ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

Journal of the Royal Statistical Society
Statistical Methodology
Series B

J. R. Statist. Soc. B (2018)
80, *Part* 5, *pp.* 995–1014

Full likelihood inference for abundance from continuous time capture–recapture data

Yang Liu and Yukun Liu,

East China Normal University, Shanghai, People's Republic of China

Pengfei Li

University of Waterloo, Canada

and Jing Qin

National Institute of Allergy and Infectious Diseases, Bethesda, USA

[Received May 2017. Final revision June 2018]

Summary. Capture–recapture experiments are widely used cost-effective sampling techniques for estimating population sizes or abundances in biology, ecology, demography, epidemiology and reliability studies. For continuous time capture–recapture data, existing estimation methods are based on conditional likelihoods and an inverse weighting estimating equation. The corresponding Wald-type confidence intervals for the abundance may have severe undercoverage, and their lower limits can be below the number of individuals captured. We propose a full likelihood inference approach by combining a parametric or partial likelihood with the empirical likelihood. Under both parametric and semiparametric intensity models, we demonstrate that the maximum likelihood estimator attains the semiparametric efficiency lower bound and that the full likelihood ratio statistic for the abundance is asymptotically χ^2 with 1 degree of freedom. Simulations indicate that compared with conditional-likelihood-based methods, the maximum full likelihood estimator has a smaller mean-square error, and the likelihood ratio confidence intervals often have remarkable gains in coverage probability. We illustrate the advantages of the proposed approach by analysing illegal immigrant data for the Netherlands and *Prinia flaviventris* data from Hong Kong.

Keywords: Abundance; Andersen–Gill model; Capture–recapture experiment; Conditional likelihood; Empirical likelihood

1. Introduction

Capture–recapture experiments are widely used cost-effective sampling techniques for estimating population sizes or abundances (Otis *et al.*, 1978), which are of fundamental importance in biology, ecology, demography, epidemiology and reliability studies (Pollock, 1991, 2000; Chao *et al.*, 2001; Borchers *et al.*, 2002, 2015). Examples of abundances include the sizes of animal populations in fisheries and wildlife biology, the number of faults in reliability testing and the frequencies of diseases in epidemiological studies. The population of interest in this paper is assumed to be closed—in other words, there is no birth, death or migration—and hence the abundance remains unchanged during the sampling experiment.

Address for correspondence: Yukun Liu, School of Statistics, East China Normal University, NO 500 Dongchuan Road, Shanghai 200241, People's Republic of China.
E-mail: ykliu @sfs.ecnu.edu.cn

 1369–7412/18/80995

A capture–recapture experiment normally consists of a number of occasions when individuals from a population are captured. They are marked, or their existing marks are noted, and then released back into the population. For each captured individual a capture history is recorded. According to whether the captures occur on a limited number of occasions or continuously, these experiments can be divided into discrete time and continuous time capture–recapture experiments. There has been extensive research into discrete time capture–recapture data. See, for example, Seber (1982), Chao (1987), Huggins (1989), Alho (1990), Chen and Lloyd (2000, 2002), Fewster and Jupp (2009), Stoklosa *et al.* (2011), Liu *et al.* (2017) and the references therein.

We focus on continuous time capture–recapture experiments. In a continuous time experiment, only one animal is caught at each trapping occasion. In addition to the marking process, we also record the exact capture times for each animal. Thus, any capture is regarded as a trapping occasion, and the exact time of each occasion is recorded. Such experiments are often used in studies of insects, sperm whales, grizzly bears and other large mammals (Wilson and Anderson, 1995). Earlier work on the estimation of abundance based on continuous time capture–recapture data includes Craig (1953) and Darroch (1958), who dealt with a homogeneous population. Becker (1984) first established a martingale-based approach for continuous time experiments. Becker and Heyde (1990) and Yip *et al.* (1993, 2000) subsequently developed a class of high efficiency martingale-based estimators, which are the solutions to a certain set of martingale-based estimating equations. Chao and Lee (1993) and Yip and Chao (1996) proposed new abundance estimators by using sample coverage and estimating function approaches.

As is well known, heterogeneity is almost always present in capture–recapture experiments. Failure to account for this heterogeneity may cause substantial bias (Otis *et al.*, 1978; Burnham and Overton, 1978; Chao, 1987). These works are free of covariates, but a better way to account for heterogeneity is to model the capture process via covariates. For continuous time capture–recapture data, the most widely used method is the Andersen–Gill intensity model (Andersen and Gill, 1982), which is a non-homogeneous Poisson process with its intensity function depending on the covariates. Yip *et al.* (1996) used a partial likelihood under this model to estimate the unknown parameters and employed the Horvitz–Thompson estimator to estimate the abundance. Lin and Yip (1999) and Hwang and Chao (2002) proposed a score-function-based estimating function approach for abundance estimation. Chen (2001) suggested a likelihood-based method and showed that his estimator achieves the semiparametric efficiency lower bound. Recent developments on continuous time capture–recapture data include considerations of measurement errors (Hwang and Huang, 2003; Yip *et al.*, 2005) and a frailty model (Xi *et al.*, 2007; Xu *et al.*, 2007).

The existing estimation methods for abundance are largely based on the conditional likelihood (CL) (Huggins and Hwang, 2011) and generally consist of two steps. In the first step, a desirable point estimator for the abundance is derived under some probability model on the capture process with or without covariates. The Horvitz–Thompson estimator, which is also known as the inverse probability weighting estimator, is usually used for this. In the second step, the abundance estimator is shown to have asymptotic normality, and a consistent estimator for its asymptotic variance is prepared. Wald-type confidence intervals are then constructed for the abundance. However, even in the simplest case, the small sample distribution of the abundance estimator is strongly skewed to the right (Evans and Bonett, 1994), which may lead to severe undercoverage of the corresponding Wald-type confidence interval. Also its lower limit can be below the number of individuals that are captured. Similar observations have been made in our simulation studies. The necessary estimation of an asymptotic variance may inflate the variation of Wald-type confidence intervals. Further, it has been widely recognized that Horvitz and Thompson's (1952) inverse probability weighting approach might not be stable since some weights might be quite small, so certain individuals become unduly influential and the final abundance estimates are too large. These shortcomings motivate our work.

In this paper, we use the empirical likelihood (EL) (Owen, 1988, 1990) to construct a novel approach for the estimation of abundance in continuous time capture–recapture experiments. As a non-parametric counterpart of the parametric likelihood, the EL has many nice properties. For example, EL confidence regions are Bartlett correctable (DiCiccio *et al.*, 1991), range preserving, transformation respecting (Hall and La Scala, 1990) and free of variance estimation. See Owen (2001) and Newey and Smith (2004) for a thorough review.

