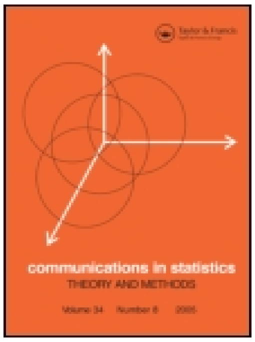

Communications in Statistics - Theory and Methods

ISSN: 0361-0926 (Print) 1532-415X (Online) Journal homepage: http://www.tandfonline.com/loi/lsta20

ASYMPTOTIC PROPERTIES OF AN OPTIMAL ESTIMATING FUNCTION APPROACH TO THE ANALYSIS OF MARK RECAPTURE DATA

Richard M. Huggins & Anne Chao

To cite this article: Richard M. Huggins & Anne Chao (2002) ASYMPTOTIC PROPERTIES OF AN OPTIMAL ESTIMATING FUNCTION APPROACH TO THE ANALYSIS OF MARK RECAPTURE DATA, Communications in Statistics - Theory and Methods, 31:4, 575-595, DOI: 10.1081/STA-120003135

To link to this article: http://dx.doi.org/10.1081/STA-120003135

Published online: 02 Sep 2006.

Submit your article to this journal

Article views: 27

View related articles

Citing articles: 1 View citing articles

Full Terms & Conditions of access and use can be found at
http://www.tandfonline.com/action/journalInformation?journalCode=lsta20

Download by: [National Tsing Hua University] **Date:** 21 June 2016, At: 06:33

COMMUN. STATIST.—THEORY METH., 31(4), 575–595 (2002)

INFERENCE

ASYMPTOTIC PROPERTIES OF AN OPTIMAL ESTIMATING FUNCTION APPROACH TO THE ANALYSIS OF MARK RECAPTURE DATA

Richard M. Huggins[1,*] **and Anne Chao**[2]

[1]Department of Statistical Science, La Trobe University, Bundoora 3086, Australia
E-mail: r.huggins@latrobe.edu.au
[2]Institute of Statistics, National Tsing Hua University, Hsin-Chu, Taiwan 30043
E-mail: chao@stat.nthu.edu.tw

ABSTRACT

A drawback of non parametric estimators of the size of a closed population in the presence of heterogeneous capture probabilities has been their lack of analytic tractability. Here we show that the martingale estimating function/sample coverage approach to estimating the size of a closed population with heterogeneous capture probabilities is mathematically tractable and develop its large sample properties.

Key Words: Heterogeneous capture probabilities; Capture–recapture experiment; Sample coverage

*Corresponding author.

Copyright © 2002 by Marcel Dekker, Inc. www.dekker.com

Downloaded by [National Tsing Hua University] at 06:33 21 June 2016

Downloaded by [National Tsing Hua University] at 06:33 21 June 2016

1. INTRODUCTION

Mark-recapture sampling is commonly used to estimate the size of a closed population. In a typical capture–recapture experiment in the biological and ecological sciences, animals are captured, uniquely marked and released into the population at each of several trapping occasion. At the end of the experiment, the complete capture history for each captured animal is known. Comprehensive reviews of this topic are provided in (1–4). The traditional models assume that on a given capture occasion all animals have the same capture probabilities. However, it has been long recognized by biologists and ecologists that the equal-catchability assumption is an unattainable ideal in natural populations (5, p. 146). Many previous studies (e.g., 6,7) have confirmed that heterogeneity among capture probabilities cause negative bias for traditional estimators based on equal-catchability.

Burnham, in an unpublished Ph.D. thesis (Oregon State University, 1972) was the first to consider heterogeneous models. This model, which is usually referred as model M_h in the literature, (6,8) assumes that each animal has its own capture probability, which remains constant over the capture occasions and is not altered by previous capture. Specifically, it is assumed that there are N animals whose capture probabilities are $\{p_1, p_2, \ldots, p_N\}$. One way to reduce the number of parameters is to assume that $\{p_1, p_2, \ldots, p_N\}$ is a random sample from some distribution. For example, Burnham in his Ph.D. thesis considered a two-parameter beta distribution, but he found that the maximum likelihood estimators were not satisfactory. The non-parametric approach makes no assumption about the form of the distribution. A jackknife estimator in a non-parametric framework has been proposed (7,9) and its practical use recommended (6,10). Other non-parametric approaches include the bootstrap estimator (11), sample coverage estimator (12,13), a loglinear model approach (14,15), and the non-parametric MLE (16).

The justification of many these estimators is through simulation studies as the estimators have tended to be analytically intractable. The simulations typically show that non-parametric estimators sometimes perform well and at other times perform poorly. However, (17) used optimal martingale estimating equations and sample coverage to propose a unified approach to the estimation of the size of a closed population which we show to be analytically tractable in the case of model M_h. For a review of the estimating function approach, see (18). The estimating function approach has also been applied to other capture–recapture models, for example, to beta-binomial (19), to the removal model (20) and to model M_{tb}, that allows capture probabilities to vary with time and in response to previous capture but not between individuals (21). The idea of sample coverage, originally due to I. J. Good and A. M. Turing (22), has been used in species and population size estimation. The basic idea is that, whereas it is difficult to estimate population size when capture probabilities vary among animals, the sample coverage can nonetheless be well estimated in such a case. Therefore, in the optimal martingale approach, we first estimate the sample coverage and then use it to estimate the population size.

This paper establishes asymptotic properties of the optimal estimating function/sample coverage approach for model M_h. We illustrate these properties when the capture probabilities have a beta distribution as this is a common model for capture probabilities. We show that the bias is reasonable for a range of beta distributions but can become quite large in some cases. These cases correspond to a proportion of individuals that are essentially uncatchable. We also examine the behaviour of a bootstrap estimator of the variance. In Section 2 we outline the derivation of the estimators for model M_h. In Section 3 our general results are presented, with technical details in Appendices A.1–A.7. In Section 4 we present results on the asymptotic biases when the capture probabilities have a beta distribution. The results are discussed in Section 5.

2. THE MARTINGALE ESTIMATORS

We first derive the optimal estimating equations and the resulting estimator of (17). Suppose the population consists of N individuals. We assume the capture probabilities of these N individuals are randomly sampled from a distribution with mean μ_p and variance σ_p^2, and assume the individuals behave independently of one another. Denote the resulting vector of probabilities by $\underset{\sim}{p} = (p_1, p_2, \ldots, p_N)^T$, the conditional mean of these capture probabilities by $\bar{p} = \sum_{i=1}^N p_i/N$ and the conditional coefficient of variation (CV) by $\gamma = [\sum_{i=1}^N (p_i - \bar{p})^2/N]^{1/2}/\bar{p}$. Consider a capture experiment conducted on occasions $j = 1, 2, \ldots, t$. Let u_j denote the number of individuals captured for the first time on occasion j, m_j the number of previously marked animals captured on occasion j, and n_j denote the total number of animals captured on occasion j. Similarly let M_j denote the number of individuals captured on occasions $1, 2, \ldots, j-1$, with $M_1 = 0$, let $f_{l,j}$ be the number of individuals captured exactly l times on occasions $1, \ldots, j$ and let $n = M_{t+1}$ denote the total number of distinct individuals captured. Let X_{ij} take the value 1 if individual i is captured on occasion j and 0 otherwise. The sample coverage of samples $1, 2, \ldots, j$ is defined by $C_j = \sum_{i=1}^N p_i I(\sum_{k=1}^j X_{ik} > 0)/\sum_{i=1}^N p_i$, with $C_0 = 0$.

