

CONTINUOUS-TIME CAPTURE–RECAPTURE MODELS WITH TIME VARIATION AND BEHAVIOURAL RESPONSE

WEN-HAN HWANG^{1,2}, ANNE CHAO^{2*} AND PAUL S.F. YIP³

Feng-Chia University, National Tsing Hua University and University of Hong Kong

Summary

This paper develops a likelihood-based inference procedure for continuous-time capture–recapture models. The first-capture and recapture intensities are assumed to be in constant proportion but may otherwise vary arbitrarily through time. The full likelihood is partitioned into two factors, one of which is analogous to the likelihood in a special type of multiplicative intensity model arising in failure time analysis. The remaining factor is free of the non-parametric nuisance parameter and is easily maximized. This factor provides an estimator of population size and an asymptotic variance under a counting process framework. The resulting estimation procedure is shown to be equivalent to that derived from a martingale-based estimating function approach. Simulation results are presented to examine the performance of the proposed estimators.

Key words: non-parametric maximum likelihood estimator; optimal estimating function; population size; quasi-likelihood.

1. Introduction

Although many capture–recapture models in discrete time have been developed for estimating parameters in biological and ecological sciences, the development of continuous-time counterparts has seldom been discussed. In a discrete-time model, a group of animals (or individuals) is captured in traps at the first trapping occasion. Each trapping day or trapping night is typically regarded as an occasion. The captured animals are uniquely tagged or marked and released back into the population. At each subsequent trapping occasion, another group of animals, containing both marked and unmarked animals, is caught. A unique tag is attached to every previously unmarked animal and information on recaptures is recorded. This sampling process continues over a number of occasions. The maximum count for each animal is the number of occasions because each animal can be counted at most once on any trapping occasion.

For a continuous-time model, typically only one animal is captured at any time. In addition to tagging the animal, the exact capture time is also recorded. Thus each capture is regarded as a trapping occasion. Earlier work includes an example of capturing butterflies (Craig, 1953) and the mathematical foundation developed by Darroch (1958). Recently, as

Received August 1999; revised October 2000; accepted October 2000.

* Author to whom correspondence should be addressed.

¹ Dept of Statistics, Feng-Chia University, Tai-Chung, Taiwan 40724.

² Institute of Statistics, National Tsing Hua University, Hsin-Chu, Taiwan 30043.

e-mail: chao@stat.nthu.edu.tw

³ Dept of Statistics and Actuarial Science, University of Hong Kong, Pokfulam Road, Hong Kong.

Acknowledgments. The authors thank an Associate Editor, the Technical Editor and a referee for thoughtful comments and suggestions, which have significantly improved the paper. This work was done while the third author was visiting National Tsing Hua University. Research was partially supported by the National Science Council of Taiwan under contracts NSC-85-2121-M007-004 and NSC-86-2115-M007-033. Part of the material is based on the PhD thesis of the first author under the supervision of the second author.

indicated by Wilson & Anderson (1995), biologists have applied this type of model to study sperm whale, grizzly bear and butterfly populations. This model has also found application in software reliability theory; see Nayak (1988) and Voas & McGraw (1998 Chapter 3). In the latter cases involving a debugging process, each bug is regarded as an animal and the detection times for each bug can be considered ‘capture times’.

In this paper, we present a likelihood-based inference for a continuous-time capture–re-capture experiment using a counting process approach. In a pioneering work, Becker (1984) established the mathematical framework of a counting process for two continuous-time models. Assuming the capture intensity varied with time (model \mathcal{M}_t), Becker constructed an estimating equation based on a martingale approach. The same paper also considered a more complicated model for which the intensity was also allowed to be heterogeneous (model \mathcal{M}_{th}). However, the estimation for model \mathcal{M}_{th} could produce unreasonable estimates for population size, as shown by Chao & Lee (1993) and Wilson & Anderson (1995). In related studies, Becker & Heyde (1990) and Yip, Fong & Wilson (1993) developed the maximum likelihood estimator (MLE) for population size and a class of high-efficiency estimators for model \mathcal{M}_t . Since then, Chao & Lee (1993), Yip & Chao (1996), Yip, Huggins & Lin (1996) and Lin & Yip (1999) have also discussed models that allow heterogeneous capture intensities.

In ecological experiments, animals usually exhibit a behavioural response to capture and become either ‘trap-happy’ or ‘trap-shy’. Therefore, the first-captures and the subsequent recaptures differ with respect to capture intensity. In software debugging processes, fault seeding is usually used; the seeded and indigenous faults may have different probabilities of being detected (Duran & Wiorkowski, 1981; Yip, 1995; Voas & McGraw, 1998 Chapter 3). This paper generalizes the model of Becker & Heyde (1990) and Yip *et al.* (1993) to a more general case. Restated, we assume that the intensity varies with time and with behavioural response. This model is usually referred to as model \mathcal{M}_{tb} in the literature. In this paper, we consider the special case that the first-capture intensity is proportional to the recapture intensity, and call it model \mathcal{M}_{tb}^* . Yip *et al.* (2000) proposed an estimation procedure using a martingale-based estimating function for this specific model. We provide an alternative and more intuitive likelihood approach. The function describing how the capture intensities vary with time is allowed to be any non-parametric function and thus there are some technical difficulties to be overcome in defining the MLE. We express the likelihood involving this time-varying function as the likelihood associated with a special type of multiplicative intensity model arising in failure-time analysis. When the cumulative time-varying function is estimated by a non-parametric maximum likelihood method, the resulting estimation procedure for population size is equivalent to that obtained by a martingale-based estimating function approach. See Lloyd (1994) and Chao, Chu & Hsu (2000) for discussion of a discrete-time analogue of the model.

Section 2 formulates the models and inference procedure for model \mathcal{M}_{tb}^* using a counting process framework. Section 3 discusses the three special cases: models \mathcal{M}_t , \mathcal{M}_b and \mathcal{M}_0 . Necessary and sufficient conditions for the existence and uniqueness of a finite MLE are derived. Section 4 presents the simulation results to show the performance of the proposed estimators.

