Statistica Sinica **13**(2003), 673-689

THE ESTIMATION OF THE SIZE OF AN OPEN POPULATION USING LOCAL ESTIMATING EQUATIONS

Hsin-Chou Yang and Richard Huggins

National Tsing Hua University and La Trobe University

Abstract: Local polynomial models for capture-recapture experiments on open populations with frequent capture occasions and heterogeneous capture probabilities are proposed. A one-step bootstrap procedure is proposed to determine optimal bandwidths. It is shown in simulations that for heterogeneous populations the proposed procedure performs better than the kernel estimator of Huggins and Yip (1999) that was developed for homogeneous populations and the locally constant estimator of Huggins, Yang, Chao and Yip (2003).

Key words and phrases: Capture-recapture, kernel function, martingale estimating function, open population, sample coverage.

1. Introduction

Traditional capture-recapture experiments have involved captures from populations on multiple occasions that are close enough together that the population could be assumed to be closed. Open population models, such as the Jolly-Seber approach, are based on relatively sparse capture occasions and involve modeling the population dynamics. However, the assumptions of the Jolly-Seber model may be easily violated. In particular, there may be heterogeneous capture probabilities. Over the twenty or more years since Burnham and Overton (1978), enormous effort has been spent on devising methods that allow individual heterogeneity of capture probabilities in closed capture-recapture experiments. See Chao and Huggins (2001) for a review of these methods. The incorporation of heterogeneity in open population models is less well studied. Hwang and Chao (1995) have previously applied sample coverage methods to the Jolly-Seber model, Pledger and Efford (1998) proposed correction to the bias arising from heterogeneity in the Jolly-Seber model and Huggins, Yang, Chao and Yip (2003) introduced kernel smoothing in a limited setting.

The natural extension of closed population methods to open populations is to consider experiments conducted over a long time period, but where the capture occasions are close together. Huggins and Yip (1999) and Huggins et al. (2003) have previously used kernel smoothing to extend the homogeneous model, where capture probabilities vary by occasion but not by individual, and the heterogeneous model for closed populations to open populations with frequent capture occasions. Their work used locally constant models and hence supposed that the population size was approximately constant over short time intervals. The estimator of Huggins and Yip (1999) was shown to have advantages over the traditional Jolly-Seber method when individuals could leave and return to the population. The methods of Huggins et al. (2003) performed better than the Huggins and Yip (1999) estimator when the capture probabilities were heterogeneous. The availability of capture data with capture occasions at frequent intervals encourages further development of the open population methodologies pioneered in Huggins and Yip (1999) and Huggins et al (2003).

The main assumption on the population size in the approaches of Huggins and Yip (1999) and Huggins et al. (2003) is that the population size is locally constant. That is, in a neighbourhood of a given capture occasion, the population size is constant. This assumption is reasonable if the capture occasions are close together but in practice this is rarely so. Moreover, it is known in other settings that violation of this assumption may result in serious bias. In non-parametric kernel estimation, local polynomial models have been shown to have good theoretical properties and to have advantages in practical applications (Fan and Gijbels (1996, p.60)). This motivates us to extend the approach of Huggins et al. (2003) so that there is no need to assume the population is locally closed. The method again utilizes the optimal martingale estimating equations of Chao, Yip, Lee and Chu (2001) and follows the general local estimating equation methodology of Carroll, Ruppert and Welsh (1998). A difficulty encountered in Huggins and Yip (1999) and Huggins et al. (2003) was the determination of the bandwidth. Here we propose the use of the bootstrap method to estimate the bias (Efron and Tibshirani, Chap. 10) and variance (e.g., Shao and Tu, pp.228-229) and hence the mean squared error. Using the criterion of minimum mean squared error, we can then select an optimal bandwidth for each time point.

A feature of the model is that the population size is regarded as a fixed deterministic function, and as such is a parameter to be estimated. However, some assumptions on how individuals are removed from the population are required, so that on a given capture occasion the nuisance parameter consisting of the number of marked individuals in the population may be estimated. This is necessary as the number of individuals marked and released will, due to removals, overestimate the number of marked individuals remaining in the population.

Our approach is based on sample coverage. The concept of sample coverage was originally proposed by I. J. Good (1953) and was defined as the proportion of capture probabilities of the captured individuals. Chao, Lee and Jeng (1992) and Lee and Chao (1994) used sample coverage and the coefficient of variation of the capture probabilities to quantify recapture information and sample dependencies, and subsequently estimated the size of a closed population. Chao et al. (2001) developed martingale estimating equations based on sample coverage. Hwang and Chao (1995) extended the sample coverage approach to the open population model. The advantage of the sample coverage approach is that whilst the number of the unobserved individuals is difficult to estimate, the sample coverage can be well-estimated. The population size may then be estimated by exploiting the relationship between population size and sample coverage.

In Section 2 we derive the estimating equations when various nuisance parameters (the sample coverage and coefficient of variation) are known. In Section 3 estimation of the sample coverage and coefficient of variation of the capture probabilities is discussed. This yields a preliminary step for estimating the number of "marked" individuals. In Section 4 we give an estimate of the number of "marked" individuals that is crucial to the sample coverage approach. In Section 5 we propose a bootstrap bandwidth selector. A simulation study is conducted in Section 6. In Section 7 the method is applied to data. Our notational convention is to take a superscript LC or LL depending on "Local Constant" or "Local Linear" and a subscript H or $\bar{H}$ for the homogeneous and heterogeneous models, respectively.

1.1. Assumptions

As in Huggins and Yip (1999) and Huggins et al. (2003), the population size is regarded as an unknown deterministic function N_t of t of the form $[N\lambda_t]$ where λ_t is some continuous function and $[x]$ denotes the closest integer to x. Suppose there are a total of τ capture occasions, $0 < t_1 < \cdots < t_\tau < T$. For simplicity it is supposed that the occasions are equally spaced but this is not necessary. Let $X_{ij} = I($ the ith individual is caught on occasion $j)$, where $I(A)$ denotes the indicator function for the event A, and let $\mathcal{F}_j$ be the capture history up to the jth capture occasion. Thus $P\{X_{ij} = 1|\mathcal{F}_{j-1}\} = p_i$. The basic assumption required in this paper is : the individual capture probabilities p_i, $i = 1, \ldots, N_t$, are supposed to be i.i.d. random variables from a distribution $F_t(p)$. This extends the classical heterogeneity model. Denote their mean at time t by $\bar{p}(t) = N_t^{-1}\sum_{i=1}^{N_t} p_i$ and their coefficient of variation at time t by $\gamma_t = \{\sum_{i=1}^{N_t}[p_i - \bar{p}(t)]^2/[N_t\bar{p}(t)^2]\}^{1/2}$. We assume throughout that the capture probability of an individual arriving into the population at time t also has distribution $F_t(p)$ and is independent of the individuals already in the population. Although the population size is not random, in order to estimate the number of marked individuals in the population it is necessary to make some assumptions on how individuals are removed from the population. It is assumed that the probability of removal is the same for individuals captured and released on a given capture occasion as for individuals captured and released before that occasion. This assumption was implicit in Huggins and Yip (1999) and was explicitly stated in the locally constant model of Huggins et al. (2003). Finally, to fit a polynomial of degree p it is supposed that the $(p+2)$nd derivative of the function λ_t, $\lambda_t^{(p+2)}$, exists and is continuous as in Fan, Heckman and Wand (1995). This assumption is reasonable in practice as long as there are not large changes in the population size in small intervals. We note that the asymptotic results of Huggins and Chao (2001) may be extended to show convergence of the estimators for fixed t, although in some cases they may be biased.