Downloaded by [National Tsing Hua University] at 06:33 21 June 2016

We show in A.1 that given $p_1, \ldots, p_N$, $E(u_k|\mathcal{F}_{k-1}) = (N - NC_{k-1})\bar{p}$, similarly, $E(m_k|\mathcal{F}_{k-1}) = NC_{k-1}\bar{p}$, $\text{var}(u_k|\mathcal{F}_{k-1}) = (1 - C_{k-1})(\sum_{i=1}^N p_i) - (1 - C'_{k-1})(\sum_{i=1}^N p_i^2)$, where $C'_k = \sum_{i=1}^N p_i^2 I[\sum_{j=1}^k X_{ij} > 0] / \sum_{i=1}^N p_i^2$. Under the assumption that $C_{k-1} \approx C'_{k-1}$, which results in suboptimal but estimable weights, the approximations $\text{var}(u_k|\mathcal{F}_{k-1}) \approx (N - NC_{k-1})\bar{p}\,[1 - (1 + \gamma^2)\bar{p}]$, and $\text{var}(m_k|F_{k-1}) \approx NC_{k-1}\,\bar{p}\,[1 - (1 + \gamma^2)\,\bar{p}\,]$ are used in (17) to derive approximately optimal weights.

If NC_{k-1} is treated as a function of N, it is only possible to estimate the confounded multiplicative effect $N\bar{p}$ and we would not be able to separately estimate the parameters N and $\bar{p}$. Let $M_k^* = NC_{k-1}, k \geq 2$, $M_1^* = 0$ and $\mathbf{g}_k = (g_{1k}, g_{2k})^T = (u_k - (N - M_k^*)\bar{p}, m_k - M_k^*\bar{p})^T$, so that $E(\mathbf{g}_k|\mathcal{F}_{k-1}) = 0$. Let $\theta = (N, \bar{p})^T$. The optimal estimating equations are given by $\mathbf{g} = \sum_{k=1}^t \mathbf{D}_k^T \mathbf{V}_k^{-1} \mathbf{g}_k = 0$, where $\mathbf{D}_k = E(\partial \mathbf{g}_k / \partial\theta|\mathcal{F}_{k-1})$ and $\mathbf{V}_k = \text{Cov}(\mathbf{g}_k|\mathcal{F}_{k-1})$. Assuming that C_{k-1} and M_k^* are known, and noting that $\text{cov}(g_{1k}, g_{2k}) = 0$, and using the approximations to the variances noted above, the optimal estimating equations are of the form

$$\sum_{j=1}^{t} (1 - C_{j-1})^{-1} \{u_j - (N - M_j^*)\bar{p}\} = 0, \tag{2.1}$$

and $\sum_{j=1}^t (n_j - N\bar{p}) = 0$. The latter equation yields

$$\bar{p} = (tN)^{-1} \sum_{j=1}^{t} n_j. \tag{2.2}$$

The asymptotic properties of the resulting estimators for known M_j^* are derived in A.3 as an intermediate step in deriving the properties of the final estimator. We follow (13,17), and estimate M_j^* by

$$\hat{M}_j^* = M_j + \hat{\gamma}^2 f_{1,\,j-1}. \tag{2.3}$$

where $\gamma^2 = N^{-1} \sum_{i=1}^N (p_i - \bar{p})^2 / \bar{p}^2$ which is estimated by

$$\hat{\gamma}^2 = \max\left\{ \frac{\hat{N}_0 t \sum_k k(k-1) f_{k,\,t}}{(t-1)\left(\sum_k n_k\right)^2} - 1, 0 \right\}$$

with $\hat{N}_0 = M_{t+1}/\hat{C}_t$, where $\hat{C}_t$ is defined below. This estimator of the CV is examined in A.2 and the approximation (2.3) in A.5. The sample coverage is estimated by $\hat{C}_{j-1} = 1 - f_{1j} / \sum_{k=1}^j n_k$. We show in A.2 that $\hat{C}_{j-1}$ is asymptotically equivalent to C_{j-1} in the sense that they both have the same

Downloaded by [National Tsing Hua University] at 06:33 21 June 2016

Downloaded by [National Tsing Hua University] at 06:33 21 June 2016

limit as $N \to \infty$. Substitution of (2.2) in (2.1) and replacing the unknown quantities by their estimates, results in the estimator

$$\hat{N} = \frac{\sum_{j=1}^{t} \left(1 - \hat{C}_{j-1}\right)^{-1} \hat{M}_j^*}{\sum_{j=1}^{t} \left(1 - \hat{C}_{j-1}\right)^{-1} \left(1 - tu_j / \sum n_k\right)}.$$

A bootstrap procedure based on the multinomial to estimate the standard errors of the estimator of (17). The bootstrap procedure proposed in (17) is equivalent to resampling $\hat{N}$ capture histories from a population of size $\hat{N}$ consisting of the n observed capture histories augmented by $\hat{N} - n$ uncaptured individuals. The simulations of (17) also indicate that in the cases they examine, the bootstrap estimate of the variance performs well with perhaps a slight tendency to overestimate the variance of the estimator for larger values of α and β and underestimate the variance for smaller values. As noted in the proof of Theorem 2, we have not been able to examine this bootstrap procedure analytically. Therefore we consider an alternative bootstrap estimator where we resample n capture histories from the n observed capture histories. This is similar to the multinomial model of (17) but we do not resample from the estimated number of uncaptured individuals. In this approach we are effectively conditioning on the number of captured individuals. We conjecture that the bootstrap procedure of (17) may eventually be found to be preferable.

Here we investigate the asymptotic properties of the estimator. These results have limited practical application but do indicate when the optimal martingale/sample coverage estimator may be expected to perform well and when it may perform badly. Our approach is based on showing that asymptotically the estimator is the weighted sum of a sum of unconditionally independently and identically distributed zero mean random vectors plus a bias term. In particular, we are able to apply the central limit theorem to the sum and use the representation to asymptotically examine a possible bootstrap estimator of the variance.

3. MAIN RESULTS

The martingale approach allows the use of standard arguments to establish a central limit theorem.

Theorem 1. *Let* $\mu_p = E(p)$, *where* p *has the common distribution of the* p_i, $\tilde{C}_j = E(p(1-(1-p)^j))/\mu_p$, $B^* = \sum_{j=1}^{t} (1-\tilde{C}_{j-1})^{-1} \tilde{C}_{j-1}$, $A_j = B^{*-1}(1-\tilde{C}_{j-1})^{-1}$,

Downloaded by [National Tsing Hua University] at 06:33 21 June 2016

$\tilde{A}_j = A_j/\mu_p$, $A = (\tilde{A}_1, \ldots, \tilde{A}_t, A_2, \ldots, A_t)^T$, $\bar{A} = (A_2, \ldots, A_t)^T$, *and* $\delta = (\delta_2, \delta_3, \ldots, \delta_t)^T$ *where*

$$\delta_j = E\left\{(1-p)^{j-2}\left(\left(1 - \frac{p}{\mu_p}\right)(1-p) - \gamma^{*2}(j-1)p\right)\right\}$$

with $\gamma^{*2} = E(1-(1-p)^t)\sigma_p^2/\{(E(p(1-(1-p)^t))\mu_p)\}$, σ_p^2 *being the variance of the capture probabilities, and* Σ *a* $(2t-1) \times (2t-1)$ *covariance matrix defined in* A.6. *Then*, $N^{-1/2}(\hat{N} - N + N\bar{A}^T\delta) \xrightarrow{d} N(0, A^T\Sigma A)$.