2. Models and estimators

Assume that the population size is ν and the experiment terminates at time τ . Let $N_i(t)$ denote the number of times that the i th animal has been captured in $[0, t]$. Each $\{N_i(t); t \geq 0\}$

is a continuous-time counting process with jumps of size 1 and no two processes can jump simultaneously. Let $\mathcal{F}(t)$ denote the capture history generated by $\{N_1(u), N_2(u), \dots, N_v(u); 0 \leq u \leq t\}$. The capture intensity for the i th animal, $\lambda_i(t)$, is defined by

$$\lambda_i(t) dt = \Pr(dN_i(t) = 1 \mid \mathcal{F}(t-)).$$

Let $\lambda(t)$ denote an arbitrary intensity function defined on $[0, \tau]$. We define the following models:

$$\begin{aligned} \text{Model } \mathcal{M}_{\text{tb}}^* & \lambda_i(t) = \begin{cases} \lambda(t) & \text{for first capture,} \\ \phi\lambda(t) & \text{for recapture,} \end{cases} \\ \text{Model } \mathcal{M}_{\text{b}} & \lambda(t) = \lambda, \\ \text{Model } \mathcal{M}_{\text{t}} & \phi = 1, \\ \text{Model } \mathcal{M}_0 & \lambda(t) = \lambda \quad \text{and} \quad \phi = 1. \end{aligned}$$

For models $\mathcal{M}_{\text{tb}}^*$ and \mathcal{M}_{b} , a basic assumption is that the first-capture intensity is proportional to the recapture intensity. Let $N(t) = \sum_{i=1}^v N_i(t)$ denote the total number of captures by time t , and $M(t)$ denote the number of animals that are marked by time t . For notational simplicity, let $M = M(\tau)$ denote the number of distinct animals caught in the experiment, $N = N(\tau)$ denote the number of total captures and $m_i = N_i(\tau)$ the number of captures of the i th animal during the time period $[0, \tau]$.

If $m_i > 0$, let $t_{i1}, t_{i2}, \dots, t_{im_i}$ denote the capture times. The likelihood function can be obtained by using an argument similar to that given by Crowder *et al.* (1991 p.165). Given small increments $(\Delta t_{i1}, \Delta t_{i2}, \dots, \Delta t_{im_i})$, consider the following independent consecutive events and their associated probabilities: no capture in $(0, t_{i1})$ with probability $\exp(-\int_0^{t_{i1}} \lambda(u) du)$, one capture in $(t_{i1}, t_{i1} + \Delta t_{i1})$ with probability $\lambda(t_{i1})\Delta t_{i1}$, no capture in $(t_{i1} + \Delta t_{i1}, t_{i2})$ with probability $\exp(-\phi \int_{t_{i1} + \Delta t_{i1}}^{t_{i2}} \lambda(u) du)$, one capture in $(t_{i2}, t_{i2} + \Delta t_{i2})$ with probability $\phi\lambda(t_{i2})\Delta t_{i2}, \dots$, and so on up to no capture in $(t_{im_i} + \Delta t_{im_i}, \tau)$ with probability $\exp(-\phi \int_{t_{im_i} + \Delta t_{im_i}}^{\tau} \lambda(u) du)$. Letting all the increments tend to 0, we find that the likelihood under model $\mathcal{M}_{\text{tb}}^*$ for the i th captured animal is

$$\begin{aligned} L_i &= \lambda(t_{i1}) \exp\left(-\int_0^{t_{i1}} \lambda(u) du\right) \\ &\quad \left(\prod_{j=2}^{m_i} \phi\lambda(t_{ij}) \exp\left(-\phi \int_{t_{i,j-1}}^{t_{ij}} \lambda(u) du\right)\right) \exp\left(-\phi \int_{t_{im_i}}^{\tau} \lambda(u) du\right). \end{aligned}$$

If the i th animal is never caught, i.e. $m_i = 0$, the likelihood becomes $\exp(-\int_0^{\tau} \lambda(u) du)$. Define the cumulative capture intensity function as $\Lambda(t) = \int_0^t \lambda(u) du$, $t \in [0, \tau]$. The full likelihood based on the capture history is given by

$$\begin{aligned} L(v, \phi, \Lambda) &= \frac{v!}{(v-M)!} \left(\prod_{i=1}^M L_i\right) (\exp(-\Lambda(\tau)))^{v-M} \\ &= \frac{v!}{(v-M)!} \left(\prod_{i=1}^M \prod_{j=1}^{m_i} \lambda(t_{ij})\right) \phi^{N-M} \\ &\quad \exp\left(-v\Lambda(\tau) - (\phi-1) \sum_{i=1}^M (\Lambda(\tau) - \Lambda(t_{i1}))\right). \end{aligned} \quad (1)$$

We define the processes

$$Q(t) = v - M(t-) + \phi M(t-) \quad \text{and} \quad R(t) = v - M(t-).$$

Using the identity

$$\sum_{i=1}^M (\Lambda(\tau) - \Lambda(t_{i1})) = \int_0^\tau M(u-) d\Lambda(u),$$

we can rewrite the likelihood as $L(v, \phi, \Lambda) = L_1(v, \phi) L_2(v, \phi, \Lambda)$, where

$$L_1(v, \phi) = \frac{v!}{(v-M)!} \phi^{N-M} \prod_{i=1}^M \prod_{j=1}^{m_i} \frac{1}{Q(t_{ij})}, \quad (2)$$

$$L_2(v, \phi, \Lambda) = \left(\prod_{i=1}^M \prod_{j=1}^{m_i} Q(t_{ij}) \lambda(t_{ij}) \right) \exp \left(- \int_0^\tau Q(t) d\Lambda(t) \right),$$

First of all, we estimate $\Lambda(t)$ (which appears only in L_2) non-parametrically. For any given v and ϕ , L_2 corresponds to a probability measure based on the observed failure times $\{t_{ij}, j = 1, 2, \dots, m_i, i = 1, 2, \dots, M\}$ from a ‘multiplicative model’ with intensity $\lambda^*(t) = Q(t) d\Lambda(t)$; see Johansen (1983), Jacobsen (1984) and Andersen *et al.* (1993 pp. 128–129, 176–183). Thus the methods of inference for a multiplicative model in the context of failure-time analysis can be applied here. Although the function $\Lambda(t)$ is continuous, we extend our model to allow the estimate of this function to be any function including step functions. For the traditional maximum likelihood approach, any probability measure involved in the model must be dominated by a common measure. In the extended model containing both the discrete and the continuous types of distributions, the MLE cannot be handled using a standard method.