2. Martingale Estimating Equations

We suppose $N_s \approx \sum_{l=0}^{p} \beta_l (s-t)^l$ for constants $\beta_0, \cdots, \beta_p$. For large N this may be justified by applying a Taylor series expansion to the smooth part $N\lambda_t$ of N_t. As usual, $\beta_0 = N_t$ and if β_0 can be estimated, the estimated population size at time t is $\hat{\beta}_0$. Let $k(t)$ denote the closest time occasion to time t. Let $Q(\cdot)$ denote a kernel function assigning weights to each capture occasion. For a fixed time t and a given bandwidth (window width) h, the window $\mathcal{W}_t$ contains all of the capture occasions in the support of $Q((t-s)/h)$, $0 \leq s \leq T$. We suppose throughout for some K, the window $\mathcal{W}_t = [k(t)-K, k(t)+K]$ includes the $2K+1$ capture occasions $t_{k(t)-K}, \ldots, t_{k(t)+K}$. The weight function $w_j(t)$ of the jth occasion in window $\mathcal{W}_t$ is defined as $w_j(t) = (\sum_{l \in \mathcal{W}_t} Q_l(t,h))^{-1} Q_j(t,h)$, where $Q_l(t,h) = Q((t-t_l)/h)$. Consider a window $\mathcal{W}_t$. Let n_j denote the number of individuals captured on occasion j, $u_j(t)$ denote the number of individuals that were captured for the first time in $\mathcal{W}_t$ on occasion j, and $m_j(t)$ denote the number of individuals that were recaptured on occasion j after being previously captured in $\mathcal{W}_t$. We may write

$$u_j(t) = \sum_{i=1}^{N_{t_j}} I\Big(\sum_{l=k(t)-K}^{j-1} X_{il} = 0, X_{ij} = 1 \Big)$$

$$\text{and} \qquad m_j(t) = \sum_{i=1}^{N_{t_j}} I\Big(\sum_{l=k(t)-K}^{j-1} X_{il} \geq 1, X_{ij} = 1 \Big).$$

Let $\boldsymbol{\theta}^T = [\beta_0, \ldots, \beta_p, \bar{p}(t)]$ be the vector of model parameters, and $g_j(t) = (g_{j1}(t), g_{j2}(t))^T$, where $g_{j1}(t) = u_j(t) - E\left(u_j(t)|\mathcal{F}_{j-1}\right)$ and $g_{j2}(t) = m_j(t) - E\left(m_j(t)|\mathcal{F}_{j-1}\right)$, be a 2×1 vector of martingale differences (See Appendix A for their explicit forms). The weighted version of the optimal estimating equations of Chao et al. (2001) are: $g(t) = \sum_{j \in \mathcal{W}_t} w_j(t) D_j(t)^T V_j(t)^{-1} g_j(t) = \mathbf{0}$, where $D_j(t) = E(\partial g_j(t)/\partial \boldsymbol{\theta}^T | \mathcal{F}_{j-1})$ is a $2 \times (p+2)$ matrix and $V_j(t) = \mathrm{Cov}\{u_j(t), m_j(t)|$

$\mathcal{F}_{j-1}\}$ is the 2×2 conditional covariance matrix. The resulting estimating functions for $\beta = (\beta_l, l = 0, \ldots, p)^T$, and $\bar{p}(t)$ are shown in Appendix A to be

$$g_{\boldsymbol{\beta}}(t) = -\sum_{j \in \mathcal{W}_t} w_j(t) \frac{\left[u_j(t) - \left\{\sum_{l=0}^{p} \beta_l (t_j - t)^l - M_j^*(t, K)\right\} \bar{p}(t)\right]}{(1 - C_{j-1}(t, K)) \sum_{l=0}^{p} \beta_l (t_j - t)^l} G_p(t_j - t), \quad (1)$$

$$g_{\bar{p}}(t) = -\sum_{j \in \mathcal{W}_t} w_j(t) \left(n_j - \bar{p}(t) \sum_{l=0}^{p} \beta_l (t_j - t)^l \right), \quad (2)$$

where $C_{j-1}(t, K) = \sum_{i=1}^{N_t} p_i I(\sum_{l=k(t)-K}^{j} X_{il} > 0) / \sum_{i=1}^{N_t} p_i$ is the local sample coverage, $G_p(v) = [1, v, v^2, \ldots, v^p]^T$ and $M_j^*(t, K) = N_{t_j} C_{j-1}(t, K)$. From the estimating function (2) we find that

$$\bar{p}(t) = \frac{\sum_{j \in \mathcal{W}_t} w_j(t) n_j}{\sum_{j \in \mathcal{W}_t} w_j(t) \sum_{l=0}^{p} \beta_l (t_j - t)^l},$$

which we may substitute into (1) to obtain the weighted estimating equations

$$\sum_{j \in \mathcal{W}_t} w_j(t) \frac{\left[u_j(t) - \left\{\sum_{l=0}^{p} \beta_l (t_j - t)^l - M_j^*(t, K)\right\} \bar{p}(t)\right]}{[1 - C_{j-1}(t, K)] \sum_{l=0}^{p} \beta_l (t_j - t)^l} G_p(t_j - t) = \mathbf{0}$$

for β. To calculate the estimates, we replace the unknown $M_j^*(t, K)$ and $C_{j-1}(t, K)$ by the estimators $\hat{M}_j^*(t, K)$ and $\hat{C}_{j-1}(t, K)$ below. Let $\hat{N}_{\bar{H}}^{LL}$ denote the local linear estimator arising when $p = 1$, which is found numerically. Details are given in Appendix B. When $p = 0$ we obtain the closed form estimator

$$\hat{N}_{\bar{H}}^{LC}(t) = \frac{\sum_{j \in \mathcal{W}_t} w_j(t) \hat{M}_j^*(t, K) / \left\{1 - \hat{C}_{j-1}(t, K)\right\}}{\sum_{j \in \mathcal{W}_t} w_j(t) [\left(1 - \eta_t^{-1} u_j(t)\right) / \left(1 - \hat{C}_{j-1}(t, K)\right)} \quad (3)$$