Proof. Let

$$\tilde{N} = \frac{\sum_{j=1}^t \left(1-\hat{C}_{j-1}\right)^{-1} M_j^*}{\sum_{j=1}^t \left(1-\hat{C}_{j-1}\right)^{-1}\left(1 - u_j/\sum_i p_i\right)} = B^{-1}\sum_{j=1}^t \left(1-\hat{C}_{j-1}\right)^{-1} M_j^*,$$

where we have also replaced $\sum n_k/t$ by $\sum p_i$ as $(Nt)^{-1}\sum_{i=1}^N p_i$ and $N^{-1}\sum_{k=1}^t n_k$ have the same limit. We show in A.3. that $\tilde{N}$ is an asymptotically unbiased estimator of N. Furthermore, from A.5, asymptotically,

$$\lim_{N\to\infty} N^{-1}(\tilde{N} - \hat{N}) = B^{*-1}\sum_{j=1}^t \left(1-\tilde{C}_{j-1}\right)^{-1} \lim_{N\to\infty} N^{-1}\left(M_j^* - \hat{M}_j^*\right)$$
$$= B^{*-1}\sum_{j=1}^t \left(1-\tilde{C}_{j-1}\right)^{-1}\delta_j.$$

Let $A = (\tilde{A}^T, \bar{A}^T)^T$, $W = (Y^T, V^T)^T$ where Y and V are vectors with elements $u_j - E(u_j|\mathcal{F}_{j-1})$ and $\hat{M}_j^* - M_j^* + N\delta_j$ respectively as defined in A.1 and A.5. Then W has mean vector zero, and unconditionally is the sum of i.i.d. vectors W_i that each have a covariance matrix denoted by Σ. This matrix is examined in A.6. Now, asymptotically,

$$\begin{aligned} N^{-1/2}\ (\hat{N} - N) &= N^{-1/2}\ (\tilde{N} - N) + N^{-1/2}\ (\hat{N} - \tilde{N}) \\ &= N^{-1/2}\tilde{A}^T Y + N^{-1/2}\bar{A}^T V - N^{1/2}\bar{A}^T\delta \\ &= N^{-1/2}A^T W - N^{1/2}\bar{A}^T\delta. \end{aligned}$$

Thus we may apply the central limit theorem to determine the asymptotic distribution of our estimator and hence,

$$N^{-1/2}(\hat{N} - N + N\bar{A}^T\delta) \xrightarrow{d} N(0, A^T\Sigma A).$$

Downloaded by [National Tsing Hua University] at 06:33 21 June 2016

Theorem 1 contains an expression for the asymptotic bias in the general case. We determine in A.6 that the asymptotic covariance matrix of the estimator is complex and difficult to estimate, and thus one cannot expect to use an analytic form of the variance in practice. We now state the general properties of our bootstrap estimator. Let Δ_{11} denote the $t \times t$ matrix with common entries $[E(p^2(1-p)^t)+E^2(p(1-p)^t)]$, Δ_{22} the $(t-1)\times(t-1)$ matrix $\delta\delta^T E((1-p)^t)\{1+E((1-p)^t)\}$, Δ_{21} the $(t-1)\times t$ matrix $\delta \underset{\sim}{1}^T E(p(1-p)^t)\{1+E((1-p)^t)\}$ and $\Delta_{12}=\Delta_{21}^T$. Define

$$\Delta = \begin{pmatrix} \Delta_{11} & -\Delta_{12} \\ -\Delta_{21} & \Delta_{22} \end{pmatrix}.$$

Theorem 2. *Let $\tilde{V}(\hat{N})$ denote the bootstrap estimate of the variance $V(\hat{N})$ of $\hat{N}$. Then, $N^{-1}(V(\hat{N})-\tilde{V}(\hat{N})) \xrightarrow{p} A^T\Delta A$.*

Proof. First let $\xi_i = 1$ if individual i is captured at least once and 0 otherwise. Let $Y_i=(Y_{i1},\ldots,Y_{it})^T$, recall $V_i=(V_{i1},\ldots,V_{it})^T$ and let $W_i=(Y_i^T,V_i^T)^T$ so that $W=\sum_{i=1}^N W_i$. We are using limiting arguments so that the vector A may be regarded as a constant. Unconditionally the $W_i\xi_i$ are independently and identically distributed, with covariance matrix $\tilde{\Sigma}\neq\Sigma$ that is examined in A.7. Hence applying the bootstrap to $W_1\xi_1,\ldots,W_N\xi_N$ allows the estimation of $\tilde{\Sigma}$. Then $n^{-1}\sum_{i=1}^n W_iW_i^T\xi_i = n^{-1}NN^{-1}\sum_{i=1}^N W_iW_i^T\xi_i$ as $\xi_i=0$ for uncaptured individuals. Now $N^{-1}n \xrightarrow{p} E\big(1-(1-p)^t\big)$ and $N^{-1}\sum_{i=1}^N W_iW_i^T\xi_i \xrightarrow{p} \tilde{\Sigma}$, hence $n^{-1}\sum_{i=1}^{\hat{N}} W_iW_i^T\xi_i \xrightarrow{p} \tilde{\Sigma}/E\big(1-(1-p)^t\big)=\Sigma^*$. We thus conclude that the bootstrap applied to the captured individuals $W_1\xi_1,\ldots,W_n\xi_n$, estimates Σ^*. Thus, asymptotically, the bootstrap estimate $\tilde{V}(\hat{N})$ of $V(\hat{N})$ is $nA^T\tilde{\Sigma}A$ and hence for large N, $V(\hat{N})-\tilde{V}(\hat{N})\approx NA^T\Sigma A - E\big(1-(1-p)^t\big)^{-1}nA^T\tilde{\Sigma}A$. As noted above, $N^{-1}\hat{n}\xrightarrow{p} E\big(1-(1-p)^t\big)$ which along with the definition of Δ, yields the Theorem.

We have not been able to apply this argument to the bootstrap procedure proposed by (17) as it is possible that $\hat{N} > N$ which requires more complex arguments.

4. THE BETA (α, β) CASE

The key to the estimator is the approximation (2.3) and the estimation of the coefficient of variation, both of which introduce some bias. The general form for the bias is given in Theorem 1 and the bias in bootstrap estimate of the variance in Theorem 2. As the beta distribution is a common model for heterogeneous probabilities and our results are most transparent in this case, we examine the bias when the capture probabilities have a beta distribution.

Theorem 3. *If the capture probabilities have a beta* (α, β) *distribution then the asymptotic bias of* $\hat{N}$ *as a proportion of* N is $-P(\alpha\beta)$ *where* $P(\alpha,\beta) = B^{*-1}\sum_{j=2}^{t}(1-\tilde{C}_{j-1})^{-1}d_j$, *with* $\tilde{C}_{j-1} = 1 - \mathcal{B}(\alpha+1, \beta+j-1)(\alpha+\beta)/\ (\mathcal{B}(\alpha,\beta)\alpha)$, $B^{*} = \sum_{j=1}^{t}(1-\tilde{C}_{j-1})^{-1}\tilde{C}_{j-1}$,

$$d_j = \frac{\mathcal{B}(\alpha, \beta+j-1)}{\mathcal{B}(\alpha,\beta)} - \frac{(\alpha+\beta)\mathcal{B}(\alpha+1, \beta+j-1)}{\alpha\mathcal{B}(\alpha,\beta)} - \frac{a\beta(j-1)\mathcal{B}(\alpha+1,\beta+j-2)}{\alpha(\alpha+\beta+1)\mathcal{B}(\alpha,\beta)}$$

with $\mathcal{B}$ *denoting the beta function and*

$$a = \frac{(\alpha/(\alpha+\beta)(1-\mathcal{B}(\alpha,\beta+t)/\mathcal{B}(\alpha,\beta)))}{(\alpha/(\alpha+\beta) - \mathcal{B}(\alpha+1,\beta+t)/\mathcal{B}(\alpha,\beta))}.$$

Proof. To obtain Theorem 3, we need to evaluate

$$E\{(1-p)^{k-2}((1-p/\mu_p)(1-p) - \gamma^{*2}(k-1)p\ )\} = E((1-p)^{k-1}) - \frac{1}{\mu_p}E(p(1-p)^{k-1}) - \gamma^{*2}(k-1)E(p(1-p)^{k-2})$$

for the beta(α, β) distribution. Now

$$E(p(1-p)^{k-1}) = \frac{\mathcal{B}(\alpha+1,\beta+k-1)}{\mathcal{B}(\alpha,\beta)},$$

$$E(p(1-p)^{k-2}) = \frac{\mathcal{B}(\alpha+1,\beta+k-2)}{\mathcal{B}(\alpha,\beta)} \qquad E(p) = \frac{\alpha}{\alpha+\beta},$$

$$E((1-p)^{k-1} = \frac{\mathcal{B}(\alpha,\beta+k-1)}{\mathcal{B}(\alpha,\beta)}, \qquad \gamma^2 = \frac{\beta}{\alpha(\alpha+\beta+1)},$$

and $\gamma^{*2} = a\gamma^2$, where a is a^* of A.4 evaluated for the beta (α, β) distribution as given in the statement of Theorem 3, which, once we observe that the asymptotic bias is the limit of $N^{-1}(\hat{N} - \tilde{N})$, yields Theorem 3.