Instead of conventional CL approaches, we develop a full likelihood set-up under the Andersen–Gill intensity model, where the capture–recapture process is a Poisson process with a covariate-dependent intensity function. When the capture–recapture data are modelled by mixture models other than non-homogeneous Poisson processes, the relationship between the conditional and full likelihoods has been extensively studied (Farcomeni and Tardella, 2012; Holzmann *et al.*, 2006; Link, 2003). However, when the capture–recapture process is a Poisson process, this relationship has not been explored. Our paper fills this gap. The full likelihood is composed of three parts. The first part is a binomial likelihood, the second part is a conditional parametric likelihood or partial likelihood and the third part is the marginal EL constructed from the covariate information. The details can be found in Section 2.2. Under both parametric and semiparametric intensity model assumptions, we establish the asymptotic normality and semiparametric efficiency of the maximum likelihood abundance estimator and show that the full likelihood ratio test statistic follows a χ^2 limiting distribution with 1 degree of freedom. When used to construct confidence intervals for the abundance, the full likelihood method has two obvious advantages over the CL methods that were discussed in Chen (2001). First, the full likelihood approach is one step and free of variance estimation. Second, the lower limit of the confidence interval that is derived from the full likelihood is always no less than the number of individuals captured. Compared with Chen's (2001) method, our simulation results indicate that the maximum full likelihood abundance estimator is more accurate in terms of mean-square error (MSE), and that the proposed full-likelihood-based confidence intervals often have remarkable gains in coverage probability. As a by-product, the approach proposed produces a consistent estimator for the marginal covariate distribution although the observed covariates are subject to biased sampling.

The rest of the paper is organized as follows. In Section 2, we introduce the Andersen–Gill intensity model, the EL and the profile EL. We also point out the close relationship between the EL and Chen's (2001) full likelihood. In Section 3, we present the maximum EL estimators and EL ratio functions, and we study their asymptotic distributions under both parametric and semiparametric intensity models. Section 4 presents a simulation study. In Section 5, we illustrate the proposed EL method by analysing illegal immigrant data for the Netherlands and *Prinia flaviventris* data from Hong Kong. A short discussion is given in Section 6. For convenience of presentation, we defer to the on-line supplementary document the technical details, some additional simulation results, the algorithms for the calculation of the method proposed, the goodness-of-fit tests and model selection for the underlying parametric and semiparametric models, and a study of the influence of additional heterogeneity via a frailty model.

The data that are analysed in the paper and the programs that were used to analyse them can be obtained from

`http://wileyonlinelibrary.com/journal/rss-datasets`

2. Full empirical likelihood

2.1. Model and data

Denote by ν the abundance of the closed population of interest. Suppose that a continuous time capture–recapture experiment with duration time $[0,\tau]$ is conducted to sample individuals from this population. Since the period of the experiment is relatively short in general, it is reasonable to assume that the covariates are time independent. Let Z^* be the time-independent covariates and $N^*(t)$, $t \in [0,\tau]$, the number of captures up to time t for a subject in the population. Denote the conditional intensity function of the counting process N^* at time t given $Z^* = z$ by $\lambda(t|z)$, i.e.

$$\lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} P\{N^*(t+\Delta t) - N^*(t) = 1 | Z^* = z, N^*(s), s \leqslant t\} = \lambda(t|z). \tag{1}$$

This implies that N^* is a non-homogeneous Poisson process with intensity $\lambda(t|z)$. The form of $\lambda(t|z)$ is assumed to be known for the time being and will be modelled by parametric and semiparametric models in the subsequent sections.

Let $D^* = I\{N^*(\tau) > 0\}$ be the indicator of a generic subject captured at least once before τ, and let

$$\pi(z) = P(D^* = 1 | Z^* = z) = 1 - \exp\left\{ -\int_0^\tau \lambda(t|z) \mathrm{d}t \right\} \tag{2}$$

and $\alpha = P(D^* = 0)$ be respectively the conditional probability of a subject being captured at least once given the covariate and the unconditional probability of never being captured. Let (N, Z) follow the conditional distribution of (N^*, Z^*) given $D^* = 1$, and let n be the random number of subjects captured throughout the study. For $i = 1, 2, \ldots, n$, denote by (N_i, Z_i) the analogue of (N, Z) for subject i being captured. Then, given n captured subjects, (N_i, Z_i), $i = 1, 2, \ldots, n$, are conditionally independent with the same distribution as (N, Z).

2.2. Empirical likelihood

The likelihood based on the full data $(n, N_1, \ldots, N_n, Z_1, \ldots, Z_n)$ consists of three parts

$$\tilde{L} = P(n)\, P(Z_1, \ldots, Z_n | n)\, P(N_1, \ldots, N_n | n, Z_1, \ldots, Z_n),$$

where

$$P(n) = \binom{\nu}{n} (1-\alpha)^n \alpha^{\nu - n} \tag{3}$$

represents the binomial probability of observing n subjects. Denote by F_Z and F_{Z^*} the distribution functions of Z and Z^* respectively. We see that $\mathrm{d}F_Z(z) = \pi(z)\mathrm{d}F_{Z^*}(z)/(1-\alpha)$, and

$$P(Z_1, \ldots, Z_n | n) = \prod_{i=1}^n \mathrm{d}F_Z(Z_i) = \prod_{i=1}^n \frac{\pi(Z_i)\mathrm{d}F_{Z^*}(Z_i)}{1-\alpha} \tag{4}$$

represents the probability of observing the covariates Z_i, given n subjects being captured.

It follows (see formula (2.7.4′) of Andersen *et al.* (1993)) that the conditional density of $N(\cdot)$ given $Z = z$ is

$$P\{N(\cdot) | Z = z\} = \frac{\exp\left\{ -\int_0^\tau \lambda(t|z) \mathrm{d}t \right\}}{\pi(z)} \exp\left[\int_0^\tau \log\{\lambda(t|z)\}\, \mathrm{d}N(t) \right].$$

Accordingly, given that n subjects are captured and given their covariates $\{Z_i : i=1,2,\ldots,n\}$, the conditional joint distribution of $\{N_i(\cdot), i=1,2,\ldots,n\}$ is

$$P(N_1,\ldots,N_n|n,Z_1,\ldots,Z_n)=\prod_{i=1}^{n} P\{N_i(\cdot)|Z_i\}$$
$$=\prod_{i=1}^{n}\frac{\exp\left\{-\int_0^\tau \lambda(t|Z_i)\mathrm{d}t\right\}}{\pi(Z_i)}\exp\left[\int_0^\tau \log\{\lambda(t|Z_i)\}\,\mathrm{d}N_i(t)\right], \qquad (5)$$

which is a CL, denoted by L^{c}.