with $\eta_t = \sum_{j \in \mathcal{W}_t} w_j(t) n_j$, and $\hat{M}_j^*(t, K)$ and $\hat{C}_{j-1}(t, K)$ being the corresponding estimators of the number of marked individuals and sample coverage under the locally constant model proposed by Huggins et al. (2003). The martingale estimator developed by Huggins and Yip (1999) under a locally constant, homogeneous capture probabilities model is

$$\hat{N}_H^{LC}(t) = \frac{\sum_{j \in \mathcal{W}_t} w_j(t) \bar{M}_j(t, K) n_j}{\sum_{j \in \mathcal{W}_t} w_j(t) \left(n_j - u_j(t)\right)}, \quad (4)$$

where the $\bar{M}_j(t, K)$ are the corresponding estimators of the numbers of marked individuals under a locally constant homogeneous model. The estimator $\hat{N}_H^{LC}(t)$ is a good starting value for iterative searching routines. In Section 6 we compare the performance of this estimator with our proposed estimators.

3. Estimating the Sample Coverage and Coefficient of Variation

The estimation of the number of marked individuals, the sample coverage and the coefficient of variation is critical to our approach. The estimators of $M_j^*(t,K)$ and $C_{j-1}(t,K)$ are different for $p=0$ and $p=1$. For $p=0$ refer to Huggins et al. (2003). For $p=1$, following Hwang and Chao (1995), $E[C_k(t,K)] \approx E[m_k(t)]/E(n_k)$, $k=k(t)-K,\ldots,k(t)+K-1$. Our proposed estimator of the local sample coverage is:

$$\hat{C}_k(t,K) = \frac{\sum_{j\in\mathcal{W}_t} w_j(t_k)m_j(t)}{\sum_{j\in\mathcal{W}_t} w_j(t_k)n_j}.$$

Due to the existence of the heterogeneity among individuals, Chao et al. (1992) employed the CV to quantify the magnitude of variation among individuals. They proposed an estimator based on capture frequencies to estimate the population size under closed population. Hwang and Chao (1995) extended the concept and proposed an estimator of the CV, $\hat{\gamma}(t,K)$, for an open population that we use, but do not formally define here. The reader is referred to Hwang and Chao (1995) for details.

4. Estimating the Number of Marked Individuals

The martingale estimating equation involves the unknown terms $M_j^*(t,K)$ and its estimation is the crucial step in the sample coverage approach. Let $M_j(t,K)$ denote the number of individuals that have been captured in the interval $k(t)-K,\ldots,j-1$, and $\tilde{M}_j(t,K)$ denote the number of individuals captured in $k(t)-K,\ldots,j-1$ that are still in the population at occasion j. Furthermore, let $M_j^*(t,K) = N_{t_j}C_{j-1}(t,K)$. In an open population, we need to estimate $\tilde{M}_j(t,K)$. A naive estimator is $M_j(t,K)$. However, some individuals marked in a window $\mathcal{W}_t$ may permanently leave the population after a specific time point $t_j \in \mathcal{W}_t$. Thus the estimator $M_l(t,K)$ overestimates the actual number of marked individuals $\tilde{M}_l(t,K)$, $j<l$. In order to adjust for the bias, Huggins and Yip (1999) and Huggins et al. (2003) applied kernel smoothing method to the estimating equation $E\{z_j(t)n_j - r_j(t)[\tilde{M}_j(t,K)-m_j(t)]|\tilde{M}_j(t,K),m_j(t),n_j\}=0$ to derive the smooth estimator

$$\bar{M}_j(t,K) = \frac{\sum_{l\in\mathcal{W}_t} w_l(t_j)[z_l(t)n_l + r_l(t)m_l(t)]}{\sum_{l\in\mathcal{W}_t} w_l(t_j)r_l(t)}, \tag{5}$$

where $r_l(t)$ denotes the number of the n_l individuals captured and released on occasion l and recaptured by occasion $k(t)+K$, and $z_l(t)$ denotes the number of individuals captured at least once in occasions $k(t)-K,\ldots,l-1$, not captured on occasion l and recaptured at least once in occasions $l+1,\ldots,k(t)+K$. If we suppose a local polynomial model for the number of marked individuals for each j, $\tilde{M}_l(t,K)=\sum_{q=0}^{p}\alpha_q(t_l-t_j)^q$, this results in the estimating equations

$$\sum_{l\in\mathcal{W}_t} w_l(t_j)\Big\{z_l(t)n_l - r_l(t)\Big[\sum_{q=0}^{p}\alpha_q(t_l-t_j)^q - m_l(t)\Big]\Big\}G_p(t_l-t_j)=0,$$

and we estimate $\tilde{M}_j(t,K)$ by $\bar{M}_j(t,K)=\hat{\alpha}_0$. However, for computational simplicity in the simulations and application below, we used the locally constant model. Furthermore, we obtain $\hat{M}_j^*(t,K)=\bar{M}_j(t,K)+[(n_j/r_j(t))R_j+f_{1,j-1}(t)]\hat{\gamma}^2(t,K)$, where $f_{1,j-1}(t)$ is the estimated number of individuals that are caught once up to occasion $j-1$, and R_j denotes the number of individuals caught at least twice in occasions $k(t)-K,\ldots,j$ and once in $j+1,\ldots,k(t)+K$.

5. Bandwidth Selection

The empirical-bias bandwidths of Ruppert (1997) was found to be computationally intractable. To choose the optimal local bandwidth, we instead employ the bootstrap to estimate the MSE at a grid of times for an initial bandwidth set $\mathcal{H}$ from which we may select optimal bandwidths. For the bandwidths not in the initial bandwidth set $\mathcal{H}$, we use interpolation. In order to construct our bootstrap samples, rather than reconstruct the entire process we construct samples for each time of interest. We follow the approach of Chao et al. (2001) for closed populations and employ a non-parametric bootstrap by resampling from the captured individuals. We suppose that the bandwidth h has been determined. Let t denote the time of the capture occasion considered, and let $k(t)-K,\ldots,k(t)+K$ be the capture occasions considered. Let $\mathcal{N}_{t,K}$ denote the capture histories of the $n_{t,K}$ individuals captured on at least one occasion in $k(t)-K,\ldots,k(t)+K$. We assign probabilities $n_{t,K}^{-1}$ to each capture history in $\mathcal{N}_{t,K}$ and construct a bootstrap sample of size $n_{t,K}$. For an estimator $\hat{N}(t)$, the number of uncaptured individuals is fixed as $\hat{N}(t)-n_{t,K}$. In order to determine the bandwidth, it is necessary to calculate the bias and variance for the bandwidths listed in Section 5.1. Thus several bootstrap samples are required at each time t considered. This procedure is repeated each time a bootstrap sample is required. Bootstrap bias and variance thus can be obtained and applied to yield the corresponding MSE, in turn the optimal local bandwidth.