Following the approach of Andersen *et al.* (1993 Section IV.1.5), we use the MLE here in the broad sense of Kiefer & Wolfowitz (1956); see also Scholz (1980) for a unification approach. This approach is based on the pairwise comparison of probability measures. The non-parametric MLE of the function $\Lambda(t)$ in this sense turns out to be the Nelson–Aalen estimator; see Nelson (1972) and Aalen (1978). That is, the MLE $\hat{\Lambda}(t)$ of $\Lambda(t)$ is a step function with jump size $1/Q(t)$ at capture time $t = t_{ij}$. This yields $L_2(v, \phi, \hat{\Lambda}) = \exp(-N)$, which is free of the parameters v and ϕ . As a result, the profile likelihood $L(v, \phi, \hat{\Lambda})$ is proportional to $L_1(v, \phi)$ and the estimation of v and ϕ can be simply based on $L_1(v, \phi)$. Therefore, the MLE of (Λ, v, ϕ) is the solution of the following system of equations:

$$d\Lambda(t) = \frac{dN(t)}{Q(t)} \quad (0 \leq t \leq \tau), \quad (3)$$

$$\frac{\partial \log L_1}{\partial v} = \int_0^\tau \frac{1}{R(t)} dM(t) - \int_0^\tau \frac{1}{Q(t)} dN(t) = 0, \quad (4)$$

$$\frac{\partial \log L_1}{\partial \phi} = \frac{N-M}{\phi} - \int_0^\tau \frac{M(t-)}{Q(t)} dN(t) = 0. \quad (5)$$

Although v is an integer in fact, it is treated here as a real variable. This is a useful approximation because the likelihood is meaningful for any real number of v . The estimator for the function Λ in (3) has an intuitive interpretation. Note that given the capture history before time t ,

there are $R(t) = \nu - M(t-)$ unmarked animals with intensity $\lambda(t)$ and $M(t-)$ marked with intensity $\phi\lambda(t)$ in the population. We then have $E(dM(t) | \mathcal{F}(t-)) = R(t) d\Lambda(t)$, $E(dN(t) | \mathcal{F}(t-)) = Q(t) d\Lambda(t)$ and the estimator for $\Lambda(t)$ can be understood. Also, it can be seen that

$$E\left(\frac{\partial \log L_1}{\partial \nu}\right) = E\left(E\left(\frac{\partial \log L_1}{\partial \nu}\right) \mid \mathcal{F}(t-)\right) = 0 \quad \text{and} \quad E\left(\frac{\partial \log L_1}{\partial \phi}\right) = 0.$$

A zero-mean martingale $\{Z(t)\}$, $dZ(t) = \phi M(t-) dM(t) - R(t) dK(t)$, where $K(t) = N(t) - M(t)$ denotes the number of recaptures by time t , was considered by Yip *et al.* (2000). Then a class of estimating equations is constructed as $\int w_1(t) dZ(t) = 0$ and $\int w_2(t) dZ(t) = 0$ where $w_1(t)$ and $w_2(t)$ are weight functions for ν and ϕ , respectively. The two optimal weight functions, following Godambe (1985) are evaluated as $w_1^*(t) = 1/(R(t)Q(t))$ and $w_2^*(t) = 1/Q(t)$. The two optimal estimating equations then become

$$g_1 = \int_0^\tau w_1^*(t) dZ(t) = \int_0^\tau \frac{\phi M(t-) dM(t) - R(t) dK(t)}{R(t)Q(t)} = 0, \tag{6}$$

$$g_2 = \int_0^\tau w_2^*(t) dZ(t) = \int_0^\tau \frac{\phi M(t-) dM(t) - R(t) dK(t)}{Q(t)} = 0. \tag{7}$$

We see that (6) and (7) are identical to (4) and (5). Therefore, the two systems of equations derived from the maximum likelihood and martingale-based approach are identical for continuous-time models. This conclusion was proved for model \mathcal{M}_t in Becker & Heyde (1990); see also Heyde (1997 p. 166). However, for discrete-time models, Chao *et al.* (2000) showed that the two systems differ.

It follows from (6) and (7) that

$$\frac{\partial g_1}{\partial \phi} = \frac{\partial g_2}{\partial \nu} = - \int_0^\tau \frac{M(t-)}{Q(t)^2} dN(t).$$

The above is a necessary condition for (g_1, g_2) being the gradient of some real-valued function which is called quasi-likelihood. That is, (6) and (7) can be regarded as the resulting derivatives of the quasi-likelihood. Direct integration yields the quasi-likelihood $L_Q(\nu, \phi)$:

$$\log L_Q(\nu, \phi) \propto \int_0^\tau \log R(t) dM(t) + (N - M) \log \phi - \int_0^\tau \log Q(t) dN(t).$$

Now $L_Q(\nu, \phi) \propto L_1(\nu, \phi)$, where L_1 is defined in (2). From this point of view, our method and the martingale-based approach are based on the same likelihood function.

On the basis of the likelihood, we have

$$\begin{aligned} \frac{\partial^2 \log L_1(\nu, \phi)}{\partial \nu^2} &= - \int_0^\tau \frac{1}{R(t)^2} dM(t) + \int_0^\tau \frac{1}{Q(t)^2} dN(t), \\ \frac{\partial^2 \log L_1(\nu, \phi)}{\partial \phi^2} &= - \frac{N - M}{\phi^2} + \int_0^\tau \frac{M(t-)^2}{Q(t)^2} dN(t), \\ \frac{\partial^2 \log L_1(\nu, \phi)}{\partial \phi \partial \nu} &= \int_0^\tau \frac{M(t-)}{Q(t)^2} dN(t). \end{aligned}$$

Using the preceding derivatives, we can obtain a variance estimator of the resulting MLE by inverting the expected Fisher information matrix. Letting $P(t) = 1 - \exp(-\Lambda(t))$, we obtain $EM(t-) = \nu P(t)$, $E(N(t-) - M(t-)) = \nu\phi(\Lambda(t) - P(t))$ and $E(dN(t)) = \nu(1 - P(t))d\Lambda(t) + \nu\phi P(t)d\Lambda(t)$. Define $\psi = \phi + (1 - \phi)\exp(-\Lambda(\tau))$. If $\phi \neq 1$, we can derive the following expectations:

$$\begin{aligned} E\left(\frac{\partial^2 \log L_1(\nu, \phi)}{\partial \nu^2}\right) &= \frac{\phi(1 - e^{\Lambda(\tau)}) + \Lambda(\tau) + \log \psi}{\nu\phi} + O(\nu^{-2}), \\ E\left(\frac{\partial^2 \log L_1(\nu, \phi)}{\partial \phi^2}\right) &= \frac{\nu(1 - \psi + \log \psi)}{\phi(1 - \phi)^2} + O(1), \\ E\left(\frac{\partial^2 \log L_1(\nu, \phi)}{\partial \phi \partial \nu}\right) &= \frac{(1 - \phi)\Lambda(\tau) + \log \psi}{\phi(1 - \phi)} + O(\nu^{-1}). \end{aligned}$$

As $\nu \rightarrow \infty$, the asymptotic variance for the resulting MLE of ν is given by

$$\sigma_{\text{tb}}^2 = \frac{\nu\phi(1 - \psi + \log \psi)}{A + B \log \psi}, \quad (8)$$

where $A = (1 - \phi)(\Lambda(\tau)^2(1 - \phi) - \phi P(\tau)(1 - \exp(\Lambda(\tau))) - \Lambda(\tau)P(\tau))$ and $B = \Lambda(\tau)(1 - 2\phi) - (1 - \phi)P(\tau) - \phi(1 - \exp(\Lambda(\tau)))$. Here the condition that one of the parameters tends to infinity ($\nu \rightarrow \infty$) is different from conventional asymptotic theory. The performance of the above asymptotic variance is examined in the simulation section.