In Figure 1 we plot $-P(\alpha, \beta)$ against α and β for $\alpha, \beta \in [0.1, 5]$ for $t = 5$. This Figure shows that for $\alpha > 2$ the bias is reasonable, being less than 10%. However, for small values of α and β the bias becomes large approaching a maximum of 90% for values of α close to 0.1. Comparisons with the simulation results of (17) suggest that the asymptotic bias slightly underestimates the true bias in finite samples.

Downloaded by [National Tsing Hua University] at 06:33 21 June 2016

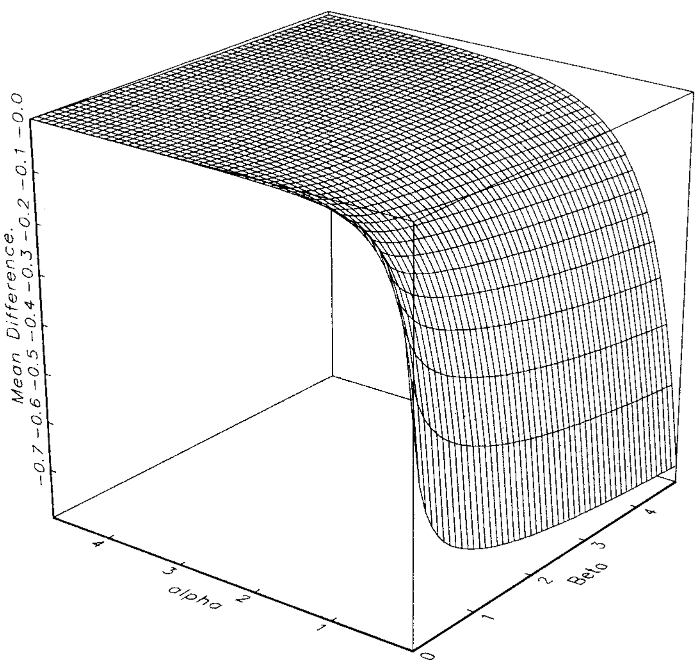

Figure 1. The bias as a proportion of the population size for $t=5$ capture occasions as a function of α and β when the capture probabilities have a beta (α, β) distribution. (The minimum plotted values of α and β are 0.1).

As a Corollary of Theorem 2 we have:

Theorem 4. *Let* $d = (d_2, \dots, d_t)^T$ *and let* b *be the vector with elements* $b_j = B^{*-1}(1-\tilde{C}_{j-1})^{-1}(\alpha+\beta)/\alpha$, $j = 1, \dots, t$ *and* $b_{t-1+k} = B^{*-1}(1-\tilde{C}_{k-1})^{-1}$, $k = 2, \dots, t$. *Further, define the matrix*

$$\tilde{\Delta} = \begin{pmatrix} \tilde{\Delta}_{11} & -\tilde{\Delta}_{12} \\ -\tilde{\Delta}_{21} & \tilde{\Delta}_{22} \end{pmatrix}$$

where $\tilde{\Delta}_{11}$ *has common entries* $\mathcal{B}(\alpha+2, \beta+t)/\mathcal{B}(\alpha, \beta) + (\mathcal{B}(\alpha+1, \beta+t)/\mathcal{B}(\alpha, \beta))^2$, $\tilde{\Delta}_{22} = dd^T \times c_1$ *where* $c_1 = \mathcal{B}(\alpha, \beta+t)/\mathcal{B}(\alpha, \beta)(1+\mathcal{B}(\alpha, \beta+t)/\mathcal{B}(\alpha, \beta))$, $\tilde{\Delta}_{21} = d\mathbf{1}^T \times c_2$ *where* $c_2 = \mathcal{B}(\alpha+1, \beta+t)/\mathcal{B}(\alpha, \beta)(1+\mathcal{B}(\alpha, \beta+t)/\mathcal{B}(\alpha, \beta))$ *and* $\tilde{\Delta}_{12} = \tilde{\Delta}_{21}^T$. *Let* $\tilde{V}(\hat{N})$ *denote our bootstrap estimate of the variance. If the capture probabilities have a beta* (α, β) *distribution then the asymptotic bias of* $\tilde{V}(\hat{N})$, *as a proportion of* N *is* $V(\alpha, \beta) = -b^T\tilde{\Delta}b$.

Downloaded by [National Tsing Hua University] at 06:33 21 June 2016

Thus asymptotically we expect the bootstrap considered here to underestimate the variance. However, as we have used asymptotic results which ignore the finite sample sources of variation involved in estimating the sample coverage and the coefficient of variation, this may not hold for smaller samples. The simulation results of (17) reveal in some cases their bootstrap slightly over estimates the variance. This may be due to the extra variability in that (17) do not condition on the number of captures. Moreover, as it is possible that $\hat{N} > N$ and a version of our argument in the proof of Theorem 2 may be applied when $\hat{N} < N$ it is feasible that the bias in the (17) bootstrap estimator may be smaller.

In Figure 2 we plot $V(\alpha, \beta)$ against α and β. This figure is consistent with the finite sample simulations of (17) in that the bias is small for larger values of α and β but becomes quite large for small values of α. The largest bias occurs when α is small and β is large and in this case the capture probabilities are uniformly low.

Downloaded by [National Tsing Hua University] at 06:33 21 June 2016

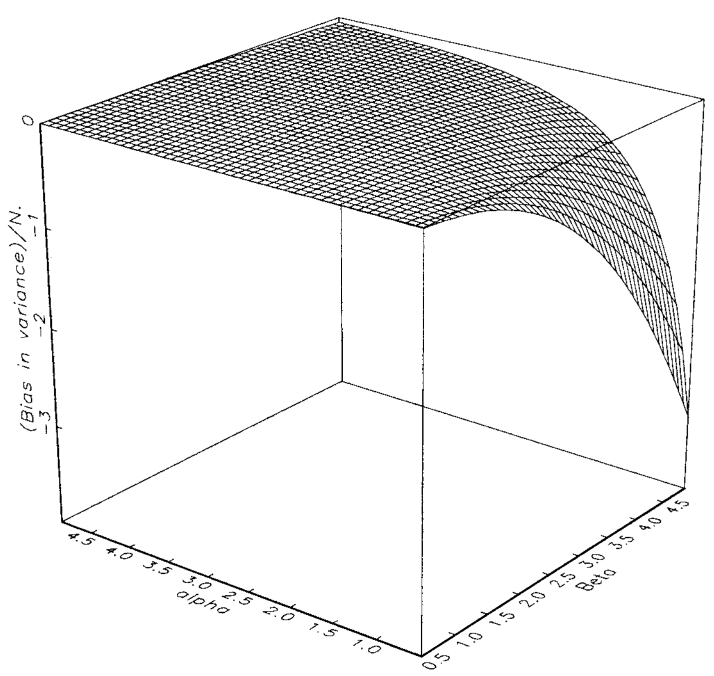

Figure 2. The bias in the bootstrap estimate of the variance as a proportion of the population size for $t = 5$ capture occasions as a function of α and β when the capture probabilities have a beta (α, β) distribution. (The minimum plotted value of α and β is 0.5).

Downloaded by [National Tsing Hua University] at 06:33 21 June 2016

5. DISCUSSION

Our asymptotic results show that as suggested by the simulations of (17) and others the sample coverage estimators perform well for a range of distributions for the capture probabilities but can perform poorly in some cases. In these cases a proportion of the individuals have small capture probabilities. The bias in the estimated population size is sensitive to these low capture probabilities. The bias in the bootstrap estimate of the variance considered here behaves a little differently in that it increases dramatically as a proportion of the population size as the capture probabilities cluster around zero.