5.1. One step bootstrap estimation of the bias and variance

Carroll et al. (1998) proposed estimating the variance of the estimators, and hence the mean squared error, using the sandwich method. However, we have reservations about this approach due to the number of nuisance parameters involved in the procedure. Note that even in closed populations the bootstrap is typically used to estimate the variance (Chao et al. (2001), Huggins and Chao (2001)). In the present setting the conventional bootstrap will involve an enormous amount of computation and we propose to use the one-step bootstrap (e.g., §5.4.7 of Shao and Tu 1995).

Let P_n denote the model that generated the data and $\hat{P}_n$ the estimate based on the sample. Let $S_n(\theta, P_n) = [S_{n,\beta}(\theta, P_n)^T, S_{n,\bar{p}(t)}(\theta, P_n)]^T$ denote the estimating equations for $\boldsymbol{\theta} = [\beta_0, \ldots, \beta_p, \bar{p}(t)]^T$. Let $g(\theta, P_n) = \theta - S_n(\theta, P_n)$ so that the estimator $\hat{\theta}_n$ is a fixed point of $g(\theta, P_n)$. Let P_n^* denote the bootstrap analogue of P_n. The computation of the one-step bootstrap estimator requires the calculation of the matrix of derivatives of $g(\theta, P_n)$. Note that

$$\frac{\partial S_{n,\beta}(\theta)}{\partial \beta_k} = \sum_{j \in \mathcal{W}_t} \frac{w_j(t)}{[1 - C_{j-1}(t,K)]} \left\{ \frac{\bar{p}(t)(t_j - t)^k [u_j(t) + M_j^*(t,K)]}{[\sum_{l=0}^p \beta_l (t_j - t)^l]^2} \right\} G_p(t_j - t),$$

$$\frac{\partial S_{n,\beta}(\theta)}{\partial \bar{p}(t)} = \sum_{j \in \mathcal{W}_t} \frac{w_j(t)}{[1 - C_{j-1}(t,K)]} \left\{ \frac{\sum_{l=0}^p \beta_l (t_j - t)^l - M_j^*(t,K)}{\sum_{l=0}^p \beta_l (t_j - t)^l} \right\} G_p(t_j - t),$$

$$\frac{\partial S_{n,\bar{p}(t)}(\theta)}{\partial \beta_k} = \bar{p}(t) \sum_{j \in \mathcal{W}_t} w_j(t)(t_j - t)^k, \text{ and } \frac{\partial S_{n,\bar{p}(t)}(\theta)}{\partial \bar{p}(t)} = \sum_{j \in \mathcal{W}_t} w_j(t) \sum_{l=0}^{p} \beta_l (t_j - t)^l.$$

Let I_{p+2} denote the $p+2$ dimensional identity matrix and let

$$\mathcal{G}_n(\theta) = \frac{\partial g(\theta, P_n)}{\partial \theta} = I_{p+2} - \frac{\partial S_n(\theta)}{\partial \theta} = I_{p+2} - \begin{bmatrix} \frac{\partial S_{n,\beta}(\theta)}{\partial \beta} & \frac{\partial S_{n,\beta}(\theta)}{\partial \bar{p}(t)} \\ \frac{\partial S_{n,\bar{p}(t)}(\theta)}{\partial \beta} & \frac{\partial S_{n,\bar{p}(t)}(\theta)}{\partial \bar{p}(t)} \end{bmatrix}.$$

Let $\hat{\theta}^{*[1]}$ be defined by $\hat{\theta}^{*[1]} = g(\hat{\theta}_n, \hat{P}_n^*) = \hat{\theta}_n - S_n(\hat{\theta}_n, \hat{P}_n^*)$. Then following (5.53) of Shao and Tu (1995), we take $\tilde{\theta}^{*[1]} = \hat{\theta}_n + [I_{p+2} - \mathcal{G}_n(\hat{\theta}_n, \hat{P}_n^*)]^{-1}(\hat{\theta}^{*[1]} - \hat{\theta}_n)$. The sample bias and variance from the one-step bootstrap can then be used to estimate the bias and variance.

5.2. MSE and optimal local bandwidth

Having determined the bootstrap bias and variance estimate over time grids $s_1 < \cdots < s_a$ for an initial bandwidth set $\mathcal{H} = \{h_1, \ldots, h_b\}$, we compute the corresponding MSE at each of these time grids. Therefore, we have a $b \times a$ MSE matrix. For other MSEs corresponding to bandwidths not in the initial bandwidth set $\mathcal{H}$, we apply a smoothing spline to interpolate using the obtained b MSEs at each time s_i, $i = 1, \ldots, a$. Therefore, for each time s_i, $i = 1, \ldots, a$, we can choose the corresponding optimal local bandwidth $h_i^*, i = 1, \ldots, a$, by selecting the minimal MSE. Eventually, we obtain the optimal local bandwidth estimates with respect to each time grid. Then these are applied to re-estimate the population size using the method discussed in the foregoing sections.

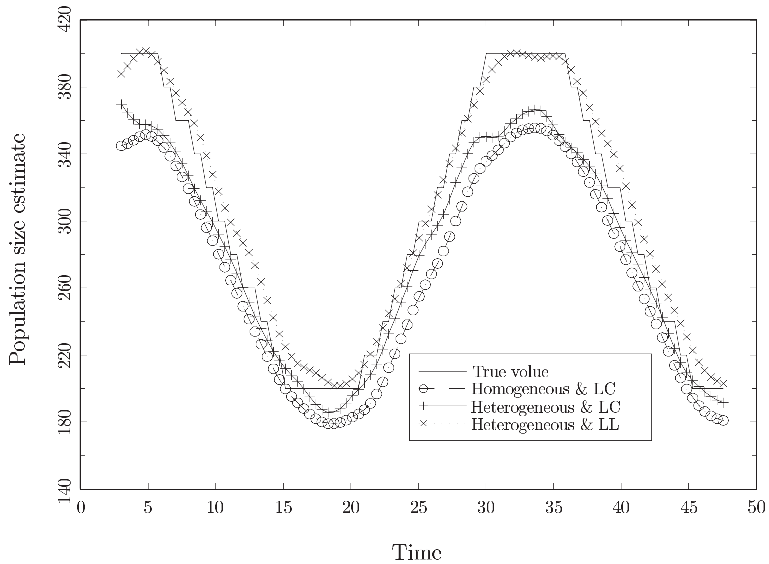

Figure 1. The average of 100 estimates arising from a population with beta(5,5) capture probabilities using bandwidth $h = 3$. The estimated population sizes were computed using the estimators derived under the three models for the capture probabilities and population size: (1) Homogeneous and locally constant; (2) Heterogeneous and locally constant; (3) Heterogeneous and locally linear.