3. Special models

3.1. Model \mathcal{M}_t

When $\phi = 1$, model $\mathcal{M}_{\text{tb}}^*$ reduces to model \mathcal{M}_t . The equation in (5) is eliminated and the other two equations become

$$\begin{aligned} d\Lambda(t) &= \frac{dN(t)}{\nu} \quad (0 \leq t \leq \tau), \\ \int_0^\tau \frac{1}{R(t)} dM(t) - \frac{N}{\nu} &= 0. \end{aligned} \quad (9)$$

Equation (9) was derived by Darroch (1958), Becker & Heyde (1990) and Yip *et al.* (1993). The profile likelihood $L(\nu, \hat{\Lambda})$ for model \mathcal{M}_t reduces to

$$\log L(\nu, \hat{\Lambda}) \propto \int_0^\tau \log R(t) dM(t) - N \log \nu,$$

which is identical to that given by Becker & Heyde (1990). If ν is treated as an integer, then the MLE can be obtained for the one-parameter case by taking a first finite difference of the above profile likelihood with respect to ν (Dahiya, 1981). It then follows that the MLE is the integer part of the solution satisfying

$$1 - \frac{M}{\nu} = \left(1 - \frac{1}{\nu}\right)^N \quad (10)$$

(see also Craig, 1953 and Seber, 1982 p. 136). When ν is large, the solutions for (9) and (10) are fairly close and both have the same asymptotic variance, given by

$$\sigma_t^2 = \frac{\nu}{\exp(\Lambda(\tau)) - 1 - \Lambda(\tau)}. \quad (11)$$

It is well known from Seber (1982 pp. 586–587) that (10) has a unique finite root greater than M if and only if $M < N$. However, under the same condition, the root of (9) in which ν is allowed to be any real number might be less than M . For example, in the case of $M = 3$, $N = 6$, the solution is 2.816. Following an argument similar to that given in Moek (1984), we can show that (9) has a unique root greater than $M - 1$ if and only if $M < N$. (When $M = N$, i.e. all animals are captured once, the profile likelihood is strictly increasing and thus no finite solution exists.) First note that

$$g(\nu) = \int_0^\tau \frac{1}{R(t)} dM(t) - \frac{N}{\nu} = \sum_{i=1}^M \frac{\nu(M - N) + (i - 1)N}{\nu(\nu - i + 1)M}.$$

As $\nu \rightarrow \infty$, we have $g(\nu) = (M - N)/(\nu M) + O(\nu^{-2})$. If $M < N$, then $g(\nu)$ increases to 0 when ν is large. Also, $g(\nu) \rightarrow \infty$ as ν approaches $M - 1$ from the right. Since $g(\nu)$ is continuous in $(M - 1, \infty)$, there exists at least one solution in this interval. To show uniqueness, notice that at any point crossing the horizontal axis such that $g(\nu) = 0$, we have

$$\begin{aligned} g'(\nu) &= - \int_0^\tau \frac{1}{R(t)^2} dM(t) + \frac{N}{\nu^2} \\ &= - \int_0^\tau \frac{1}{R(t)^2} dM(t) + \frac{1}{N} \left(\int_0^\tau \frac{1}{R(t)} dM(t) \right)^2 \\ &< - \int_0^\tau \frac{1}{R(t)^2} dM(t) + \frac{1}{M} \left(\int_0^\tau \frac{1}{R(t)} dM(t) \right)^2 < 0. \end{aligned}$$

The last inequality follows from the Cauchy–Schwarz inequality. Consequently, $g(\nu)$ must cross the horizontal axis from above and thus cannot cross more than once. This implies that there exists a unique solution in $(M - 1, \infty)$ if $M < N$. We remark that if the solution of (9) is restricted to be an integer, then the solution cannot be less than M , because the likelihood tends to zero as ν approaches $M - 1$. Therefore, the nearest integer to the solution with higher likelihood is at least M . The resulting integer must be identical to the integer part of the solution of (10).

From the behaviour of the function $g(\nu)$, it is seen that when $g(M) < 0$ the resulting population size estimate is less than M . For large M , this condition $g(M) < 0$ implies $M(\log M - 1) < N - M$, i.e. there are relatively more recaptures than first captures.

3.2. Model \mathcal{M}_b

In the special case that $\lambda(t) = \lambda$ for a first capture and $\lambda(t) = \phi\lambda$ for a recapture, model \mathcal{M}_{tb}^* reduces to model \mathcal{M}_b . For model \mathcal{M}_b , there are three parameters but no unknown time varying function is involved. The standard inference procedure is applicable. Under model \mathcal{M}_b , the log likelihood from (1) reduces to

$$\int_0^\tau \log R(t) dM(t) + N \log \lambda + (N - M) \log \phi - \lambda \int_0^\tau Q(t) dt.$$

Taking derivatives of the above likelihood with respect to ϕ , λ and ν , respectively, leads to

$$\frac{N - M}{\phi} - \lambda \int_0^\tau M(t-) dt = 0 \Leftrightarrow \phi\lambda = \frac{N - M}{\sum_{i=1}^M (\tau - t_{i1})}, \quad (12)$$

$$\frac{M}{\lambda} - \int_0^\tau R(t) dt = 0 \Leftrightarrow \lambda = \frac{M}{(\nu - M)\tau + \sum_{i=1}^M t_{i1}}, \quad (13)$$

$$\int_0^\tau \frac{1}{R(t)} dM(t) - \lambda\tau = 0. \quad (14)$$

The first equation is irrelevant in the estimation of ν because it is used only to estimate the recapture intensity $\phi\lambda$. If a first finite difference is taken with respect to ν instead of a derivative, then (14) is replaced by

$$\nu = \frac{M}{1 - \exp(-\lambda\tau)}. \quad (15)$$

Inverting a 3×3 expected Fisher information matrix, both equations for ν yield the following identical asymptotic variance

$$\sigma_b^2 = \frac{\nu(1 - \exp(-\lambda\tau))}{\exp(\lambda\tau) - 2 + \exp(-\lambda\tau) - \lambda^2\tau^2}. \quad (16)$$