Combining equations (3), (4) and (5) and taking the logarithm, we arrive at the log-likelihood

$$\tilde{l}=\log(\tilde{L})=\log\binom{\nu}{n}+(\nu-n)\log(\alpha)+\sum_{i=1}^{n}\left[\log\{\mathrm{d}F_{Z^*}(Z_i)\}-\int_0^\tau \lambda(t|Z_i)\mathrm{d}t\right.$$
$$\left.+\int_0^\tau \log\{\lambda(t|Z_i)\}\mathrm{d}N_i(t)\right].$$

We note here that α, $\lambda(t|z)$ and $F_{Z^*}(z)$ are the only unknowns, and they satisfy

$$\alpha=\int\exp\left\{-\int_0^\tau \lambda(t|z)\mathrm{d}t\right\}\mathrm{d}F_{Z^*}(z), \qquad (6)$$

which follows from $\alpha=\int\{1-\pi(z)\}\mathrm{d}F_{Z^*}(z)$ and equation (2).

2.3. *Profile empirical likelihood*

According to the principle of the EL (Owen, 1988, 1990), we need to consider only the distributions F_{Z^*} having supports on $\{Z_1,Z_2,\ldots,Z_n\}$. Let $p_i=\mathrm{d}F_{Z^*}(Z_i)$ $(i=1,2,\ldots,n)$ such that $p_i\geqslant 0$ and $\Sigma_{i=1}^{n} p_i=1$. Maximizing $\tilde{l}$ with respect to the p_is under constraint (6) leads to

$$p_i=\frac{1}{n}\frac{1}{1+\xi\left[\exp\left\{-\int_0^\tau \lambda(t|Z_i)\mathrm{d}t\right\}-\alpha\right]},$$

where the Lagrange multiplier ξ, which is an implicit function of α and $\lambda(\cdot)$, is the solution to

$$\sum_{i=1}^{n}\frac{\exp\left\{-\int_0^\tau \lambda(t|Z_i)\mathrm{d}t\right\}-\alpha}{1+\xi\left[\exp\left\{-\int_0^\tau \lambda(t|Z_i)\mathrm{d}t\right\}-\alpha\right]}=0. \qquad (7)$$

Thus, the profile empirical log-likelihood function of $(\nu,\alpha,\lambda(\cdot))$ is

$$l\{\nu,\alpha,\lambda(\cdot)\}=\log\binom{\nu}{n}+(\nu-n)\log(\alpha)+\sum_{i=1}^{n}\left[-\int_0^\tau \lambda(t|Z_i)\,\mathrm{d}t+\int_0^\tau \log\{\lambda(t|Z_i)\}\,\mathrm{d}N_i(t)\right]$$
$$-\sum_{i=1}^{n}\log\left(1+\xi\left[\exp\left\{-\int_0^\tau \lambda(t|Z_i)\mathrm{d}t\right\}-\alpha\right]\right), \qquad (8)$$

which is the foundation of our subsequent statistical inference.

Our EL has a close relationship with Chen's full likelihood $L(\nu,\theta,f_Z)$. It can be verified that $p_i=\mathrm{d}F_{Z^*}(Z_i)=q_i(1-\alpha)/\pi(Z_i)$, where $q_i=\mathrm{d}F_Z(Z_i)$ is defined in Chen (2001), and equation (6) is equivalent to $1=\Sigma_{i=1}^{n} q_i(1-\alpha)/\pi(Z_i)$. In terms of the q_is, the empirical log-likelihood function becomes

$$\tilde{l}=\log\binom{\nu}{n}+(\nu-n)\log(\alpha)+n\log(1-\alpha)+\sum_{i=1}^{n}\log(q_i)+\log(L^{c}),$$

where L^{c} is defined in equation (5). By noting that $1-\alpha$ and $\log(L^{c})$ are equal to Chen's p and $l_n(\theta)$ respectively, we can show that $\tilde{l}$ is exactly his full likelihood $\log\{L(\nu,\theta,f_Z)\}$. Although Chen (2001) proposed a maximum likelihood estimation procedure for the abundance based on this full likelihood, his recommended estimation procedure is based on the conditional likelihood L^{c} or his $l_n(\theta)$. In this paper, we instead propose to perform inference for the abundance based directly on the profile full EL.

3. Estimation and asymptotics

3.1. Parametric intensity model

It is natural to postulate a parametric model $\lambda(t|z)=\lambda(t,z,\beta)$ on the intensity function, where $\lambda(t,z,\beta)$ is known up to a parameter β. Simple examples of $\lambda(t,z,\beta)$ include $\exp(z^{T}\beta)$ and $t\exp(z^{T}\beta)$. Parametric intensity models have been used by Lin and Yip (1999) to propose a martingale-based estimation function approach to abundance estimation.

Under the parametric intensity model $\lambda(t|z)=\lambda(t,z,\beta)$, let $l_p(\nu,\alpha,\beta)$ denote the profile empirical log-likelihood function, which is the same as equation (8) with $\lambda(t,z,\beta)$ in place of $\lambda(t|z)$. We use the subscript p to highlight the *parametric* intensity model. Denote the maximum likelihood estimators by

$$(\hat{\nu}_p,\hat{\alpha}_p,\hat{\beta}_p)=\underset{(\nu,\alpha,\beta)}{\arg\max}\, l_p(\nu,\alpha,\beta). \tag{9}$$

We propose to estimate the abundance ν by $\hat{\nu}_p$ and to estimate $\lambda(t|z)$ by $\lambda(t,z,\hat{\beta}_p)$. The empirical log-likelihood ratio functions of (ν,α,β) and ν are defined as

$$R_p(\nu,\alpha,\beta)=2\{l_p(\hat{\nu}_p,\hat{\alpha}_p,\hat{\beta}_p)-l_p(\nu,\alpha,\beta)\},$$

$$R'_p(\nu)=2\{l_p(\hat{\nu}_p,\hat{\alpha}_p,\hat{\beta}_p)-\sup_{\alpha,\beta} l_p(\nu,\alpha,\beta)\}.$$

Next, we study the large sample properties of the maximum likelihood estimators and the empirical log-likelihood ratio functions. Let (ν_0,α_0,β_0) be the true value of (ν,α,β) with $\alpha_0\in(0,1)$, which excludes the trivial cases $\alpha_0=0$ and $\alpha_0=1$. Define $\pi_p(z,\beta)=1-\exp\{-\int_0^{\tau}\lambda(t,z,\beta)\mathrm{d}t\}$ to be the parametric counterpart of $\pi(z)$.