6. Simulations

6.1. The population size

We carry out a limited simulation to assess the performance of the proposed estimator. We simulate the capture probabilities as random variables from a beta distribution with mean $\alpha/(\alpha+\beta)$ and squared CV $\beta/\{\alpha(\alpha+\beta+1)\}$. The ten beta distributions listed in Table 1 were considered. A population with cyclic-variation in population size (emigration(reduction) $\rightarrow$ immigration(addition) $\rightarrow$ emigration(reduction)) was generated.

We considered 51 evenly spaced occasions and seven different periods. The population size, which changed relatively smoothly from one period to another period, is displayed in Figure 1. To construct this population, in the first period (from occasion 1 to 6), there were 400 individuals in the population. In the second period (occasions 7 to 15), the individuals emigrated sequentially and the population size dropped from 380 to 220 in steps of 20. In the third period (occasions 16 to 21), the population size was kept at a fixed level of 200. In the fourth period (occasions 22 to 30), the individuals that left the population in the third period reentered the population and hence the population size increases from 220 to 380 by sequential increments of 20. In the fifth period (occasions 31 to 36), the population size returned to the original population size of 400 and maintained the constant level within this period. In the sixth period (occasions 37 to 45), the population size started again to drop sequentially from 380 to 220 by reductions of 20. It stopped dropping until the beginning of the next period. In the last period (occasions 46 to 51), the population size was reduced to 200 and kept constant within this period.

Due to the complexity of the computations for selecting optimal bandwidths, we first examined the performance of the proposed estimators under fixed bandwidths, $h = 3$, 4 and 5. Computation of the weighted martingale estimators requires a choice of an appropriate kernel function. Usually, a symmetric and unimodal kernel function is used. In the simulations, we considered three kernel functions: $Q(v) = (15/16)(1 - v^2)^2$, $-1 \leq v \leq 1$; the Epanechnikov kernel $Q(v) = (3/4)(1 - v^2)$, $-1 \leq v \leq 1$; the triangular kernel $Q(v) = (1 - |v|)$, $-1 \leq v \leq 1$. The values of each kernel function were set to zero for $v < -1$ or $v > 1$. Estimates obtained from solving a smooth estimating equation need not be smooth. We produced a smoothing.

We first partitioned the whole capture period, from $h + 0.01$ to $h + 46.01$, in steps of 0.45 to yield 100 evenly spaced time grids. We generated the capture history matrix with 51 capture occasions using various beta-distributed capture probabilities. In total, 100 simulated experiments were conducted for each distribution. At each time grid, we estimated the population size using three estimators: (i) $\hat{N}_H^{LC}$ (Huggins and Yip, (1999)) from (4); (ii) $\hat{N}_{\bar{H}}^{LC}$ (Huggins et al. (2003)) from (3); (iii) the proposed local linear estimator $\hat{N}_{\bar{H}}^{LL}$. Moreover, we computed the error, standard error and root mean squared error estimates for each trial. Finally, we took the average over these 100 trials, they are in Table 1 as BIAS, S.E. and RMSE.

In Table 1, for brevity, we only show the results obtained under the fixed bandwidths $h = \{3, 4, 5\}$ for the quartic kernel. The three kernel functions produced similar results although the RMSE may be somewhat different. However, in general, the estimates arising from the quartic kernel perform better than the others, in RMSE sense. In Figure 1 we plot the average of 100 estimates of the population size in the beta(5,5) case with bandwidth $h = 3$.

In Table 1 observe that, as measured by the MSE, our proposed estimator generally performs better than the other two estimators, with the estimators $\hat{N}_{\bar{H}}^{LC}$ and $\hat{N}_{\bar{H}}^{LL}$ being superior to $\hat{N}_H^{LC}$ as they take the individual heterogeneity effect into consideration. The estimator $\hat{N}_{\bar{H}}^{LL}$ typically has smaller bias than $\hat{N}_{\bar{H}}^{LC}$. However, the penalty is a larger standard error in some cases. Even though the proposed estimator has better performance in RMSE sense, the proposed estimator is not unbiased either and negative bias occurs when the beta-distributed capture probabilities have a high coefficient of variation. When the CV is modest, the two estimators that allow individual heterogeneity perform well for bandwidths of 4 or 5. However, in cases with high CV, for example beta(1,1), beta(2,10) and beta(0.5,0.5), all estimators seriously underestimate the true population size. As noted by Chao et al. (2001) and demonstrated in Huggins and Chao (2001), in these cases there exist some individuals that are essentially uncatchable, which results in a negative bias. We also simulated the other population from a death-only model with survival rate 0.9 (the same simulation considered in Huggins et al. (2003)). The same ten beta models for capture probability settings were considered. The results, not reported in detail here, show that the performance of the proposed estimator $\hat{N}_{\bar{H}}^{LL}$ was not superior to $\hat{N}_{\bar{H}}^{LC}$ in this case.

6.2. Optimal bandwidths

We also studied the effect of the different kernel functions on the optimal bandwidths. We adopted the same cyclic-variation population discussed in the previous simulation study, but to reduce the amount of computations only the optimal bandwidths at the time points, $t = \{8, 16, 24, 32, 40\}$, were calculated with the capture probability taken to be beta(5,5). We conducted 100 simulation experiments and computed optimal bandwidths for the three various kernel functions at each point. Due to the generally better performance of the local linear estimator under the cyclic-variation population, we only focused on the estimator $\hat{N}_{\bar{H}}^{LL}$. The averages and standard errors, in parentheses, of the optimal bandwidths obtained from 100 simulations at the five time points were very similar. This suggests that the effect of the kernel functions on determination of the optimal bandwidth may be not critical. This results are similar to the results of ordinary kernel smoothing analyses.

The relationship between the capture probability and the optimal bandwidth is also of interest (an associate editor conjectures that a smaller optimal bandwidth is needed if the mean capture probability is higher). In order to investigate this, we again used the cyclicly-varying population and the quartic kernel in simulations. Four capture probability distributions, beta(2,0.667), beta(3,1.5), beta(5,5) and beta(7,14), were considered − they have the same coefficient of variation (CV= 0.302) but different mean capture probabilities $\bar{p}(t)$. We focused on the estimator $\hat{N}_{\bar{H}}^{LL}$ and calculated the corresponding optimal bandwidths in 100 simulations at the five time points $t = \{8, 16, 24, 32, 40\}$. The pattern of optimal bandwidths for different capture probabilities indeed suggested that the associate editor's conjecture is correct and that the higher the capture probabilities, the smaller the optimal bandwidth. Moreover, the increase was larger when the population was smaller.

Table 1. Simulation results for different bandwidths under quartic kernel.