Following a suggestion from a referee, we connect the present model to the discrete-time model \mathcal{M}_b . The latter model has also been discussed in the literature as a removal model. In a discrete-time framework, there are τ trapping occasions with corresponding time units $1, 2, \dots, \tau$. The parameters λ and $\phi\lambda$ are interpreted as the probability of first-capture and recapture on each occasion. The minimal sufficient statistics for (λ, ν) are $(M, \sum_{s=1}^\tau su_s)$, where u_s is the number of unmarked first-captures (or removals in a removal model) on occasion s ; see Otis *et al.* (1978 p.29). As shown by Moran (1951), Zippin (1956) and Seber & Whale (1970), the approximate MLE of (λ, ν) satisfies

$$\lambda = \frac{M}{(\nu - M)\tau + \sum_{s=1}^\tau su_s}, \quad (17)$$

$$\nu = \frac{M}{1 - (1 - \lambda)^\tau}. \quad (18)$$

In this discrete-time system, there are u_s animals with first-capture time s . The statistic $\sum_{s=1}^\tau su_s$ then corresponds to $\sum_{i=1}^M t_{i1}$ in a continuous-time set-up. Let \bar{t}_1 denote the average first-capture time, so that

$$\bar{t}_1 = \frac{1}{M} \sum_{i=1}^M t_{i1} = \frac{1}{M} \sum_{s=1}^\tau su_s.$$

Comparing (13)–(15) and (17)–(18), we find that the two systems of equations differ slightly only in the equation for ν . The chance of an animal being captured over τ discrete occasions is $1 - (1 - \lambda)^\tau$, while in a continuous-time interval it becomes $1 - \exp(-\lambda\tau)$.

Seber & Whale (1970) showed that in a discrete-time model the system of equations (17)–(18) has a unique solution $\hat{\lambda} > 0$ and $\hat{\nu} > M$ if and only if $\sum_{s=1}^\tau (\tau + 1 - 2s)u_s > 0$.

Chaiyapong & Lloyd (1997) and Lloyd & Chaiyapong (1999) interpreted this as a negative correlation between u_s and s . In other words, the number of first-captures should decrease with time because of the relationship $E(u_s) = \nu\lambda(1-\lambda)^{s-1}$. We now offer another interpretation so that a link between the conditions for the continuous-time model and discrete-time model can be established. Observe that $\sum_{s=1}^{\tau}(\tau+1-2s)u_s > 0$ can be rewritten as $\bar{t}_1 < \frac{1}{2}(\tau+1)$. That is, the average of the first-capture times should be less than half of the duration of the experiment. Here, the midpoint of the discrete times $\{1, 2, \dots, \tau\}$ is $\frac{1}{2}(\tau+1)$.

We consider the necessary and sufficient condition for the existence of a finite solution to (13) and (15). Substituting (15) into (13), we can show that $\hat{\lambda}$ is a root of

$$f(\lambda) = \frac{1}{\lambda} - \frac{\tau}{e^{\lambda\tau} - 1} - \bar{t}_1 = 0.$$

The function is strictly decreasing in $(0, \infty)$ because $f'(\lambda) < 0$ for $\lambda \in (0, \infty)$. Note that $\lim_{\lambda \rightarrow \infty} f(\lambda) = -\bar{t}_1 < 0$ and $\lim_{\lambda \rightarrow 0} f(\lambda) = \frac{1}{2}\tau - \bar{t}_1$. Therefore, a unique solution satisfying $\hat{\lambda} > 0$ and $\hat{\nu} > M$ exists if and only if $\bar{t}_1 < \frac{1}{2}\tau$, which also means that the average of the first-capture times is less than half of the experiment duration. Therefore, this provides a unified interpretation of the two sets of conditions. The following approximate probability of estimation breakdown is derived in the same way as that in Lloyd & Chaiyapong (1999) for a discrete-time model:

$$\Pr(\bar{t}_1 \geq \frac{1}{2}\tau) \approx \Phi\left(-\frac{\nu^{1/2}(2\alpha - (2+\alpha)(1-e^{-\alpha}))}{(4 - 4\alpha e^{-\alpha} - 3\alpha^2 e^{-\alpha} - (2+\alpha)^2 e^{-2\alpha})^{1/2}}\right),$$

where $\alpha = \lambda\tau$ and Φ denotes the standard normal cumulative distribution function. For a given fixed proportion of the population captured in the experiment, the failure probabilities for the continuous-time and discrete-time models are very close. For example, if $\nu = 500$ and 41.8% of the population is captured when the stopping time $\tau = 5$, then the failure probability is approximately 1.2%, whereas Lloyd & Chaiyapong (1999 Table 1) obtained an approximate value of 1.4% for a corresponding discrete-time model.

We now derive the condition for the existence of a finite solution to (13) and (14). For simplicity, write (13) as $\lambda = M/(\tau(\nu - a))$, where $a = \sum_{i=1}^M(\tau - t_{i1}) = M(1 - \bar{t}_1/\tau)$. After the substitution of (13) into (14), the MLE of ν is any $\nu > a$ (otherwise $\lambda < 0$) satisfying

$$h(\nu) = \sum_{i=1}^M \left(\frac{1}{\nu - i + 1} - \frac{1}{\nu - a} \right) = 0.$$

So the MLE of (ν, λ) depends on the value of M and the first-capture times, but it is independent of the value of ϕ .

As $\nu \rightarrow \infty$, we have $h(\nu) = \nu^{-2}M(\frac{1}{2}(M-1) - a) + O(\nu^{-3})$. Arguments similar to that discussed previously for the function $f(\nu)$ lead to the necessary and sufficient condition for the existence of a unique finite MLE such that $\hat{\lambda} > 0$ and $\hat{\nu} > M-1$ is $\frac{1}{2}(M-1) < a < M-1$. Here the condition $a < M-1$ is needed because if $a \geq M-1$, then $h(\nu) < 0$ for $\nu \in (a, \infty)$ and the MLE of ν does not exist.

The condition $\frac{1}{2}(M-1) < a < M-1$ is equivalent to $\tau/M < \bar{t}_1 < \frac{1}{2}\tau + \frac{1}{2}\tau/M$. That is, the average of the first-capture times should be restricted to some proper range. A population size estimate smaller than M occurs when $h(M) < 0$, which for large M means $\bar{t}_1 < \tau/(\log M + 1)$, i.e. the average of the first-capture times is relatively small in the restricted range.