To remedy the bias in the estimator, one suggestion could be that extra terms be included in the approximation (2.3). However, for $\alpha < 1$ there are a number of essentially uncatchable individuals in the population, which would appear to be the source of the bias in the sample coverage estimators. A possible resolution of this problem is to note that if there are a number of individuals with low capture probabilities then the catchable population increases with effort suggesting that a comparison of estimates based on subsets of the capture occasions may provide a useful test. Alternatively a sequential procedure may be considered. Moreover, refinements of the estimator proposed by (17) are possible. For example, we may derive an alternate closed form estimator if we treat N in the CV formula as an unknown and solve the resulting estimating equations. An examination of the properties of this estimator is beyond the scope of the present work and shall be conducted elsewhere.

APPENDICES

A.1. Behaviour of u_j

The asymptotic distribution of the estimators depends on that of $M_j^* - \hat{M}_j^*$ and of $u_j = \sum_{i=1}^{N} I(\sum_{k=1}^{j-1} X_{ik} = 0)I(X_{ij} = 1)$. Note that $N^{-1}u_j \xrightarrow{a.s.} E(p(1-p)^{j-1})$. Let $\mathcal{F}_{j-1}$ denote the σ-field generate by the capture histories up to occasion j. Then it is easily seen that given $p_1, \ldots, p_N$,

$$E\big(u_j|\mathcal{F}_{j-1}\big) = E\left\{\sum_{i=1}^{N} I \sum_{k=1}^{j-1} X_{ik} = 0, X_{ij} = 1)|\mathcal{F}_{j-1}\right\}$$

$$= \sum_{i=1}^{N} I\left(\sum_{k=1}^{j-1} X_{ik} = 0\right) p_i = (N - M_j^*)\ \bar{p} = N(1 - C_{j-1})\bar{p}$$

Let $Y_{ij} = I(\sum_{k=1}^{j-1} X_{ik} = 0)(I(X_{ij} = 1) - p_i)$, $Y_{(j)} = u_j - E(u_j|\mathcal{F}_{j-1}) = \sum_{i=1}^{N} Y_{ij}$, $Y = (Y_{(1)}, \ldots, Y_{(t)})^T$, and note that given p_i, $E(Y_{ij}|\mathcal{F}_{j-1}) = 0$, $E(Y_{ij}^2|\mathcal{F}_{j-1}) = p_i(1-p_i)I(\sum_{k=1}^{j-1} X_{ik} = 0)$ so that $E(Y_{ij}^2) = p_i(1-p_i)^j$, and for fixed j, the Y_{ij} are independent $i = 1, \ldots, n$. Moreover, unconditionally, for fixed j, the Y_{ij} are independently and identically distributed. Thus the central limit theorem may be applied to $N^{-1/2}\sum_{i=1}^{N} Y_{ij}$. Furthermore, for fixed i, it is easily seen that the Y_{ij} are martingale differences and are hence uncorrelated. Hence, the joint distribution of $N^{-1/2}Y$ is multivariate normal with mean vector 0 and the covariance matrix Λ is diagonal with kth diagonal element $E(p(1-p)^k)$.

A.2. Asymptotic Behaviour of the Estimated Sample Coverage

Note that (12) derived the estimator of the sample coverage from an examination of the moments of various quantities. Here we show that as the population size increases, the sample coverage estimator and the sample coverage converge a.s. to the same limit and are thus asymptotically equivalent. Note that $f_{1j} = \sum_{i=1}^{N} I(\sum_{k=1}^{j} X_{ik} = 1)$ so that $E(f_{1j}) = \sum_{i=1}^{N} jE(p_i(1-p_i)^{j-1}) = NjE(p(1-p)^{j-1})$. Hence, as the individuals are assumed to behave independently, the law of large numbers implies that $N^{-1}\times f_{1j} \xrightarrow{a.s.} jE(p(1-p)^{j-1})$. As $\sum_{k=1}^{j} n_k = \sum_{i=1}^{N}\sum_{k=1}^{j} I(X_{ik} = 1)$, once again the law of large numbers implies that $N^{-1}\sum_{k=1}^{j} n_k \xrightarrow{a.s.} jE(p)$. Thus,

$$\hat{C}_{j-1} \xrightarrow{a.s.} 1 - \frac{E\big(p(1-p)^{j-1}\big)}{E(p)} = \frac{E\big(p(1-(1-p)^{j-1})\big)}{E(p)} = \frac{E\Big(p\sum_{k=1}^{j-1} I(X_{ik} > 0)\Big)}{E(p)} = \widetilde{C}_{j-1}$$

but $N^{-1}\sum_{i=1}^{N} p_i \xrightarrow{a.s.} E(p)$ and $N^{-1}\sum_{i=1}^{N} p_i I(\sum_{k=1}^{j-1} X_{ik} > 0) \xrightarrow{a.s.} E(p_i \times (\sum_{k=1}^{j-1} I(X_{ik} > 0))$. Thus, unconditionally, $\hat{C}_{j-1}$ and C_{j-1} are asymptotically equivalent.

A.3. Estimators with Known M_j^*

Consider

$$\tilde{N} = \frac{\sum_{j=1}^{t}\big(1-\hat{C}_{j-1}\big)^{-1} M_j^*}{\sum_{j=1}^{t}\big(1-\hat{C}_{j-1}\big)^{-1}\big(1-u_j/\sum_i p_i\big)} = B^{-1}\sum_{j=1}^{t}\big(1-\hat{C}_{j-1}\big)^{-1} M_j^*,$$

Downloaded by [National Tsing Hua University] at 06:33 21 June 2016

where we have also replaced $\sum n_k/t$ by $\sum p_i$ as $(Nt)^{-1}\sum_{i=1}^{N} p_i$ and $N^{-1}\sum_{k=1}^{t} n_k$ have the same limit. Consider the denominator B. The crucial term is $u_j/\sum_i p_i$ but this is just $N^{-1}u_j/\bar{p}_t$ which as is shown in A.1 and A.2 converges a.s. to $1-\tilde{C}_{j-1}$ defined in A.2, so that $B \xrightarrow{p} B^* = \sum_{j=1}^{t}(1-\tilde{C}_{j-1})^{-1}\tilde{C}_{j-1}$ as $N \to \infty$. Also let μ_p denote the limit of $\bar{p}$ as $N \to \infty$. Now, $N^{-1}(\tilde{N}-N) = \bar{p}^{-1}B^{-1}\sum_{j=1}^{t}(1-\tilde{C}_{j-1})^{-1}N^{-1}\{u_j-(N-M_j^*)\bar{p}\}$ which is asymptotically the same as $\mu_p^{-1}B^{*-1}\sum_{j=1}^{t}(1-\tilde{C}_{j-1})^{-1}N^{-1}\{u_j-N(1-C_{j-1})\bar{p}\} \xrightarrow{p} 0$. Hence, $\tilde{N}$ is an asymptotically unbiased estimator of N. Thus, the asymptotic bias is due to the estimation of M_j^* by $\hat{M}_j^*$. If we let $\tilde{A} = (\tilde{A}_1, \ldots, \tilde{A}_t)^T$, where $\tilde{A}_j$ is as defined in Theorem 1, then following A.1 we may write $N^{-1/2}(\tilde{N}-N) = N^{-1/2}\tilde{A}Y$ which is asymptotically normal with zero mean and variance $\tilde{A}^T\Lambda\tilde{A}$ which reduces to $\mu_p^{-2}B^{*-2}\sum_{j=1}^{t}(1-\tilde{C}_{j-1})^{-2}E(p(1-p)^j)$.