Bandwidth		$h=3$			$h=4$			$h=5$		
Estimates		$\hat{N}_H^{LC}$	$\hat{N}_{\bar{H}}^{LC}$	$\hat{N}_{\bar{H}}^{LL}$	$\hat{N}_H^{LC}$	$\hat{N}_{\bar{H}}^{LC}$	$\hat{N}_{\bar{H}}^{LL}$	$\hat{N}_H^{LC}$	$\hat{N}_{\bar{H}}^{LC}$	$\hat{N}_{\bar{H}}^{LL}$
$beta(10,10)$	BIAS	-22.5	-14.8	14.0	-13.3	-6.9	12.5	-8.2	-4.7	11.5
$\bar{p}=0.500$	S.E.	13.8	14.9	12.1	12.6	14.6	11.7	15.1	19.3	15.5
$cv=0.218$	RMSE	26.4	21.0	18.5	18.3	16.2	17.1	17.2	19.9	19.3
$beta(10,20)$	BIAS	-35.0	-18.6	5.8	-24.0	-9.4	6.1	-17.1	-4.8	5.5
$\bar{p}=0.333$	S.E.	18.3	17.9	18.8	15.6	16.7	16.0	16.3	18.6	15.1
$cv=0.254$	RMSE	39.5	25.8	19.7	28.6	19.2	17.1	23.6	19.2	16.1
$beta(5,5)$	BIAS	-32.2	-19.9	5.4	-21.9	-11.8	4.9	-15.6	-8.4	5.1
$\bar{p}=0.500$	S.E.	15.8	16.2	12.3	14.3	19.8	12.9	16.4	19.8	14.6
$cv=0.302$	RMSE	35.9	25.7	13.4	26.1	15.9	11.9	22.6	21.5	15.4
$beta(5,8)$	BIAS	-42.9	-23.8	-1.8	-31.5	-14.6	-0.5	-24.1	-9.7	0.1
$\bar{p}=0.385$	S.E.	18.2	17.9	15.5	16.2	17.3	13.6	17.3	20.1	14.4
$cv=0.338$	RMSE	46.6	29.8	15.6	35.4	22.7	13.6	29.7	22.3	14.4
$beta(4,8)$	BIAS	-53.7	-29.1	-10.8	-42.0	-19.3	-8.4	-34.4	-13.9	-7.6
$\bar{p}=0.333$	S.E.	20.8	19.4	18.1	18.5	18.8	15.6	19.2	20.6	16.4
$cv=0.392$	RMSE	57.6	34.9	21.1	45.9	26.9	17.8	39.4	24.9	18.1
$beta(3,5)$	BIAS	-57.4	-32.4	-15.5	-45.6	-22.8	-13.1	-37.4	-17.1	-11.4
$\bar{p}=0.375$	S.E.	21.2	19.9	17.1	18.9	18.8	15.0	19.6	21.6	15.6
$cv=0.430$	RMSE	61.2	38.1	23.1	49.4	29.5	19.9	42.3	27.5	19.3
$beta(3,10)$	BIAS	-75.2	-40.9	-26.3	-63.4	-31.3	-25.9	-56.0	-23.7	-25.0
$\bar{p}=0.231$	S.E.	28.0	24.8	27.3	24.3	22.4	23.8	24.0	23.5	22.7
$cv=0.489$	RMSE	80.2	47.8	37.9	67.9	38.5	35.2	60.9	33.4	33.8
$beta(1,1)$	BIAS	-75.5	-53.5	-44.3	-64.5	-43.8	-41.9	-56.1	-37.3	-38.1
$\bar{p}=0.500$	S.E.	24.9	23.8	18.9	23.7	23.7	19.6	24.3	25.6	20.5
$cv=0.557$	RMSE	79.5	58.6	48.2	68.7	49.8	46.3	61.1	45.3	43.2
$beta(2,10)$	BIAS	-102.5	-61.9	-49.4	-89.3	-48.7	-46.5	-82.3	-41.5	-48.5
$\bar{p}=0.167$	S.E.	32.3	31.6	37.7	31.4	28.4	32.2	30.1	28.5	30.4
$cv=0.600$	RMSE	108.4	69.5	62.2	94.6	56.4	56.5	87.6	50.3	57.2
$beta(0.5,0.5)$	BIAS	-99.2	-80.5	-76.2	-89.4	-71.1	-73.3	-81.8	-64.3	-68.6
$\bar{p}=0.500$	S.E.	30.0	29.1	24.9	29.4	29.4	26.2	29.5	30.8	26.5
$cv=0.707$	RMSE	103.7	85.6	80.1	94.2	77.0	77.9	86.9	71.3	73.6

7. Example

Huggins and Yip (1999) and Huggins et al. (2003) have previously examined a data set concerning captures of Prinia flaviventris at Mai Po in Hong Kong. The banding data was collected weekly on the bird species Prinia flaviventris at the Mai Po bird sanctuary in Hong Kong over 34 weeks from September 1991-April 1992. A total of 216 birds were captured in the period considered. More details of this data are available in Huggins and Yip (1999) and Huggins et al.

(2003). We examine this data set using a local linear model.

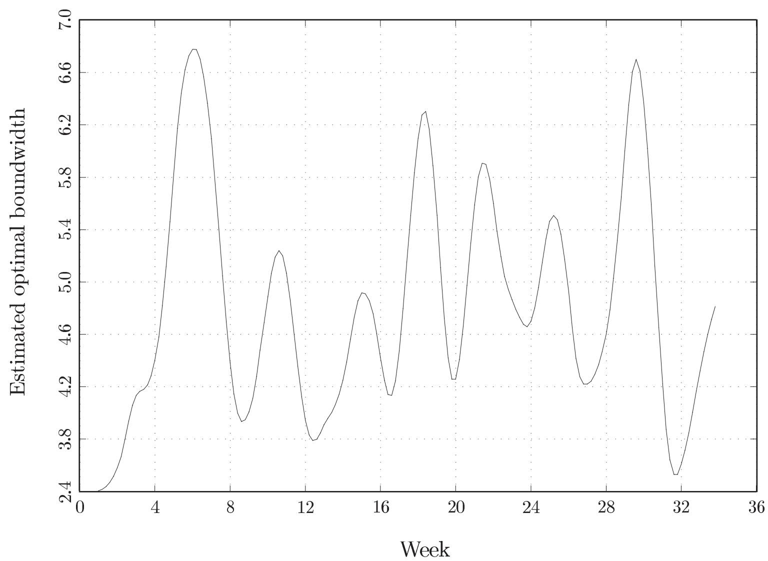

Figure 2. The optimal local bandwidth over the whole experiment.