3.3. Model \mathcal{M}_0

We discuss model \mathcal{M}_0 only briefly because of its limited practical use and its equivalence to model \mathcal{M}_t via a transformation, as discussed later. Letting $\phi = 1$ in the likelihood of model \mathcal{M}_b , we can show that the estimating equation for model \mathcal{M}_0 is exactly the same as (9), the equation for model \mathcal{M}_t . The asymptotic variance is given by $\sigma_0^2 = \nu / (\exp(\lambda\tau) - 1 - \lambda\tau)$. Comparing this variance with (11), we find these two formulas are identical if $\Lambda(\tau) = \lambda\tau$. Therefore, under model \mathcal{M}_0 , there is no loss of efficiency at all in allowing for time variation. This conclusion was first noted by Lloyd (1994) for the discrete-time models and we have shown it is valid for the continuous-time models as well.

The above finding can be intuitively explained by a proper transformation. Notice that in both models \mathcal{M}_t and \mathcal{M}_0 , only the statistics M and N are used in the estimation. Under model \mathcal{M}_t , consider a time transformation $t^* = \Lambda^{-1}(\lambda t)$. It is known (e.g. Ross, 1983 p.53) that the capture process $N^*(t) = N(t^*)$ is model \mathcal{M}_0 with constant intensity λ . Although the transformation is unknown because $\lambda(t)$ is unknown, the statistics M and N remain the same after the transformation. Therefore, the analyses for the two models are exactly the same on the two different time scales. Similarly, if $\lambda(t)$ is known, we can transform model \mathcal{M}_{tb}^* to \mathcal{M}_b ; the analysis for the complicated model \mathcal{M}_{tb}^* can be performed in the transformed process using model \mathcal{M}_b . However, for unknown $\lambda(t)$, the two models are not equivalent through a transformation because the first-capture times are used in the equations for model \mathcal{M}_b and those transformed capture times are unknown in the new time scale.

Lloyd (1994) also showed that under model \mathcal{M}_0 there is a large loss of efficiency in allowing for behavioural response for discrete-time models. Based on (8), (11) and (16), a similar conclusion holds and various efficiency comparisons can be made provided that these asymptotic results properly describe the variations of the estimators. It is interesting to note that all these asymptotic variances are functions of ν , $\Lambda(\tau)$ and/or ϕ , which makes the efficiency comparisons much simpler than those in discrete-time models; see Lloyd (1994). More variance comparisons when the underlying model is \mathcal{M}_{tb}^* are discussed using simulation results in Section 4.

4. Simulation study

We performed a simulation study to investigate the performance of the proposed estimators. The true population size was fixed at 400. We report the results for two types of time varying function $\lambda(t)$ (i.e. $1 + \sin 2\pi t$, $(t + 0.5)^{-1}$), and two values of ϕ (0.75, 1.5). See Hwang (1997) for other types of functions and more simulation results. Tables 1 and 2 present the simulation results for $\phi = 0.75$ and 1.5, respectively. For each combination of $\lambda(t)$ and ϕ , five stopping times were chosen: $\tau = 1, 1.25, 1.5, 2$ and 4. Then for each fixed stopping time, 500 datasets were generated. Based on each generated dataset, we calculated iteratively the following three estimators and their estimated standard errors:

$\hat{\nu}_t$: the MLE of ν under model \mathcal{M}_t , see (9).

$\hat{\nu}_b$: the MLE of ν under model \mathcal{M}_b , see (12)–(14).

$\hat{\nu}_{tb}$: the MLE of ν under model \mathcal{M}_{tb}^* , see (3)–(5).

The standard error (se) for each estimator was obtained by inverting the Fisher information matrix under the corresponding model. The iterative steps for computing estimates diverged in a negligible number of trials. Here ‘divergence’ means that either the steps increased without a limit (the upper bound was set to be $10M$) or the iterations failed to converge to a stable

TABLE 1

Comparison of various estimators (est.) of population size for model \mathcal{M}_{tb}^*
 (500 trials, $\phi = 0.75$, $\nu = 400$, $M =$ number of distinct animals captured in the experiment)

Stopping time	Estimator	Average est.	sd est.	Average se	rms est.	95% CI coverage
$\lambda(t) = 1 + \sin 2\pi t$						
$\tau = 1, M = 252$	$\hat{\nu}_t$	468	32.7	34.4	75.5	36.0
	$\hat{\nu}_b$	264	11.0	5.1	136.2	0.0
	$\hat{\nu}_{tb}$	465	268.5	319.5	276.1	92.6
$\tau = 1.25, M = 303$	$\hat{\nu}_t$	454	21.3	22.3	58.0	16.4
	$\hat{\nu}_b$	344	15.5	12.7	58.2	14.8
	$\hat{\nu}_{tb}$	413	83.4	66.5	84.2	94.0
$\tau = 1.5, M = 335$	$\hat{\nu}_t$	440	13.2	15.6	42.0	12.4
	$\hat{\nu}_b$	383	15.6	14.1	22.7	80.8
	$\hat{\nu}_{tb}$	401	31.6	29.2	31.6	93.8
$\tau = 2, M = 346$	$\hat{\nu}_t$	436	12.0	13.7	38.0	10.0
	$\hat{\nu}_b$	361	8.4	5.6	40.0	1.2
	$\hat{\nu}_{tb}$	403	27.0	23.7	27.2	91.4
$\tau = 4, M = 393$	$\hat{\nu}_t$	410	3.3	4.2	10.1	15.2
	$\hat{\nu}_b$	395	2.8	1.9	5.6	76.0
	$\hat{\nu}_{tb}$	399	3.7	3.5	3.7	91.8
$\lambda(t) = (t + 0.5)^{-1}$						
$\tau = 4, M = 266$	$\hat{\nu}_t$	463	28.6	30.6	69.4	31.0
	$\hat{\nu}_b$	295	14.2	10.0	105.5	0.2
	$\hat{\nu}_{tb}$	454	230.6	236.6	236.6	92.4
$\tau = 1.25, M = 286$	$\hat{\nu}_t$	459	24.5	25.9	64.3	21.4
	$\hat{\nu}_b$	308	11.2	7.5	92.8	0.0
	$\hat{\nu}_{tb}$	430	126.5	112.1	129.8	95.0
$\tau = 1.5, M = 299$	$\hat{\nu}_t$	453	21.9	22.7	57.6	21.6
	$\hat{\nu}_b$	315	10.1	5.8	85.6	0.0
	$\hat{\nu}_{tb}$	414	81.0	69.7	82.0	94.6
$\tau = 2, M = 320$	$\hat{\nu}_t$	446	17.3	18.5	49.0	17.0
	$\hat{\nu}_b$	329	8.8	4.0	71.6	0.0
	$\hat{\nu}_{tb}$	405	51.2	42.2	51.4	93.4
$\tau = 4, M = 355$	$\hat{\nu}_t$	432	10.8	11.9	33.3	11.4
	$\hat{\nu}_b$	357	6.3	1.5	43.7	0.0
	$\hat{\nu}_{tb}$	402	18.5	18.3	18.6	94.0

value in 100 steps. The associated 95% confidence intervals were calculated using a log transformation presented in Chao (1987). The resulting 500 estimates and standard errors were averaged to give the results under the headings ‘Average est.’ and ‘Average se’ in Tables 1 and 2. The sample standard deviation and the sample root mean squared error (RMSE) for the 500 estimates were also computed; see ‘sd est.’ and ‘rms est.’ in the tables. The percentage of trials in which the 95% confidence intervals covered the true parameter value is also shown, in the last column of each table, and we also give the average of the number of distinct animals captured in the experiment (M).