A.4. Estimating the CV

The estimator $\hat{M}_j^*$ involves an estimate of the coefficient of variation. We have estimated the CV, $\gamma^2 = N^{-1}\sum_{i=1}^{N}(p_i-\bar{p})^2/\bar{p}^2$, by

$$\hat{\gamma}^2 = \max\left\{\frac{\hat{N}_0 t\sum_k k(k-1)f_{k,t}}{(t-1)\left(\sum_k n_k\right)^2} - 1, 0\right\}$$

where $\hat{N}_0 = M_{t+1}/\hat{C}_t$. First consider

$$\hat{\gamma}^{*2} = \max\left\{\frac{Nt\sum_k k(k-1)f_{k,t}}{(t-1)\left(\sum_k n_k\right)^2} - 1, 0\right\}$$

Now,

$$\frac{Nt\sum_k k(k-1)f_k}{(t-1)\left(\sum_k n_k\right)^2} - 1 = \frac{\sum_k k(k-1)N^{-1}f_{k,t}}{t(t-1)\bar{p}^2} - 1. \tag{A.4.1}$$

However,

$$\begin{aligned}\sum_k k(k-1)N^{-1}f_{k,t} &\xrightarrow{p} \sum_k k(k-1)E\left(\binom{t}{k}p^k(1-p)^{t-k}\right)\\ &= E\left(\sum_k k(k-1)N^{-1}f_{k,t}\right) = N^{-1}E\left(t(t-1)\sum_{i=1}^{N}p_i^2\right)\\ &= t(t-1)E(p^2)\end{aligned}$$

Downloaded by [National Tsing Hua University] at 06:33 21 June 2016

using (3.15) of (12). Thus, (A.4.1) reduces to $(E(p^2) - \mu_p^2)/\mu_p^2 = \sigma_p^2/\mu_p^2$. Next note that

$$N^{-1}\hat{N}_0 = \frac{N^{-1}M_{t+1}}{\hat{C}_t} \xrightarrow{p} \frac{E\big(1-(1-p)^t\big)\mu_p}{E\big(p(1-(1-p)^t)\big)} = a^*$$

When the capture probabilities have a beta (α, β) distribution this latter quantity is equal to a given in Theorem 3. Let $\gamma^{*2} = a^*\sigma_p^2/\mu_p^2$ denote the limit in probability of $\hat{\gamma}^2$.

A.5. Estimating M_j^*

The estimating equations involve an approximation to estimation of $NC_{k-1} = M_k^*$ which we now examine. Let $\underset{\sim}{p} = (p_1, \ldots, p_N)^T$. Now,

$$M_k^* = \frac{\sum_{i=1}^N p_i I\big(\sum_{j=1}^{k-1} X_{ij} > 0\big)}{\bar{p}} = \sum_{i=1}^N \frac{p_i}{\bar{p}} I\left(\sum_{j=1}^{k-1} X_{ij} > 0\right), \text{ and}$$

$$M_k = \sum_{i=1}^N I\left(\sum_{j=1}^{k-1} X_{ij} > 0\right)$$

with $M_1 = M_1^* = 0$ so that

$$\begin{aligned} E\big(M_k^* - M_k|\underset{\sim}{p}\big) &= \bar{p}^{-1}\sum_{i=1}^N (p_i - \bar{p})\big(1-(1-p_i)^{k-1}\big) \\ &= -\bar{p}^{-1}\sum_{i=1}^N (p_i - \bar{p})(1-p_i)^{k-1} \end{aligned}$$

and hence as in (3.12) of p. 205 of (12), for a remainder term R, $E(M_k^* - M_k|\underset{\sim}{p}) = f_{1,k-1}\gamma^2 + R$. This motivated the approximation $\hat{M}_k^* = M_k + \widetilde{f}_{1,k-1}\hat{\gamma}^2$ of (17). To determine the bias arising from the approximation note that for $k \geq 2$,

$$\begin{aligned} \delta_k &= \lim_{N\to\infty} N^{-1}\big(M_k^* - \hat{M}_k^*\big) = \lim_{N\to\infty} N^{-1}\big(M_k^* - M_k - f_{1,k-1}\gamma^{*2}\big) \\ &= \lim_{N\to\infty} N^{-1}\sum_{i=1}^N \left\{ I\left(\sum_{j=1}^{k-1} X_{ij} > 0\right)\left(\frac{p_i}{\mu_p} - 1\right) - \gamma^{*2} I\left(\sum_{j=1}^{k-1} X_{ij} = 1\right)\right\} \\ &= E\Big\{(1-p)^{k-2}\big((1-p/\mu_p)(1-p) - \gamma^{*2}(k-1)p\big)\Big\}. \end{aligned}$$

Downloaded by [National Tsing Hua University] at 06:33 21 June 2016

To determine the asymptotic distribution of the approximation, write

$$\begin{aligned} M_j^* - \hat{M}_j^* &= \sum_{i=1}^{N} \left\{ I\left(\sum_{k=1}^{j-1} X_{ik} > 0 \right) \left(\frac{p_i}{\mu_p} - 1 \right) - \gamma^{*2} I\left(\sum_{k=1}^{j-1} X_{ik} = 1 \right) \right\} \\ &= \sum_{i=1}^{N} Z_{ij} \end{aligned}$$

Let $\bar{A} = (A_2, \ldots, A_t)^T$, where A_j is as defined in Theorem 1, let $V_{(j)} = \hat{M}_j^* - M_j^* + N\delta_j = \sum_{i=1}^{N}(Z_{ij} + \delta_j)$, and let $V = (V_{(2)}, \ldots, V_{(t)})^T$. Then V has mean vector zero, and unconditionally is the sum of i.i.d. vectors $V_i = (V_{i2}, \ldots, V_{it})^T$ that each have a covariance matrix that we denote by Γ. Let $\delta = (\delta_2, \ldots, \delta_t)^T$. Now, $N^{-1/2}(\hat{N} - \tilde{N}) = N^{-1/2}\bar{A}^T V - N^{1/2}\bar{A}^T\delta$. Hence, $N^{-1/2}(\hat{N} - \tilde{N} + N\bar{A}^T\delta) \xrightarrow{d} N(0, \bar{A}^T\Gamma\bar{A})$.

A.6. The Covariance Matrix of W_i

Recall $W = (Y^T, V^T)^T$. Write

$$\Sigma = \begin{pmatrix} \Lambda & \Omega \\ \Omega^T & \Gamma \end{pmatrix},$$

where we have already determined Λ in A.1. In Lemmas 1–2 we consider the covariance matrix Γ of the V_i, and in Lemma 3 the covariance matrix Ω of Y and V, which allows the construction of the theoretical covariance matrix Σ of the W_i.

Lemma 1.

$$\begin{aligned} E(V_{ij}^2) &= E\left(\{1 - (1-p)^{j-1}\} \left(\frac{p_i}{\mu_p} - 1 \right)^2 \right) \\ &\quad + \gamma^{*4}(j-1)E\left(p_i(1-p_i)^{j-2}\right) + \delta_j^2 \\ &\quad + 2\delta_j E\left(\left(1 - (1-p_i)^{j-1}\right) \left(\frac{p_i}{\mu_p} - 1 \right) \right) \\ &\quad - 2\gamma^{*2}\delta_j(j-1)E\left(p_i(1-p_i)^{j-2}\right) \\ &\quad - 2\gamma^{*2}(j-1)E\left(p_i(1-p_i)^{j-2} \left(\frac{p_i}{\mu_p} - 1 \right) \right). \end{aligned}$$

Downloaded by [National Tsing Hua University] at 06:33 21 June 2016

Proof. Note that

$$V_{ij}^2 = I\left(\sum_{k=1}^{j-1} X_{ik} > 0\right)\left(\frac{p_i}{\mu_p} - 1\right)^2 + \gamma^{*4} I\left(\sum_{k=1}^{j-1} X_{ik} = 1\right) + \delta_j^2$$
$$+ 2I\left(\sum_{k=1}^{j-1} X_{ik} > 0\right)\left(\frac{p_i}{\mu_p} - 1\right)^2 \delta_j - 2\gamma^{*2}\delta_j I\left(\sum_{k=1}^{j-1} X_{ik} = 1\right)$$
$$- 2\gamma^{*2} I\left(\sum_{k=1}^{j-1} X_{ik} = 1\right)\left(\frac{p_i}{\mu_p} - 1\right).$$

and taking expectations yields the lemma.

