\theta=\sum_{i=1}^{n}\left[\sum_{j=1}^{t}\left(\frac{d\gamma_{ij}}{d\theta}\right)^{\otimes2}\left\{\frac{\Delta_{ij}}{\gamma_{ij}^2(1-\gamma_{ij})}\right\}+\left(\frac{d\Lambda_{it}}{d\theta}\right)^{\otimes2}\left\{\frac{\Delta_{ij}I(z_{it}=1)}{\Lambda_{it}^2(1-\Lambda_{it})}\right\}\right].$$

In many cases an estimate of I_θ can be recovered from the algorithms used to compute $\hat\theta$; that is it is often convenient to use the estimator

$$\hat I_\theta=-\{dS_r(\hat\theta)/d\hat\theta\}.$$

Similarly

$$\hat\nu_r(\theta)-\nu_r=\sum_{i=1}^{\nu_r}\left\{\frac{I(C_i)-p_i(\theta)}{p_i(\theta)}\right\}$$

is the sum of ν_r independently and identically distributed random variables with means 0 and variance $a(\theta)=p_i(\theta)^{-1}\{1-p_i(\theta)\}$.

Note that the covariance between $S_r(\theta)$ and $\hat\nu_r(\theta)$ is given by

$$\operatorname{cov}\{S_r(\theta),\hat\nu(\theta)\}=E\left[\sum_{i=1}^{\nu_r}\frac{U_iI(C_i)\{I(C_i)-p_i(\theta)\}}{p_i(\theta)}\right]$$
$$=E\left\{\sum_{i=1}^{\nu_r}\frac{U_i(C_i)}{p_i(\theta)}\right\}-E\left\{\sum_{i=1}^{\nu_r}U_iI(C_i)\right\}=0,$$

for as we have observed above, $E\{U_iI(C_i)\}=0$.

Using standard methods for the sums of independently and identically distributed random vectors it is apparent that the limiting distribution of the vector

$$\nu_r^{-\frac{1}{2}}(S_r^{\mathrm{T}}(\theta),\hat\nu_r-\nu_r)^{\mathrm{T}}$$

will be multivariate normal with covariance matrix, diag $\{i(\theta), a(\theta)\}$ and this convergence may be taken to be mixing.

Standard likelihood type methods may now be applied to show that the asymptotic distribution of $(\nu_r^{\frac{1}{2}}(\hat{\theta}_r-\theta)^{\mathrm{T}}, \nu_r^{-\frac{1}{2}}(\hat{\nu}_r-\nu_r))^{\mathrm{T}}$ is multivariate normal with covariance matrix diag $\{i^{-1}(\theta), a(\theta)\}$.

Now denote by $\hat{\nu}_r(\theta)$ our estimator of ν_r when θ is known and by $\hat{\nu}_r(\hat{\theta}_r)$ the corresponding estimator when θ is replaced by $\hat{\theta}_r$. Then as long as the Λ_{ij} are smooth enough functions of θ we may write using a first-order approximation,

$$\hat{\nu}_r(\hat{\theta}_r)-\nu_r=\hat{\nu}_r(\theta)-\nu_r+(\hat{\theta}_r-\theta)^{\mathrm{T}}\left[\frac{d}{d\theta}\hat{\nu}_r(\theta)\right]_{\theta^*},$$

where θ^* lies between $\hat{\theta}_r$ and θ.

Now

$$\nu_r^{-1}\frac{d\hat{\nu}_r(\theta)}{d\theta}=-\nu_r^{-1}\sum_{i=1}^{\nu_r}p_i^{-2}(\theta)I(C_i)\frac{dp_i(\theta)}{d\theta},$$

which converges in probability to some quantity $d(\theta)$ by the law of large numbers. Hence the limiting distribution of $\nu_r^{-1/2}\{\hat{\nu}_r(\hat{\theta}_r)-\nu_r\}$ will be normal with variance $c(\theta)=a(\theta)+d(\theta)^{\mathrm{T}}i(\theta)^{-1}d(\theta)$, so that, for ν_r large enough, $\hat{\nu}_r(\hat{\theta}_r)$ has approximately a multivariate normal distribution with mean ν_r and variance

$$s^2+\hat{D}_\theta^{\mathrm{T}}\hat{I}_\theta^{-1}\hat{D}_\theta,$$

where $\hat{D}_\theta$ is the unbiased estimator of $\nu_r d(\theta)$ given by $d\nu_r(\theta)/d\theta$ evaluated at $\hat{\theta}_r$.

Using this asymptotic normality it is now possible to construct confidence intervals for ν as discussed in the introduction.

In the case where the Λ_{ij} or the a_i are not smooth enough to use the mean value theorem as above it may be necessary to consider the approach of Randles (1982).

REFERENCES

BURNHAM, K. P. & OVERTON, W. S. (1978). Estimation of the size of a closed population when capture probabilities vary among animals. *Biometrika* **65**, 625-33.

OTIS, D. L., BURNHAM, K. P., WHITE, G. C. & ANDERSON, D. R. (1978). Statistical inference from capture data on closed animal populations. Wildlife Monograph No. 62. Washington, D.C.: The Wildlife Soc.

RANDLES, R. H. (1982). On the asymptotic normality of statistics with estimated parameters. *Ann. Statist.* **10**, 462-74.

SEBER, G. A. F. (1982). *The Estimation of Animal Abundance*, 2nd ed. New York: Macmillan.

[*Received October* 1986. *Revised October* 1988]