When the underlying model is \mathcal{M}_{tb}^* , the MLE $\hat{\nu}_t$ is positively biased in the trap-shy cases (Table 1) and negatively biased in the trap-happy cases (Table 2). This is a typical performance for an estimator that wrongly ignores the behaviour effect; see Otis *et al.* (1978).

TABLE 2

Comparison of various estimators (est.) of population size for model \mathcal{M}_{tb}^*
(500 trials, $\phi = 1.5$, $\nu = 400$, $M =$ number of distinct animals captured in the experiment)

Stopping time	Estimator	Average est.	sd est.	Average se	rms est.	95% CI coverage
$\lambda(t) = 1 + \sin 2\pi t$						
$\tau = 1, M = 252$	$\hat{\nu}_t$	333	15.7	14.2	69.1	3.4
	$\hat{\nu}_b$	265	10.3	5.1	135.9	0.0
	$\hat{\nu}_{tb}$	424	129.5	136.2	131.5	94.6
$\tau = 1.25, M = 302$	$\hat{\nu}_t$	350	11.0	9.3	51.1	0.8
	$\hat{\nu}_b$	343	16.0	12.5	59.1	12.8
	$\hat{\nu}_{tb}$	405	53.4	47.9	53.6	93.0
$\tau = 1.5, M = 335$	$\hat{\nu}_t$	363	8.6	6.7	37.5	0.4
	$\hat{\nu}_b$	385	17.1	14.4	22.7	80.6
	$\hat{\nu}_{tb}$	402	26.5	25.2	26.5	95.2
$\tau = 2, M = 346$	$\hat{\nu}_t$	368	7.7	5.8	32.8	2.0
	$\hat{\nu}_b$	361	8.3	5.6	39.8	1.6
	$\hat{\nu}_{tb}$	400	20.5	19.6	20.5	94.2
$\tau = 4, M = 393$	$\hat{\nu}_t$	394	2.8	1.4	7.1	56.6
	$\hat{\nu}_b$	395	2.9	1.9	5.8	76.8
	$\hat{\nu}_{tb}$	399	3.6	3.4	3.6	92.8
$\lambda(t) = (t + 0.5)^{-1}$						
$\tau = 1, M = 267$	$\hat{\nu}_t$	338	13.4	12.7	63.3	1.8
	$\hat{\nu}_b$	297	13.3	10.1	103.5	0.2
	$\hat{\nu}_{tb}$	427	129.0	115.2	131.7	95.2
$\tau = 1.25, M = 286$	$\hat{\nu}_t$	344	12.8	10.9	57.0	1.8
	$\hat{\nu}_b$	307	11.3	7.4	93.8	0.0
	$\hat{\nu}_{tb}$	413	72.0	68.7	73.2	94.4
$\tau = 1.5, M = 300$	$\hat{\nu}_t$	350	11.3	9.6	51.4	0.8
	$\hat{\nu}_b$	316	10.1	5.9	84.4	0.0
	$\hat{\nu}_{tb}$	405	50.4	49.3	50.6	95.4
$\tau = 2, M = 320$	$\hat{\nu}_t$	357	9.5	7.9	43.7	1.6
	$\hat{\nu}_b$	329	8.5	4.0	71.4	0.0
	$\hat{\nu}_{tb}$	405	38.5	35.1	38.8	95.2
$\tau = 4, M = 355$	$\hat{\nu}_t$	373	6.9	5.0	28.3	1.8
	$\hat{\nu}_b$	357	6.4	1.5	43.7	0.0
	$\hat{\nu}_{tb}$	401	16.8	16.2	16.8	93.2

The direction of the bias for the estimator $\hat{\nu}_b$, which does not take the time-varying effects into account, depends on the behaviour of the time-varying functions $\lambda(t)$. In the two reported cases, the bias is negative, but for other increasing functions $\lambda(t)$ (not reported here), the bias becomes positive. In both directions, the bias is generally very large. As discussed in Section 3, if $\lambda(t)$ is known, we can transform model \mathcal{M}_{tb}^* to \mathcal{M}_b by a proper transformation of the capture process. This implies that the model assumption for the estimator $\hat{\nu}_b$ is upheld in the transformed time scale but not in the original one. Using a wrong time scale in the analysis results in serious bias for the population size estimator $\hat{\nu}_b$. Almost none of the associated confidence intervals cover the true parameter (the estimated coverage probabilities for $\hat{\nu}_b$ are nearly 0). See Hwang (1997) for more simulation results.

Although both the estimators $\hat{\nu}_t$ and $\hat{\nu}_b$, which do not consider both behavioural response and time variation, exhibit less variation, their biases are substantial. At smaller stopping

times, either $\hat{\nu}_t$ or $\hat{\nu}_b$ of the associated 95% confidence intervals are much lower than the nominal level. The criterion of RMSE has the danger of selecting precise but seriously biased estimates, especially when the mean capture rate is low or capture information is insufficient. Thus neither $\hat{\nu}_t$ nor $\hat{\nu}_b$ can be used under model \mathcal{M}_{tb}^* .

The proposed MLE $\hat{\nu}_{tb}$ based on (3)–(5) has the smallest bias in all cases considered. When the stopping time is relatively short, so that the data provide little information about the population size, the estimator $\hat{\nu}_{tb}$ behaves unstably. When the stopping time is increased and more data are available as in the trap-happy cases, the proposed estimator outperforms the others with respect to both bias and RMSE, although it unavoidably has larger variation than the other two simpler estimators. The standard method, inverting the Fisher information matrix, produces reasonable standard errors for the trap-happy models, but it works less well in some cases for the trap-shy models as the amount of recapture information decreases. The associated 95% confidence intervals perform well in the interval coverage probability. Most of the estimated coverage probabilities for $\hat{\nu}_{tb}$ differ little from the nominal rate.