Theorem 1. Suppose that $\int\pi_p(z,\beta)^{-1}\mathrm{d}F_{Z^*}(z)<\infty$ for β in a neighbourhood of β_0. If the matrix W_p that is defined in equation (4) of the on-line supplementary document is non-singular, then, as $\nu_0\to\infty$,

(a) $\sqrt{\nu_0}(\log(\hat{\nu}_p/\nu_0),\,\hat{\alpha}_p-\alpha_0,\,\hat{\beta}_p-\beta_0)^{T}\to^{d} N(0,W_p^{-1})$, where $\to^{d}$ stands for convergence in distribution, and
(b) $R_p(\nu_0,\alpha_0,\beta_0)\to^{d}\chi^2_{k+2}$ and $R'_p(\nu_0)\to^{d}\chi^2_1$, where k is the dimension of β_0.

We suggest constructing a confidence interval for ν_0 at level $1-\alpha$ as $\{\nu: R'_p(\nu)\leqslant\chi^2_{1,1-\alpha}\}$, where $\chi^2_{1,1-\alpha}$ is the $(1-\alpha)$-quantile of the χ^2-distribution with 1 degree of freedom. Theorem 1 indicates that this confidence interval has an asymptotically correct coverage probability.

As an alternative to $\hat{\nu}_p$, Chen's (2001) CL estimator of ν is $\tilde{\nu}_p = \Sigma_{i=1}^n \pi_p(Z_i, \tilde{\beta}_p)^{-1}$, where $\tilde{\beta}_p = \arg\max_\beta L^c(\beta)$. We find that the proposed maximum EL estimator $\hat{\nu}_p$ has the same asymptotic behaviour as the CL estimator $\tilde{\nu}_p$.

Let $\dot{\lambda}(t, z, \beta) = \partial\lambda(t, z, \beta)/\partial\beta$. Define $\varphi_p = \mathbb{E}\{\pi_p(Z^*, \beta_0)^{-1}\}$, $V_{p22} = \alpha_0^{-1} - \varphi_p$ and

$$V_{p32} = V_{p23}^{\mathrm{T}} = -\mathbb{E}\left\{ \frac{1-\pi_p(Z^*, \beta_0)}{\pi_p(Z^*, \beta_0)} \int_0^\tau \dot{\lambda}(t, Z^*, \beta_0)\,\mathrm{d}t \right\},$$

$$V_{p33} = \mathbb{E}\left[\int_0^\tau \frac{\dot{\lambda}(t, Z^*, \beta_0)^{\otimes 2}}{\lambda(t, Z^*, \beta_0)} \mathrm{d}t - \frac{1-\pi_p(Z^*, \beta_0)}{\pi_p(Z^*, \beta_0)} \left\{ \int_0^\tau \dot{\lambda}(t, Z^*, \beta_0)\mathrm{d}t \right\}^{\otimes 2} \right].$$

The expectation operator $\mathbb{E}$ is with respect to F_{Z^*} and $A^{\otimes 2} = AA^{\mathrm{T}}$.

Theorem 2. Under the assumptions in theorem 1, as $\nu_0 \to \infty$, we have the following results:

(a) $\hat{\beta}_p - \tilde{\beta}_p = O_p(\nu_0^{-1})$ and $\hat{\nu}_p - \tilde{\nu}_p = O_p(1)$;
(b) both $\surd\nu_0(\hat{\beta}_p - \beta_0)$ and $\surd\nu_0(\tilde{\beta}_p - \beta_0)$ converge in distribution to $N(0, V_{p33}^{-1})$;
(c) both $\nu_0^{-1/2}(\hat{\nu}_p - \nu_0)$ and $\nu_0^{-1/2}(\tilde{\nu}_p - \nu_0)$ converge in distribution to $N(0, \sigma_p^2)$, where $\sigma_p^2 = \varphi_p - 1 + V_{p23} V_{p33}^{-1} V_{p32}$.

In the on-line supplementary document, we show that σ_p^2 is exactly equal to the asymptotic variance in equation (3.5) of Chen (2001). Since Chen showed that his estimator achieves the semiparametric efficiency lower bound, theorem 2 implies that the proposed EL estimator $\hat{\nu}_p$ also achieves the semiparametric efficiency lower bound.

When constructing confidence intervals for ν based on the CL estimator $\tilde{\nu}_p$, we need a consistent estimator of its asymptotic variance σ_p^2, e.g. $\tilde{\sigma}_p^2 = \tilde{\varphi}_p - 1 + \tilde{V}_{p23}\tilde{V}_{p33}^{-1}\tilde{V}_{p32}$, where $\tilde{\varphi}_p = (1/\tilde{\nu}_p)\Sigma_{i=1}^n \pi_p(Z_i, \tilde{\beta}_p)^{-2}$, and

$$\tilde{V}_{p32} = \tilde{V}_{p23}^{\mathrm{T}} = -\frac{1}{\tilde{\nu}_p} \sum_{i=1}^n \frac{1-\pi_p(Z_i, \tilde{\beta}_p)}{\pi_p(Z_i, \tilde{\beta}_p)^2} \int_0^\tau \dot{\lambda}(t, Z_i, \tilde{\beta}_p)\mathrm{d}t,$$

$$\tilde{V}_{p33} = \frac{1}{\tilde{\nu}_p} \sum_{i=1}^n \left[\frac{1}{\pi_p(Z_i, \tilde{\beta}_p)} \int_0^\tau \frac{\dot{\lambda}(t, Z_i, \tilde{\beta}_p)^{\otimes 2}}{\lambda(t, Z_i, \tilde{\beta}_p)} \mathrm{d}t - \frac{1-\pi_p(Z_i, \tilde{\beta}_p)}{\pi_p(Z_i, \tilde{\beta}_p)^2} \left\{ \int_0^\tau \dot{\lambda}(t, Z_i, \tilde{\beta}_p)\mathrm{d}t \right\}^{\otimes 2} \right].$$

It is worth noting that, in the expression of $\tilde{\sigma}_p^2$, we use not the maximum EL estimator $\hat{\beta}_p$ but the maximum CL estimator $\tilde{\beta}_p$. This is because a variance estimator is needed by the Wald-type confidence intervals but not by the EL confidence intervals. We can verify that $\tilde{\sigma}_p^2$ is indeed a $\surd\nu_0$-consistent estimator of σ_p^2 by the consistency of $(\tilde{\beta}_p, \tilde{\nu}_p)$ and the central limit theorem.