Lemma 2. *For* $k < j$,

$$E(V_{ij}V_{ik}) = E\left(\left\{1-(1-p_i)^{k-1}\right\}\left(\frac{p_i}{\mu_p}-1\right)^2\right) - \gamma^{*2}(k-1)E$$
$$\times\left(p_i(1-p_i)^{k-2}\left(\frac{p_i}{\mu_p}-1\right)\right) + \delta_k E\left(\left\{1-(1-p)^{j-1}\right\}\left(\frac{p_i}{\mu_p}-1\right)\right)$$
$$-\gamma^{*2}(k-1)E\left(p_i(1-p_i)^{k-2}(1-p)^{j-k+1}\left(\frac{p_i}{\mu_p}-1\right)\right)$$
$$+\gamma^{*4}(k-1)E\left(p(1-p)^{k-2}(1-p)^{j-k+1}\right)$$
$$+\delta_k\gamma^{*2}(j-1)E\left(p(1-p)^{j-2}\right)$$
$$+\delta_j E\left(\left\{1-(1-p)^{k-1}\right\}\left(\frac{p_i}{\mu_p}-1\right)\right)$$
$$-\delta_j\gamma^{*2}(k-1)E(p(1-p)^{k-2}) + \delta_j\delta_k.$$

Proof.

$$V_{ij}V_{ik} = I\left(\sum_{l=1}^{k-1} X_{il} > 0\right)\left(\frac{p_i}{\mu_p} - 1\right)^2 - \gamma^{*2} I\left(\sum_{l=1}^{k-1} X_{il} = 1\right)\left(\frac{p_i}{\mu_p} - 1\right)$$
$$+ \delta_k I\left(\sum_{l=1}^{j-1} X_{il} > 0\right)\left(\frac{p_i}{\mu_p} - 1\right)$$
$$- \gamma^{*2} I\left(\sum_{l=1}^{k-1} X_{il} > 0\right) I\left(\sum_{l=1}^{j-1} X_{il} = 1\right)\left(\frac{p_i}{\mu_p} - 1\right)$$

Downloaded by [National Tsing Hua University] at 06:33 21 June 2016

$$+ \gamma^{*4} I\left(\sum_{l=1}^{j-1} X_{il} = 1\right) I\left(\sum_{l=1}^{k-1} X_{il} = 1\right) + \delta_k \gamma^{*2} I\left(\sum_{l=1}^{j-1} X_{il} = 1\right)$$

$$+ \delta_j \left(\frac{p_i}{\mu_p} - 1\right) I\left(\sum_{l=1}^{k-1} X_{il} > 0\right) - \delta_j \gamma^{*2} I\left(\sum_{l=1}^{k-1} X_{il} = 1\right) + \delta_j \delta_k$$

and taking expectations yields the lemma.

Lemmas 1 and 2 now yield Γ. Next we need to consider the covariance between the Y_i and V_i. Note that $V_{ij} Y_{ij} = \delta_j Y_{ij}$ so that $E(Y_{ij} V_{ij}) = 0$ and similarly $E(Y_{ik} V_{ij}) = 0$ for $k > j$. However, if $k < j$ we have

Lemma 3. *If* $k < j$ *then*,

$$E\big(Y_{ik} V_{ij}\big) = E\left((1-p)^{k-1} p \left(\frac{p}{\mu_p} - 1\right)\right) - E\left(p(1-p)^{j\gamma^{*2}}\right)$$

$$+ \delta_j E\left((1-p)^{k-1} p\right) - E\left(p(1-p)^{k-1}\left(1 - (1-p)^{j-k}\right)\right.$$

$$\left.\times \left(\frac{p}{\mu_p} - 1\right)\right) + \gamma^{*2} E\big(p^2 (1-p)^j (j-k)\big)$$

$$- \delta_j E\left(p(1-p)^{k-1}\right).$$

Proof. For $k < j$,

$$Y_{ik} V_{ij} = I\left(\sum_{l=1}^{k-1} X_{il} = 0\right) I(X_{ik} = 1) \left(\frac{p_i}{\mu_p} - 1\right)$$

$$- I\left(\sum_{l=1}^{k-1} X_{il} = 0\right) I(X_{ik} = 1) I\left(\sum_{l=k+1}^{j-1} X_{ik} = 0\right) \gamma^{*2}$$

$$+ \delta_j I\left(\sum_{l=1}^{k-1} X_{il} = 0\right) I(X_{ik} = 1)$$

$$- p_i I\left(\sum_{l=1}^{k-1} X_{il} = 0\right) I\left(\sum_{l=k}^{j-1} X_{il} > 0\right) \left(\frac{p_i}{\mu_p} - 1\right)$$

$$+ p_i \gamma^{*2} I\left(\sum_{l=1}^{k-1} X_{il} = 0\right) I\left(\sum_{l=k}^{j-1} X_{il} = 1\right) - p_i \delta_j I\left(\sum_{l=1}^{k-1} X_{il} = 0\right)$$

and taking expectations yields the lemma.

Downloaded by [National Tsing Hua University] at 06:33 21 June 2016

A.7. The Covariance Matrix of W_i for the Captured Individuals

Now,

$$W = \sum_{i=1}^{N} W_i = \sum_{i=1}^{N} W_i \xi_i + \sum_{i=1}^{N} W_i (1 - \xi_i)$$

The bootstrap conditional on the n captured individuals allows us to approximate the variance of $\sum_{i=1}^{N} W_i \xi_i$. We wish to examine how well this approximates the true variance. Let $\tilde{\Sigma}$ denote the covariance matrix of $W_i \xi_i$. Then we may write

$$\tilde{\Sigma} = \begin{pmatrix} \tilde{\Lambda} & \tilde{\Omega} \\ \tilde{\Omega}^T & \tilde{\Gamma} \end{pmatrix} \text{ and } \Delta = \Sigma - \tilde{\Sigma} = \begin{pmatrix} \Delta_{11} & -\Delta_{12} \\ -\Delta_{21} & \Delta_{22} \end{pmatrix}.$$

Lemma 4. *Let Δ_{11} be a $t \times t$ matrix with constant elements $E(p^2(1-p)^t) + E^2(p(1-p)^t)$. Then $\tilde{\Lambda} = \mathrm{Cov}(Y_i \xi_i) = \Lambda - \Delta_{11}$.*

Proof. Recall that $Y_{ij} = I(\sum_{k=1}^{j-1} X_{ik} = 0)(I(X_{ij} = 1) - p_i)$. First,

$$\begin{aligned} E(Y_{ij}\xi_i|p_i) &= E\left[I\left(\sum_{k=1}^{j-1} X_{ik} = 0\right) \{I(X_{ij}=1) - p_i\} I\left(\sum_{k=1}^{t} X_{ik} > 0\right) \Big| p_i \right] \\ &= E\left[I\left(\sum_{k=1}^{j-1} X_{ik} = 0\right) I(X_{ij} = 1) \right. \\ &\qquad \left. - p_i I\left(\sum_{k=1}^{j-1} X_{ik} = 0\right) I\left(\sum_{k=j}^{t} X_{ik} > 0\right) \Big| p_i \right] \\ &= (1-p)^{j-1} p_i - p_i (1-p_i)^{j-1} \left(1 - (1-p_i)^{t-(j-1)}\right) = p_i (1-p_i)^t. \end{aligned}$$

and as

$$\begin{aligned} E\left(Y_{ij}^2 \xi_i | p_i\right) &= E\left\{ I\left(\sum_{k=1}^{j-1} X_{ik} = 0\right) I\left(\sum_{k=1}^{t} X_{ik} > 0\right) \left(I(X_{ij} = 1) - p_i\right)^2 \Big| p_i \right\} \\ &= E\left\{ I\left(\sum_{k=1}^{j-1} X_{ik} = 0\right) I\left(\sum_{k=1}^{t} X_{ik} > 0\right) \right. \\ &\qquad \left. \times \left(I(X_{ij} = 1) - 2I(X_{ij} = 1)p_i + p_i^2\right) \Big| p_i \right\} \\ &= (1-p_i)^{j-1} p_i (1 - 2p_i) + (1-p_i)^{j-1} \left(1 - (1-p_i)^{t-(j-1)}\right) p_i^2 \\ &= p_i (1-p_i)^{j-1} - p_i^2 (1-p_i)^{j-1} - p_i^2 (1-p_i)^t \end{aligned}$$