We took 164 evenly spaced time points from 1 to 33.8 in increments of 0.2. We again used the quartic kernel and computed the estimators $\hat{N}_H^{LC}$, $\hat{N}_{\bar{H}}^{LC}$ and $\hat{N}_{\bar{H}}^{LL}$. Using the approach of Section 5, we selected an optimal local bandwidth. An initial bandwidth set $\mathcal{H} = \{h_j, j = 1, \ldots, 10\} = \{3, 3.5, \ldots, 7.5\}$ was considered. We first chose a bandwidth from the bandwidth set $\mathcal{H}$. For this bandwidth h_j, we constructed a bootstrap sample with $2h_j + 1$ occasions and calculated the estimate $\hat{N}_{\bar{H}}^{LL}$ in this window. By repeating this procedure 200 times, we estimated the bias, variance and hence mean squared error at each time point. Eventually, we had a 10×164 array of the MSEs, where the rows denote the bandwidth set $\mathcal{H}$ and the columns the time axis. As illustrated in Section 5.2, we applied a smoothing spline to interpolate the MSE for other bandwidths not in the initial bandwidth set $\mathcal{H}$. To do this we divided the intervals between the bandwidths into ten evenly spaced subintervals and interpolated the MSE. Then we choose the optimal bandwidth based on minimal MSE criterion and obtain the optimal bandwidths for all time points. The results are shown in Figure 2.

Applying the selected optimal bandwidths, we can estimate the population size and the slope term in the local polynomial model as discussed in Section 2. After selecting the optimal bandwidth, we applied the new optimal bandwidths to all three estimators to show the final smoothed curve of estimated population size in Figure 3. Thus the bandwidth is optimal for $\hat{N}_{\bar{H}}^{LL}$, but not for the other two estimators.

All three estimators give the same overall pattern for the population dynamics. The estimated coefficients of variation ranged from zero to 1.81 (mean 0.44 and standard error 0.54), and was largest between weeks 16 and 28, giving a quite different pattern to Figure 2. This demonstrates heterogeneity among individuals in the population and, in this region, the estimator $\hat{N}_H^{LC}$ gave noticeably lower estimates than the other two. Under the local linear model, the estimated slope calculated for evenly spaced time points in increments of 0.2 weeks ranged between -628.80 and 510.57 with mean -8.13. The large positive peak was around 21 weeks and the negative peak near 24 weeks. At the large negative peak, the locally constant and locally linear estimators produced different results, but they are similar where the estimated slope was positive. In Figure 3, we see that the local linear and local constant models are generally similar, with the local linear model differing the most from the locally constant model where the estimated slope was large and negative.

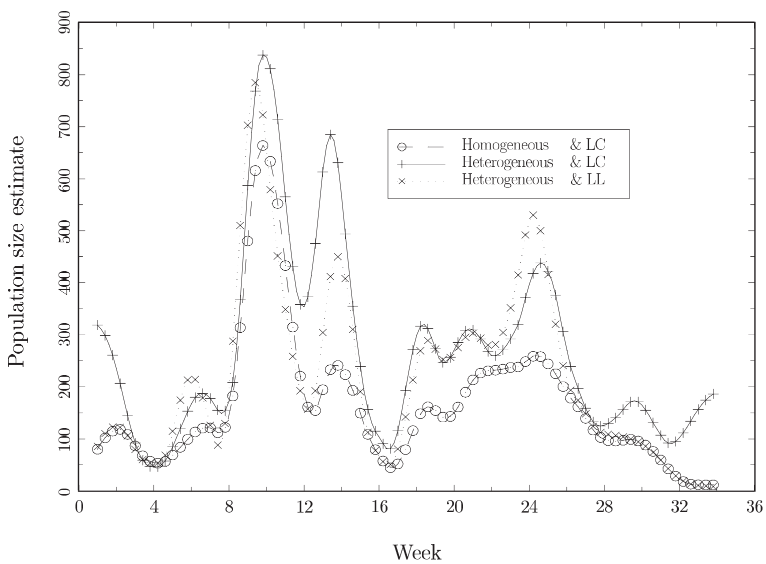

Figure 3. The estimated population sizes under the three models considered for the optimal bandwidth.

8. Discussion

The use of local polynomials allows more realistic assumptions on the population size than does the locally constant approach. In the simulation study, we have shown that the local polynomial approach also generally improves the bias and the MSE. This being particularly so when the changes in the population size were the largest. Other simulations, not reported here, show that even if the population changes more sharply than that in the simulations reported here, the local linear estimator still performs quite well. In other simulations, also not reported in detail here, we considered a population with individuals entering and leaving at the same rate so the population size was approximately constant. In this case the local linear estimator $\hat{N}_{H}^{LL}$ performed better than the other two estimators.

The proposed method of the selection of optimal local bandwidth is conceptually simple and computationally feasible, where other possible approaches are not. We saw in simulations that there was little difference between the optimal bandwidths for some common kernel functions and that the higher the capture probability is, the smaller the optimal bandwidth. The same data driven procedure could be applied to determine the degree of polynomial as, in a similar fashion to the selection of the optimal bandwidth in Section 5.2, we could calculate the MSE of the proposed estimator at each time point based on polynomial models of different orders. Hence, the degree with minimum MSE could theoretically be selected at each time and the optimal degree could even be obtained locally, or a fixed degree of polynomial over whole experimental period could be obtained. Thus our extensions of the local linear kernel smoothing approach and our data driven bandwidth selection method, whilst computationally intensive, have the potential to greatly enhance the utility of the local estimating method in capture-recapture studies.

Acknowledgements

This research was carried out when the second author visited National Tsing Hua University, Hsin-Chu, Taiwan. This visit was supported by an exchange program between Australian Academy of Science and Taiwan National Science Council.

Appeddix A. The Estimating Equations

Let $\mathcal{F}_{j-1}$ denote the capture histories on occasions $k(t)-K,\ldots,j-1$. Then it is easily shown that $E[u_j(t)|\mathcal{F}_{j-1}] = [N_{t_j} - M_j^*(t,K)]\bar{p}(t_j) = [\sum_{l=0}^{p}\beta_l(t_j-t)^l - M_j^*(t,K)]\bar{p}(t_j)$, $E[m_j(t)|\mathcal{F}_{j-1}] = M_j^*(t,K)\bar{p}(t_j)$ and $E[n_j|\mathcal{F}_{j-1}] = N_{t_j}\bar{p}(t_j) = \sum_{l=0}^{p}\beta_l(t_j-t)^l\bar{p}(t_j)$. Under our assumptions on the removals from the population and arrivals into the population, $\bar{p}(t_j)$ and $\bar{p}(t)$ estimate the same quantities. Thus we replace $\bar{p}(t_j)$ by $\bar{p}(t)$ in the estimating functions and use a locally constant approach to estimate $\bar{p}(t)$. Hence the estimating functions are $g_{j1}(t) = u_j(t) - [\sum_{l=0}^{p}\beta_l(t_j-t)^l - M_j^*(t,K)]\bar{p}(t)$ and $g_{j2}(t) = m_j(t) - M_j^*(t,K)\bar{p}(t)$. In our case, $\theta^T = [\beta_0,\ldots,\beta_p,\bar{p}(t)]$. Now $\partial g_{j1}(t)/\partial\beta_l = -(t_j-t)^l\bar{p}(t)$, $l=0,\ldots,p$, $\partial g_{j1}(t)/\partial\bar{p}(t) = -[\sum_{l=0}^{p}\beta_l(t_j-t)^l - M_j^*(t,K)]$, $\partial g_{j2}(t)/\partial\beta_l = 0$, $l=0,\ldots,p$, and