3.2. Semiparametric intensity model

If we are not sure about the form of the intensity function $\lambda(t|z)$, a completely parametric intensity model would be risky. To alleviate the risk, we consider semiparametric models, and the proportional hazard model (Cox, 1972) is probably the most popular. Cox's model assumes that $\lambda(t|z) = \lambda_0(t)\exp(z^{\mathrm{T}}\beta)$, where $\lambda_0(t)$, which is independent of the covariate z, is an unknown baseline intensity function. Under this model, the empirical log-likelihood function in equation (8) becomes

$$l\{\nu,\alpha,\beta,\lambda_0(\cdot)\} = \log\binom{\nu}{n} + (\nu-n)\log(\alpha) + \sum_{i=1}^{n}\Big(-\exp(Z_i^{\mathrm{T}}\beta)\int_0^{\tau}\lambda_0(t)\mathrm{d}t + \int_0^{\tau}[\log\{\lambda_0(t)\} + Z_i^{\mathrm{T}}\beta]\mathrm{d}N_i(t)\Big) - \sum_{i=1}^{n}\log\Big(1+\xi\Big[\exp\Big\{-\exp(Z_i^{\mathrm{T}}\beta)\int_0^{\tau}\lambda_0(t)\mathrm{d}t\Big\} - \alpha\Big]\Big),$$

where $\xi = \xi\{\alpha,\beta,\lambda_0(\cdot)\}$ is the solution to

$$\sum_{i=1}^{n}\frac{\exp\Big\{-\exp(Z_i^{\mathrm{T}}\beta)\int_0^{\tau}\lambda_0(t)\,\mathrm{d}t\Big\} - \alpha}{1+\xi\Big[\exp\Big\{-\exp(Z_i^{\mathrm{T}}\beta)\int_0^{\tau}\lambda_0(t)\mathrm{d}t\Big\} - \alpha\Big]} = 0$$

for fixed α, β and $\lambda_0(\cdot)$.

To facilitate our statistical inference for ν, we need to profile out the infinite dimensional baseline intensity function $\lambda_0(\cdot)$ in $l\{\nu,\alpha,\beta,\lambda_0(\cdot)\}$. Let $0 = t_0 < t_1 < t_2 < \ldots < t_K \leqslant \tau$ be distinct time points for which there is at least one capture. Denote $h_k = \int_{t_{k-1}}^{t_k}\lambda_0(s)\mathrm{d}s$ for $k = 1,2,\ldots,K$. When K is sufficiently large, $\max_{1\leqslant k\leqslant K}(t_k - t_{k-1})$ will be very small. Therefore, $\lambda_0(t_k) \approx h_k/(t_k - t_{k-1})$ for all k, by the mean value theorem. With this reasoning, the empirical log-likelihood becomes

$$\log\binom{\nu}{n} + (\nu-n)\log(\alpha) + \sum_{i=1}^{n}\sum_{k=1}^{K}[-\exp(Z_i^{\mathrm{T}}\beta)h_k + \{\log(h_k) + Z_i^{\mathrm{T}}\beta\}\Delta N_i(t_k)] + C - \sum_{i=1}^{n}\log\Big(1+\xi\Big[\exp\Big\{-\exp(Z_i^{\mathrm{T}}\beta)\sum_{k=1}^{K}h_k\Big\} - \alpha\Big]\Big), \qquad (10)$$

where $\Delta N_i(t_k) = N_i(t_k) - N_i(t_k-)$ is the number of captures at the time points t_k, and $C = -\Sigma_{i=1}^{n}\Sigma_{k=1}^{K}\Delta N_i(t_k)\log(t_k - t_{k-1})$ is independent of the unknown parameters.

Since the capture process is continuous in a closed time interval $[0,\tau]$, it is reasonable to assume that each component process N_i has a finite number of jumps, each positive and of size 1, and two component processes N_i and N_j $(i \neq j)$ cannot jump at the same time point; see Andersen and Gill (1982). This implies that $\Sigma_{i=1}^{n}\Delta N_i(t_k) = 1$ holds at any time point $k = 1,2,\ldots,K$.

Setting the partial derivatives of the log-likelihood in expression (10) with respect to h_k to 0, we have $h_1 = h_2 = \ldots = h_K$. Denote their common value by h and let $\phi = Kh$. Up to quantities that are independent of the unknown parameters, the log-likelihood in expression (10) has the same maximizer as

$$l_{\mathrm{s}}(\nu,\alpha,\beta,\phi) = \log\binom{\nu}{n} + (\nu-n)\log(\alpha) - \phi\sum_{i=1}^{n}\exp(Z_i^{\mathrm{T}}\beta) + \sum_{i=1}^{n}\{Z_i^{\mathrm{T}}\beta + \log(\phi)\}N_i(\tau) - \sum_{i=1}^{n}\log(1+\xi[\exp\{-\exp(Z_i^{\mathrm{T}}\beta)\phi\} - \alpha]), \qquad (11)$$

where $\xi = \xi(\alpha,\beta,\phi)$ is the solution to

$$\sum_{i=1}^{n}\frac{\exp\{-\exp(Z_i^{\mathrm{T}}\beta)\phi\} - \alpha}{1+\xi[\exp\{-\exp(Z_i^{\mathrm{T}}\beta)\phi\} - \alpha]} = 0.$$

We use the subscript s to highlight the *semiparametric* intensity model. The maximum likelihood estimator of (ν,α,β,ϕ) is $(\hat{\nu}_{\mathrm{s}},\hat{\alpha}_{\mathrm{s}},\hat{\beta}_{\mathrm{s}},\hat{\phi}_{\mathrm{s}}) = \arg\max l_{\mathrm{s}}(\nu,\alpha,\beta,\phi)$. Accordingly, we define the likelihood ratio functions of (ν,α,β,ϕ) and ν as

$$R_s(\nu,\alpha,\beta,\phi)=2\{l_s(\hat\nu_s,\hat\alpha_s,\hat\beta_s,\hat\phi_s)-l_s(\nu,\alpha,\beta,\phi)\},$$
$$R'_s(\nu)=2\{l_s(\hat\nu_s,\hat\alpha_s,\hat\beta_s,\hat\phi_s)-\sup_{\alpha,\beta,\phi} l_s(\nu,\alpha,\beta,\phi)\}.$$

Let $\phi_0=\int_0^\tau \lambda_0(t)\mathrm{d}t$ when $\lambda_0(\cdot)$ takes its true value, and suppose that (ν_0,α_0,β_0) is the true value of (ν,α,β) with $\alpha_0\in(0,1)$. Define $\pi_s(z,\beta,\phi)=1-\exp\{-\phi\exp(z^{\mathrm{T}}\beta)\}$ with $\pi_s(z)=\pi_s(z,\beta_0,\phi_0)$. Similarly to the parametric case, we find that the maximum likelihood estimator $\hat\nu_s$ is asymptotically normal and the EL ratio statistic follows an asymptotic χ^2_1-distribution.