In our estimation procedure, we have assumed that the ratios of the recapture to first-capture probabilities are a constant. To investigate the robustness of the proposed procedure to this assumption, we also considered trials where the proportion constant is a random variable from a normal distribution. Simulation results (not reported here) show that the relative merits of the three estimators remain the same as in Tables 1 and 2 if ϕ is a random variable from a normal distribution with mean 0.75 (sd 0.2) and mean 1.5 (sd 0.2), respectively. When a sufficiently large proportion of the population is captured, the estimates $\hat{\nu}_{tb}$ are comparable to those in Tables 1 and 2. Thus our proposed likelihood approach shows some degree of robustness with respect to the proportional constant assumption.

References

- AALEN, O. (1978). Nonparametric inference for a family of counting processes. *Ann. Statist.* **6**, 701–726.
- ANDERSEN, P.K., BORGAN, Ø., GILL, R.D. & KEIDING, N. (1993). *Statistical Models Based on Counting Processes*. New York: Springer-Verlag.
- BECKER, N.G. (1984). Estimating population size from capture–recapture experiments in continuous time. *Austral. J. Statist.* **26**, 1–7.
- BECKER, N.G. & HEYDE, C.C. (1990). Estimating population size from multiple recapture experiments. *Stochastic Process. Appl.* **36**, 77–83.
- CHAIYAPONG, Y. & LLOYD, C.J. (1997). Accurate inference for recapture experiments with behavioural response. *J. Statist. Comput. Simulation* **56**, 97–115.
- CHAO, A. (1987). Estimating the population size for capture–recapture data with unequal catchability. *Biometrics* **43**, 783–791.
- CHAO, A. & LEE, S-M. (1993). Estimating population size for continuous-time capture–recapture models via sample coverage. *Biometrical J.* **35**, 29–45.
- CHAO, A., CHU, W. & HSU, C-H. (2000). Capture–recapture when time and behavioral response affect capture probabilities. *Biometrics* **56**, 427–433.
- CRAIG, C.C. (1953). On the utilization of marked specimens in estimating populations of flying insects. *Biometrika* **40**, 170–176.
- CROWDER, M.J., KIMBER, A.C., SMITH, R.L. & SWEETING, T.J. (1991). *Statistical Analysis of Reliability Data*. London: Chapman and Hall.
- DAHIYA, R.C. (1981). An improved method of estimating an integer-parameter by maximum likelihood. *Amer. Statist.* **35**, 34–37.
- DARROCH, J.N. (1958). The multiple-recapture census, I. Estimation of a closed population. *Biometrika* **45**, 343–359.
- DURAN, J.W. & WIORKOWSKI, J.J. (1981). Capture–recapture sampling for estimating software content. *IEEE Trans. Software Eng.* **SE-7**, 147–148.

- GODAMBE, V.P. (1985). The foundations of finite sample estimation in stochastic processes. *Biometrika* **72**, 419–428.
- HEYDE, C.C. (1997). *Quasi-Likelihood and Its Application: A General Approach to Optimal Parameter Estimation*. New York: Springer-Verlag.
- HWANG, W.-H. (1997). Estimating Population Size for Continuous Time Capture–Recapture Experiments. PhD Dissertation, Institute of Statistics, National Tsing Hua University, Taiwan.
- JACOBSEN, M. (1984). Maximum likelihood estimation in the multiplicative intensity models: a survey. *Internat. Statist. Rev.* **52**, 193–207.
- JOHANSEN, S. (1983). An extension of Cox’s regression model. *Internat. Statist. Rev.* **51**, 165–174.
- KIEFER, J. & WOLFOWITZ, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many nuisance parameters. *Ann. Math. Statist.* **27**, 887–906.
- LIN, D.Y. & YIP, P. (1999). Parametric regression models for continuous time removal and recapture studies. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **61**, 401–411.
- LLOYD, C.J. (1994). Efficiency of martingale methods in recapture studies. *Biometrika* **81**, 305–315.
- LLOYD, C.J. & CHAIYAPONG, Y. (1999). Choosing sampling effort for recapture experiments with behavioural response. *Aust. N. Z. J. Stat.* **41**, 299–304.
- MOEK, G. (1984). Comparison of some software reliability models for simulated and real failure data. *Internat. J. Model. Simulation* **4**, 29–41.
- MORAN, P.A.P. (1951). A mathematical theory of animal trapping. *Biometrika* **38**, 307–311.
- NAYAK, T.K. (1988). Estimating population size by recapture sampling. *Biometrika* **75**, 113–120.
- NELSON, W. (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics* **14**, 945–965.
- OTIS, D.L., BURNHAM, K.P., WHITE, G.C. & ANDERSON, D.R. (1978). Statistical inference from capture data on closed animal populations. *Wildl. Monog.* **62**, 1–135.
- ROSS, S.M. (1983). *Stochastic Processes*. New York: Wiley.
- SCHOLZ, F.W. (1980). Towards a unified definition of maximum likelihood. *Canad. J. Statist.* **8**, 193–203.
- SEBER, G.A.F. (1982). *The Estimation of Animal Abundance*, 2nd edn. London: Griffin.
- SEBER, G.A.F. & WHALE, J.F. (1970). The removal method for two and three samples. *Biometrics* **26**, 393–400.
- VOAS, J.M. & MCGRAW, G. (1998). *Software Fault Injection: Inoculating Programs Against Errors*. New York: Wiley.
- WILSON, K.R. & ANDERSON, D.R. (1995). Continuous-time capture–recapture population estimation when capture probabilities vary over time. *Environ. Ecol. Statist.* **2**, 55–69.
- YIP, P. (1995). Estimating the number of errors in a system, using a martingale approach. *IEEE Trans. Reliab.* **44**, 322–326.
- YIP, P. & CHAO, A. (1996). Estimating population size from capture–recapture studies via sample coverage and estimating functions. *Commun. Statist. Stochastic Models* **12**, 17–35.
- YIP, P.S.F., FONG, D.Y.T. & WILSON, K. (1993). Estimating population size by recapture sampling via estimating function. *Comm. Statist. Stochastic Models* **9**, 179–193.
- YIP, P.S.F., HUGGINS, R.M. & LIN, D.Y. (1996). Inference for capture–recapture experiments in continuous time with variable capture rates. *Biometrika* **83**, 477–483.
- YIP, P.S.F., XI, L., CHAO, A. & HWANG, W.-H. (2000). Estimating the population size with a behavioural response in a capture–recapture experiment. *Environ. Ecol. Statist.* **7**, 405–414.
- ZIPPIN, C. (1956). An evaluation of the removal method of estimating animal populations. *Biometrics* **12**, 163–189.