$\partial g_{j2}(t)/\partial \bar{p}(t) = -M_j^*(t,K)$. Thus for $j \in \mathcal{W}_t$, we obtain the conditional expectation D_j and conditional covariance V_j as follows:

$$D_j = \begin{bmatrix} -\bar{p}(t) & -(t_j-t)\bar{p}(t) & \dots & -(t_j-t)^p\bar{p}(t) & -[\sum_{l=0}^p \beta_l(t_j-t)^l - M_j^*(t,K)] \\ 0 & 0 & \dots & 0 & -M_j^*(t,K) \end{bmatrix},$$

$$V_j = N_{t_j}\bar{p}(t)[1-\bar{p}(t)(1+\gamma_t^2)] \begin{bmatrix} 1-C_{j-1}(t,K) & 0 \\ 0 & C_{j-1}(t,K) \end{bmatrix}.$$

After removing terms that do not depend on j, the estimating functions for the β is the weighted sum of the terms

$$\left(D_j^T V_j^{-1} g_j\right)_\beta = -\frac{\{u_j(t) - [\sum_{l=0}^p \beta_l(t_j-t)^l - M_j^*(t,K)]\bar{p}(t)\}}{[1-C_{j-1}(t,K)]\sum_{l=0}^p \beta_l(t_j-t)^l} G_p(t_j-t),$$

where $G_p^T(v) = [1, v, v^2, \dots, v^p]$. This yields (1). The estimating function for $\bar{p}(t)$ is the weighted sum of the terms $\left(D_j^T V_j^{-1} g_j\right)_{\bar{p}(t)} = -[n_j - \bar{p}(t)\sum_{l=0}^p \beta_l(t_j-t)^l]$, which yields (2).

Appeddix B. Computation of the Estimates

Let $H_{j-1}(t,K) = [1-C_{j-1}(t,K)][\beta_0^{(n-1)} + \beta_1^{(n-1)}(t_j-t)]$. Given an initial value $(\beta_0^{(0)}, \beta_1^{(0)})$, we can obtain the estimates by solving the following iterative formulae until convergence:

$$\bar{p}^{(n-1)}(t) = \frac{\sum_{j\in\mathcal{W}_t} w_j(t) n_j}{\sum_{j\in\mathcal{W}_t} w_j(t)[\beta_0^{(n-1)} + \beta_1^{(n-1)}(t_j-t)]},$$

$$\beta_0^{(n)} = \frac{\sum_{j\in\mathcal{W}_t} w_j(t) H_{j-1}^{-1}(t,K)\{u_j - [\beta_1^{(n-1)}(t_j-t) - \hat{M}_j^*(t,K)]\bar{p}^{(n-1)}(t)\}}{\sum_{j\in\mathcal{W}_t} w_j(t) H_{j-1}^{-1}(t,K)\bar{p}^{(n-1)}(t)},$$

$$\beta_1^{(n)} = \frac{\sum_{j\in\mathcal{W}_t} w_j(t) H_{j-1}^{-1}(t,K)\{u_j - [\beta_0^{(n-1)} - \hat{M}_j^*(t,K)]\bar{p}^{(n-1)}(t)(t_j-t)\}}{\sum_{j\in\mathcal{W}_t} w_j(t) H_{j-1}^{-1}(t,K)\bar{p}^{(n-1)}(t_j-t)^2}.$$

The resulting estimated population size, slope and mean capture probability thus are $N_{\hat{H}}^{LL}(t) = \hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\bar{p}}(t)$ respectively.

References

Burnham, K. P. and Overton, W. S. (1978). Estimation of the size of a closed population when capture probabilities vary among animals. *Biometrika* **65**, 625-633.

Carroll, R. J., Ruppert, D. and Welsh, A. H. (1998). Local estimating equations. *J. Amer. Statist. Assoc.* **93**, 214-227.

Chao, A., Lee, S.-M. and Jeng, S.-L. (1992). Estimating population size for capture-recapture data when capture probabilities vary by time and individual animal. *Biometrics*, **48**, 201-216.

Chao, A. and Huggins, R. M. (2001). Closed capture-recapture models. To appear in *The Handbook of Capture-Recapture Methods* (Edited by B. Manly, T. L. McDonald and S. C. Amstrup). Princeton University Press.

Chao, A., Yip, P. S. F., Lee, S.-M. and Chu, W. (2001). Population size estimation based on estimating functions for closed capture-recapture models. *J. Statist. Plann. Inference* **92**, 213-232.

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Chapman and Hall, New York.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modeling and its Applications*. Chapman and Hall, London.

Fan, J., Heckman, N. E., and Wand, M. P. (1995). Local polynomial kernel regression for generalized linear models and quasi-likelihood functions. *J. Amer. Statist. Assoc.* **90**, 141-150.

Good, I. J. (1953). The population frequencies of species and the estimation of population parameters. *Biometrika* **40**, 237-264.

Huggins, R. M. and Chao, A. (2002). Asymptotic properties of an optimal estimating function approach to the analysis of mark recapture data. *Comm. Statist.* **31**, 575-595.

Huggins, R. M. and Yip, P. S. F. (1999). Estimating of the size of an open population from capture-recapture data using weighted martingale methods. *Biometrics* **55**, 387-395.

Huggins, R. M., Yang, H.-C., Chao, A. and Yip, P. S. F. (2003). Population size estimation using local sample coverage for open populations. *J. Statist. Plann. Inference* **113**, 699-714.

Hwang, W.-D. and Chao, A. (1995). Quantifying the effects of un-equal catchabilities on Jolly-Seber estimators via sample coverage. *Biometrics* **51**, 128-141.

Lee, S.-M. and Chao, A. (1994). Estimating population size via sample coverage for closed capture-recapture models. *Biometrics* **50**, 88-97.

Pledger, S. and Efford, M. (1998). Correction of bias due to heterogeneous capture probability in capture-recapture studies of open populations. *Biometrics* **54**, 888-898.

Ruppert, D. (1997). Empirical-bias bandwidths for local polynomial nonparametric regression and density estimation. *J. Amer. Statist. Assoc.* **92**, 1049-1062.

Shao, J. and Tu, D. (1995). *The Jackknife and Bootstrap*. Springer-Verlag, New York.

Institute of Statistics, National Tsing Hua University, Hsin-Chu 30043, Taiwan.

Department of Statistical Science, La Trobe University, Bundoora, Victoria 3086, Australia.

E-mail: r.huggins@latrobe.edu.au

(Received October 2001; accepted April 2003)