Theorem 3. Suppose that $\int\pi_s(z,\beta,\phi)^{-1}\mathrm{d}F_{Z^*}(z)<\infty$ for (β,ϕ) in a neighbourhood of (β_0,ϕ_0). If W_s defined in equation (5) of the on-line supplementary document is non-singular, then, as $\nu_0\to\infty$,

(a) $\sqrt{\nu_0}(\log(\hat\nu_s/\nu_0),\ \hat\alpha_s-\alpha_0,\ \hat\beta_s-\beta_0,\ \hat\phi_s-\phi_0)^{\mathrm{T}}\to^{\mathrm{d}} N(0,W_s^{-1})$, and
(b) $R_s(\nu_0,\alpha_0,\beta_0,\phi_0)\to^{\mathrm{d}}\chi^2_{k+3}$ and $R'_s(\nu_0)\to^{\mathrm{d}}\chi^2_1$, where k is the dimension of β_0.

With the same reasoning as in the derivation of $l_s(\nu,\alpha,\beta,\phi)$, we can show that the conditional log-likelihood $\log(L^{\mathrm{c}})$ in equation (5) is equivalent, up to a constant, to

$$l_s(\beta,\phi)=\sum_{i=1}^n(-\phi\exp(Z_i^{\mathrm{T}}\beta)-\log[1-\exp\{-\phi\exp(Z_i^{\mathrm{T}}\beta)\}]+\{\log(\phi)+Z_i^{\mathrm{T}}\beta\}N_i(\tau)).$$

Denote the maximum CL estimator as $(\tilde\beta_s,\tilde\phi_s)=\arg\max_{\beta,\phi} l_s(\beta,\phi)$. Accordingly, Chen's (2001) CL estimator of ν is $\tilde\nu_s=\Sigma_{i=1}^n\pi_s(Z_i,\tilde\beta_s,\tilde\phi_s)^{-1}$. Like those in the parametric case, the maximum EL estimators and the CL estimators are asymptotically equivalent in the semiparametric case. Let $\varphi_s=\mathbb{E}\{\pi_s(Z^*)^{-1}\}$, and define

$$V_{s22}=\alpha_0^{-1}-\varphi_s,$$

$$V_{s32}=V_{s23}^{\mathrm{T}}=-\phi_0\,\mathbb{E}\left\{\frac{1-\pi_s(Z^*)}{\pi_s(Z^*)}\exp(Z^{*\mathrm{T}}\beta_0)Z^*\right\},$$

$$V_{s33}=\mathbb{E}\left(\left[\phi_0\exp(Z^{*\mathrm{T}}\beta_0)-\frac{1-\pi_s(Z^*)}{\pi_s(Z^*)}\{\phi_0\exp(Z^{*\mathrm{T}}\beta_0)\}^2\right]Z^*Z^{*\mathrm{T}}\right),$$

$$V_{s24}=V_{s42}=-\mathbb{E}\left\{\frac{1-\pi_s(Z^*)}{\pi_s(Z^*)}\exp(Z^{*\mathrm{T}}\beta_0)\right\},$$

$$V_{s34}=V_{s43}^{\mathrm{T}}=\mathbb{E}\left[\left\{\exp(Z^{*\mathrm{T}}\beta_0)-\frac{1-\pi_s(Z^*)}{\pi_s(Z^*)}\phi_0\exp(2Z^{*\mathrm{T}}\beta_0)\right\}Z^*\right],$$

$$V_{s44}=\mathbb{E}\left\{\frac{1}{\phi_0}\exp(Z^{*\mathrm{T}}\beta_0)-\frac{1-\pi_s(Z^*)}{\pi_s(Z^*)}\exp(2Z^{*\mathrm{T}}\beta_0)\right\}.$$

Theorem 4. Assume the conditions in theorem 3. Let $\hat\Theta=(\hat\beta_s^{\mathrm{T}},\hat\phi_s)^{\mathrm{T}}$, $\tilde\Theta=(\tilde\beta_s^{\mathrm{T}},\tilde\phi_s)^{\mathrm{T}}$ and $\Theta=(\beta^{\mathrm{T}},\phi)^{\mathrm{T}}$ with true value $\Theta_0=(\beta_0^{\mathrm{T}},\phi_0)^{\mathrm{T}}$. As $\nu_0\to\infty$:

(a) $\hat\Theta-\tilde\Theta=O_p(\nu_0^{-1})$ and $\hat\nu-\tilde\nu=O_p(1)$;
(b) both $\sqrt{\nu_0}(\hat\Theta-\Theta_0)$ and $\sqrt{\nu_0}(\tilde\Theta-\Theta_0)$ converge in distribution to $N(0,V_\Theta^{-1})$, where $V_\Theta=(V_{sij})_{3\leqslant i,j\leqslant 4}$;

(c) both $\nu_0^{-1/2}(\hat{\nu}_s - \nu_0)$ and $\nu_0^{-1/2}(\tilde{\nu}_s - \nu_0)$ converge in distribution to $N(0, \sigma_s^2)$, where $\sigma_s^2 = \varphi_s - 1 + V_{s2}^T V_\Theta^{-1} V_{s2}$, $V_{s2} = (V_{s23}, V_{s24})^T$.

In the on-line supplementary document, we also show that σ_s^2 is exactly equal to the asymptotic variance at the bottom of page 613 of Chen (2001) when the covariate is time independent. Since Chen's abundance estimator achieves the semiparametric efficiency lower bound, theorem 4 implies that so does the proposed EL estimator $\hat{\nu}_s$.

For the construction of Wald-type confidence intervals based on $\tilde{\nu}_s$, we estimate σ_s^2 by $\tilde{\sigma}_s^2 = \tilde{\varphi}_s - 1 + \tilde{V}_{s2}^T \tilde{V}_\Theta^{-1} \tilde{V}_{s2}$. Here $\tilde{\varphi}_s = (\tilde{\nu}_s)^{-1} \Sigma_{i=1}^n \pi_s(Z_i, \tilde{\beta}_s, \tilde{\phi}_s)^{-2}$, $\tilde{V}_{s2}^T = (\tilde{V}_{s23}, \tilde{V}_{s24})$ and $\tilde{V}_\Theta^T = (\tilde{V}_{s2ij})_{3 \leqslant i,j \leqslant 4}$, where

$$\tilde{V}_{s23} = -\frac{\tilde{\phi}_s}{\tilde{\nu}_s} \sum_{i=1}^n \left\{ \frac{1 - \pi_s(Z_i, \tilde{\beta}_s, \tilde{\phi}_s)}{\pi_s(Z_i, \tilde{\beta}_s, \tilde{\phi}_s)^2} \exp(Z_i^T \tilde{\beta}_s) Z_i^T \right\},$$