Downloaded by [National Tsing Hua University] at 06:33 21 June 2016

we see that $V(Y_{ij}\xi_i|p_i)$ is just

$$E\Big\{\big(Y_{ij}\xi_i - p_i(1-p_i)^t\big)^2\big|p_i\Big\} = p_i(1-p_i)^{j-1} - p_i^2(1-p_i)^{j-1} \\ - p_i^2(1-p_i)^t - p_i^2(1-p_i)^{2t}$$

and hence, $V(Y_{ij}\xi_i) = E((1-p)^j p) - E(p^2(1-p)^t) - E^2(p(1-p)^t) = \widetilde{\lambda}_{jj}$. However, for $k < j$, we lose the martingale property and

$$Y_{ij}Y_{ik}\xi_i = -p_i I\left(\sum_{l=1}^{j-1} X_{il} = 0\right) I\big(X_{ij} = 1\big) \\ + p_i^2 I\left(\sum_{l=1}^{j-1} X_{il} = 0\right) I\left(\sum_{l=j}^{t} X_{ij} > 0\right)$$

so that $E(Y_{ij}Y_{ik}\xi_i|\, p_i) = -p_i^2(1-p_i)^t$ and hence $\mathrm{Cov}(Y_{ij}\xi_i, Y_{ik}\xi_i) = -E(p^2(1-p)^t) - E^2(p(1-p)^t) = \widetilde{\lambda}_{jk}$.

Lemma 5.

$$\tilde{\Gamma} = \mathrm{Cov}(V_i\xi_i) = \Gamma - \delta\delta^T E\big((1-p)^t\big)\big\{1 + E\big((1-p)^t\big)\big\} = \Gamma - \Delta_{22}.$$

Proof. Recall

$$V_{ij} = I\left(\sum_{k=1}^{j-1} X_{ik} > 0\right)\left(\frac{p_i}{\mu_p} - 1\right) - \gamma^{*2} I\left(\sum_{k=1}^{j-1} X_{ik} = 1\right) + \delta_j$$

Then $V_{ij}\xi_i = V_{ij} - \delta_j I(\sum_{k=1}^t X_{ik} = 0)$, so that, $E(V_{ij}\xi_i) = -\delta_j E((1-p)^t)$, $V_i\xi_i = V_i - \delta(1-\xi_i)$ and $V_i(1-\xi_i) = \delta(1-\xi_i)$. Thus, $V_iV_i^T\xi_i = V_iV_i^T - V_i\delta^T(1-\xi_i) - \delta V_i^T(1-\xi_i) + \delta\delta^T(1-\xi_i)$ and hence $E(V_iV_i^T\xi_i) = E(V_iV_i^T) - \delta\delta^T E((1-p)^t)$. Subtraction of $E(V_i\xi_i)$ from the l.h.s. yields the lemma.

Lemma 6.

$$\tilde{\Omega}^T = \mathrm{Cov}(V_i\xi_i, Y_i\xi_i) = \Omega^T + \delta\underset{\sim}{1}^T E\big(p(1-p)^t\big)\big\{1 + E\big((1-p)^t\big)\big\} = \Omega^T + \Delta_{21}.$$

Proof. Note that $E(Y_{ij}(1-\xi_i)) = -E(p_i I(\sum_{k=1}^t X_{ik} = 0))$ so that

$$E\big(V_iY_i^T\xi_i\big) = E\big(V_iY_i^T\big) - \delta E\big(Y_i^T(1-\xi_i)\big) = E(V_iY_i^T) + \delta\underset{\sim}{1}^T E(p(1-p)^t)$$

Subtracting the product of appropriate mean vectors (i.e. $-\delta\underset{\sim}{1}^T E(p(1-p))\times E(1-p)^t)$ yields the lemma.

Downloaded by [National Tsing Hua University] at 06:33 21 June 2016

ACKNOWLEDGMENT

This work was initiated while the first author was visiting the Institute of Statistics at the National Tsing Hua University. The authors are grateful to an anonymous referee whose comments greatly improved the presentation of the results.

Downloaded by [National Tsing Hua University] at 06:33 21 June 2016

REFERENCES

1. Seber, G.A.F. *The Estimation of Animal Abundance and Related Parameters*, 2nd Ed.; Griffin: London, 1982.
2. Seber, G.A.F. A Review of Estimating Animal Abundance. Biometrics **1986**, *42*, 267–292.
3. Seber, G.A.F. A Review of Estimating Animal Abundance II. International Statistical Review **1992**, *60*, 129–166.
4. Schwarz, C.J.; Seber, G.A.F. A Review of Estimating Animal Abundance III. 1999. To appear in Statistical Science.
5. Carothers, A.D. Capture–Recapture Methods Applied to a Population with Known Parameters. Journal of Animal Ecology **1973**, *42*, 125–146.
6. Otis, D.L.; Burnham, K.P.; White, G.C.; Anderson, D.R. Statistical Inference from Capture Data on Closed Animal Populations. Wildlife Monographs **1978**, *62*, 1–135.
7. Burnham, K.P.; Overton, W.S. Robust Estimation of Population Size When Capture Probabilities Vary Among Animals. Ecology **1979**, *60*, 927–936.
8. White, G.C.; Anderson, D.R.; Burnham, K.P.; Otis et al. *Capture–Recapture and Removal Methods for Sampling Closed Populations*. Publication LA-8787-NERP. Loa Alamos National Laboratory: Los Alamos, New Mexico, 1982.
9. Burnham, K.P.; Overton, W.S. Estimation of the Size of a Closed Population When Capture Probabilities Vary Among Animals. Biometrika **1978**, *65*, 625–633.
10. Pollock, K.H.; Otto, M.C. Robust Estimation of Population Size in Closed Animal Populations from Capture–Recapture Experiments. Biometrics **1983**, *39*, 1035–1049.
11. Smith, E.P.; van Belle, G. Nonparametric Estimation of Species Richness. Biometrics **1984**, *40*, 119–129.
12. Chao, A.; Lee, S.-M.; Jeng, S.-L. Estimating Population Size for Capture–Recapture Data When Capture Probabilities Vary by Time and Individual Animal. Biometrics **1992**, *48*, 201–216.

13. Lee, S.-M.; Chao, A. Estimating Population Size Via Sample Coverage for Closed Capture–Recapture Models. Biometrics **1994**, *50*, 88–97.
14. Darroch, J.N.; Fienberg, S.E.; Glonek, G.F.V.; Junker, B.W. A Three-Sample Multiple-Recapture Approach to Census Population Estimation with Heterogeneous Catchability. Journal of the American Statistical Association **1993**, *88*, 1137–1148.
15. Agresti, A. Simple Capture–Recapture Models Permitting Unequal Catchability and Variable Sampling Effort. Biometrics **1994**, *50*, 494–500.
16. Norris, J.L.; Pollock, K.H. Nonparametric MLE Under Two Closed Capture–Recapture Models with Heterogeneity. Biometrics **1996**, *52*, 639–649.
17. Chao, A.; Yip, P.; Lee, S.-M.; Chu, W. Population Size Estimation Based on Estimating Functions for Closed Capture–Recapture Models. Journal of Statistical Planning and Inference, **2001**, *92*, 213–232.
18. Liang, K.-Y.; Zeger, S.L. Inference Based on Estimating Functions in the Presence of Nuisance Parameters (with Discussion). Statistical Science **1995**, *10*, 158–199.
19. Lloyd, C.J.; Yip, P. A Unification of Inference for Capture–Recapture Studies Through Martingale Estimating Functions. In *Estimating Equations*; Godambe, V.P., Ed.; Clarendon Press: Oxford, 1991; pp. 65–88.
20. Wang, Y.-G. Estimating Equations for Removal Data Analysis. Biometrics **1999**, *55*, 1263–1268.
21. Chao, A.; Chu, W.; Hsu, C.-H. Capture–Recapture When Time and Behavioral Response Affect Capture Probabilities. Biometrics **2000**, *56*, 427–433.
22. Good, I.J. The Population Frequencies of Species and the Estimation of Population Parameters. Biometrika **1953**, *40*, 237–264.

Downloaded by [National Tsing Hua University] at 06:33 21 June 2016

Downloaded by [National Tsing Hua University] at 06:33 21 June 2016