$$\tilde{V}_{s24} = -\frac{1}{\tilde{\nu}_s} \sum_{i=1}^n \left\{ \frac{1 - \pi_s(Z_i, \tilde{\beta}_s, \tilde{\phi}_s)}{\pi_s(Z_i, \tilde{\beta}_s, \tilde{\phi}_s)^2} \exp(Z_i^T \tilde{\beta}_s) \right\},$$

$$\tilde{V}_{s33} = \frac{\tilde{\phi}_s}{\tilde{\nu}_s} \sum_{i=1}^n \left\{ \frac{1}{\pi_s(Z_i, \tilde{\beta}_s, \tilde{\phi}_s)} \exp(Z_i^T \tilde{\beta}_s) - \frac{1 - \pi_s(Z_i, \tilde{\beta}_s, \tilde{\phi}_s)}{\pi_s(Z_i, \tilde{\beta}_s, \tilde{\phi}_s)^2} \tilde{\phi} \exp(2 Z_i^T \tilde{\beta}_s) \right\} Z_i Z_i^T,$$

$$\tilde{V}_{s34} = \frac{1}{\tilde{\nu}_s} \sum_{i=1}^n \left\{ \frac{1}{\pi_s(Z_i, \tilde{\beta}_s, \tilde{\phi}_s)} \exp(Z_i^T \tilde{\beta}_s) - \frac{1 - \pi_s(Z_i, \tilde{\beta}_s, \tilde{\phi}_s)}{\pi_s(Z_i, \tilde{\beta}_s, \tilde{\phi}_s)^2} \tilde{\phi} \exp(2 Z_i^T \tilde{\beta}_s) \right\} Z_i,$$

$$\tilde{V}_{s44} = \frac{1}{\tilde{\phi}_s \tilde{\nu}_s} \sum_{i=1}^n \left\{ \frac{1}{\pi_s(Z_i, \tilde{\beta}_s, \tilde{\phi}_s)} \exp(Z_i^T \tilde{\beta}_s) - \frac{1 - \pi_s(Z_i, \tilde{\beta}_s, \tilde{\phi}_s)}{\pi_s(Z_i, \tilde{\beta}_s, \tilde{\phi}_s)^2} \tilde{\phi} \exp(2 Z_i^T \tilde{\beta}_s) \right\}.$$

The consistency of $\tilde{\beta}_s$ and $\tilde{\phi}_s$ implies that $\tilde{\sigma}_s^2$ is a $\sqrt{\nu_0}$-consistent estimator of σ_s^2.

4. Simulation study

We carry out simulations to study the finite sample performance of the proposed EL inference approach for point and interval estimation under both parametric and semiparametric intensity models. The numerical procedure for implementing the EL-based methods is discussed in the on-line supplementary document.

For the point estimation, we compare the proposed abundance estimator $\hat{\nu}$ ($\hat{\nu}_p$ or $\hat{\nu}_s$) with Chen's (2001) estimator $\tilde{\nu}$ ($\tilde{\nu}_p$ or $\tilde{\nu}_s$). The proposed EL confidence interval for ν is

$$\mathcal{I}_1 = \{\nu : R'(\nu) \leqslant \chi^2_{1,1-\alpha}\},$$

where $R'(\nu) = R'_p(\nu)$ (parametric case) or $R'_s(\nu)$ (semiparametric case). A Wald-type confidence interval based on Chen's estimator $\tilde{\nu}$ is

$$\mathcal{I}_2 = \{\nu : (\tilde{\nu} - \nu)^2 / (\tilde{\nu}\tilde{\sigma}^2) \leqslant \chi^2_{1,1-\alpha}\},$$

where $(\tilde{\nu}, \tilde{\sigma}) = (\tilde{\nu}_p, \tilde{\sigma}_p)$ (parametric case) or $(\tilde{\nu}_s, \tilde{\sigma}_s)$ (semiparametric case). Theorems 1–4 indicate that all the above confidence intervals have asymptotically correct coverage probabilities. For a generic two-sided confidence interval $\mathcal{I} = [\nu_l, \nu_u]$, we also study the performance of the corresponding one-sided confidence intervals $[\nu_l, \infty]$ (lower limit) and $[n, \nu_u]$ (upper limit).

4.1. *Simulation set-up*

We fix the population size to $\nu_0 = 100$ or $\nu_0 = 200$, set the period of the recapture study to [0,2] and consider a bivariate covariate $Z^* = (Z^*_{(1)}, Z^*_{(2)})$, where $Z^*_{(1)}$ and $Z^*_{(2)}$ are independent of each other. We generate data from the following two scenarios.

(a) Scenario A is borrowed from Chen (2001). Here $Z^*_{(1)}$ and $Z^*_{(2)}$ follow a uniform distribution on [0,1] and a binomial distribution $B(1, 0.5)$ respectively. The true value of β is $\beta_0 = (0.3, -0.2)^{\mathrm{T}}$ and the intensity function $\lambda(t, Z, \beta) = t \exp(Z^{\mathrm{T}}\beta)$.
(b) Scenario B is scenario A with $\beta_0 = (-3.2, 0.8)^{\mathrm{T}}$.

The overall probabilities of being captured in scenarios A and B are 87.5% and 49.0% respectively. Our simulation results are based on 5000 simulated data sets. When interval estimation is studied, the confidence levels are set to 90%, 95% and 99%.

4.2. *Simulation results*

When modelling the intensity function by a parametric model, we choose $\lambda(t|Z) = t \exp(Z^{\mathrm{T}}\beta)$ with β unknown, whereas, when modelling it by a semiparametric model, we choose $\lambda(t|Z) = \lambda_0(t) \exp(Z^{\mathrm{T}}\beta)$, where both β and $\lambda_0(\cdot)$ are unknown.

4.2.1. *Point estimation comparison*

We first examine the performance of the point estimators $\hat{\nu}_{\mathrm{p}}$ and $\tilde{\nu}_{\mathrm{p}}$. Table 1 lists their simulated medians, averages and MSEs. We observe that the MSEs of the proposed estimator $\hat{\nu}_{\mathrm{p}}$ are uniformly smaller than those of Chen's estimator $\tilde{\nu}_{\mathrm{p}}$. Compared with $\tilde{\nu}_{\mathrm{p}}$, $\hat{\nu}_{\mathrm{p}}$ has a noticeably increasing gain in MSE as the overall capture probability decreases; see the comparisons of scenarios A and B.

We display the plots of $\hat{\nu}_{\mathrm{p}}$ *versus* $\tilde{\nu}_{\mathrm{p}}$ under the parametric intensity model in the on-line supplementary document. The plots for the semiparametric case are similar and so have been omitted. Together with the medians and averages in Table 1, these plots indicate that, although the two estimators are generally close to each other, the proposed estimator $\hat{\nu}$ is usually closer to the true abundance and has smaller MSEs (particularly in scenario B). In addition, when we relax the intensity model from parametric to semiparametric, the MSEs become larger since less information is available for the point estimation.

4.2.2. *Interval estimation comparison*

We report in Table 2 the simulated coverage probabilities of the EL confidence interval $\mathcal{I}_1$ and the Wald-type confidence interval $\mathcal{I}_2$, which is based on the asymptotic normality of $\tilde{\nu}$. We display in Fig. 1 the *QQ*-plots of the empirical likelihood ratio $R'_{\mathrm{p}}(\nu_0)$ *versus* the χ^2_1-distribution, and the pivotal statistic $(\tilde{\nu}_{\mathrm{p}} - \nu_0)/(\tilde{\nu}_{\mathrm{p}}^{1/2}\tilde{\sigma}_{\mathrm{p}})$ *versus* $N(0, 1)$ under the parametric intensity model with $\nu_0 = 100$. The *QQ*-plots for $\nu_0 = 200$ under the parametric model and for both $\nu_0 = 100$ and $\nu_0 = 200$ under the semiparametric model are similar and are in the on-line supplementary document for brevity.

Let us first examine the two-sided coverage probabilities of the two confidence intervals. In all cases the coverage probabilities of the EL confidence interval $\mathcal{I}_1$ are very close to the nominal levels, and the departure is at most 1.04%; see the case in scenario B under the parametric intensity model with $\nu_0 = 100$ at the nominal level 90%. Although the Wald-type interval $\mathcal{I}_2$ has acceptable performance in scenario A, its coverage probability is well below the nominal levels in scenario B, and the undercoverage can be as large as 4.5%. From the *QQ*-plots in Fig. 1, we observe that the distribution of the EL ratio $R'(\nu_0)$ is much closer to its limiting χ^2_1-distribution

Table 1. Medians, averages and MSEs of proposed estimate $\hat{\nu}$ and Chen's (2001) estimate $\tilde{\nu}$

		Results for parametric intensity model				*Results for semiparametric intensity model*			
		Scenario A		*Scenario B*		*Scenario A*		*Scenario B*	
		$\nu_0 = 100$	$\nu_0 = 200$	$\nu_0 = 100$	$\nu_0 = 200$	$\nu_0 = 100$	$\nu_0 = 200$	$\nu_0 = 100$	$\nu_0 = 200$
Median	$\hat{\nu}$	99.90	200.03	96.50	197.27	100.17	200.17	99.45	199.89
	$\tilde{\nu}$	100.59	200.71	100.25	201.56	100.94	200.92	103.65	204.08
Average	$\hat{\nu}$	99.96	200.02	109.69	207.76	100.25	200.28	113.99	210.79
	$\tilde{\nu}$	100.66	200.72	114.35	212.12	101.02	201.04	119.02	215.28
MSE	$\hat{\nu}$	20.32	39.77	5184.84	2731.30	22.95	43.84	5775.60	2968.56
	$\tilde{\nu}$	21.04	40.55	5690.30	2988.17	24.30	45.15	6419.77	3267.31

Table 2. Simulated coverage probabilities for the EL confidence interval $\mathcal{I}_1$ and the Wald-type confidence interval $\mathcal{I}_2$

Type	*Level*		*Results for parametric intensity model*				*Results for semiparametric intensity model*			
			Scenario A		*Scenario B*		*Scenario A*		*Scenario B*	
			$\nu_0 = 100$	$\nu_0 = 200$	$\nu_0 = 100$	$\nu_0 = 200$	$\nu_0 = 100$	$\nu_0 = 200$	$\nu_0 = 100$	$\nu_0 = 200$
Two sided	90%	$\mathcal{I}_1$	90.23	90.13	88.96	89.04	89.93	90.37	89.40	89.11
		$\mathcal{I}_2$	90.64	90.48	87.26	90.08	90.78	91.00	89.48	91.12
	95%	$\mathcal{I}_1$	94.98	95.18	94.32	94.52	94.89	95.18	94.66	94.63
		$\mathcal{I}_2$	95.62	95.10	90.50	92.46	95.74	95.46	92.44	93.32
	99%	$\mathcal{I}_1$	98.94	98.86	98.66	99.04	99.10	98.92	98.92	99.16
		$\mathcal{I}_2$	98.70	98.88	94.74	96.22	98.64	98.94	95.64	96.84
Lower limit	90%	$\mathcal{I}_1$	90.23	90.01	86.78	88.78	90.91	90.49	89.12	90.03
		$\mathcal{I}_2$	90.60	90.32	82.64	85.42	91.20	90.76	85.38	86.98
	95%	$\mathcal{I}_1$	95.06	95.10	93.30	93.94	95.29	95.66	94.74	94.67
		$\mathcal{I}_2$	94.34	94.74	87.26	90.08	94.48	95.06	89.48	91.12
	99%	$\mathcal{I}_1$	98.94	99.06	98.28	98.62	99.00	99.12	98.72	98.92
		$\mathcal{I}_2$	98.14	98.40	93.42	95.00	98.02	98.60	94.68	95.76
Upper limit	90%	$\mathcal{I}_1$	90.43	90.05	91.22	90.50	89.29	89.27	89.96	89.51
		$\mathcal{I}_2$	90.72	90.16	99.98	98.16	89.70	89.42	100.0	98.00
	95%	$\mathcal{I}_1$	95.18	95.04	95.66	95.10	94.63	94.71	94.66	94.43
		$\mathcal{I}_2$	96.30	95.74	100.0	100.0	96.30	95.94	100.0	100.0
	99%	$\mathcal{I}_1$	99.22	98.88	99.04	99.24	99.12	98.88	98.78	99.14
		$\mathcal{I}_2$	99.82	99.48	100.0	100.0	99.86	99.50	100.0	100.0

than that of the pivotal statistic $(\tilde{\nu} - \nu_0)/(\tilde{\nu}^{1/2}\tilde{\sigma})$ is to its limiting distribution $N(0, 1)$. This explains why the EL confidence interval $\mathcal{I}_1$ has more accurate coverage probabilities than the Wald-type confidence interval $\mathcal{I}_2$.

We now investigate the one-sided coverage probabilities of the two confidence intervals. The coverage probabilities of both the lower and the upper limits of the EL confidence interval $\mathcal{I}_1$ are again the closest to the nominal levels in most cases. The lower limits of $\mathcal{I}_2$ often produce undercoverage, but its upper limits often produce overcoverage. As indicated by Fig. 1, a possible

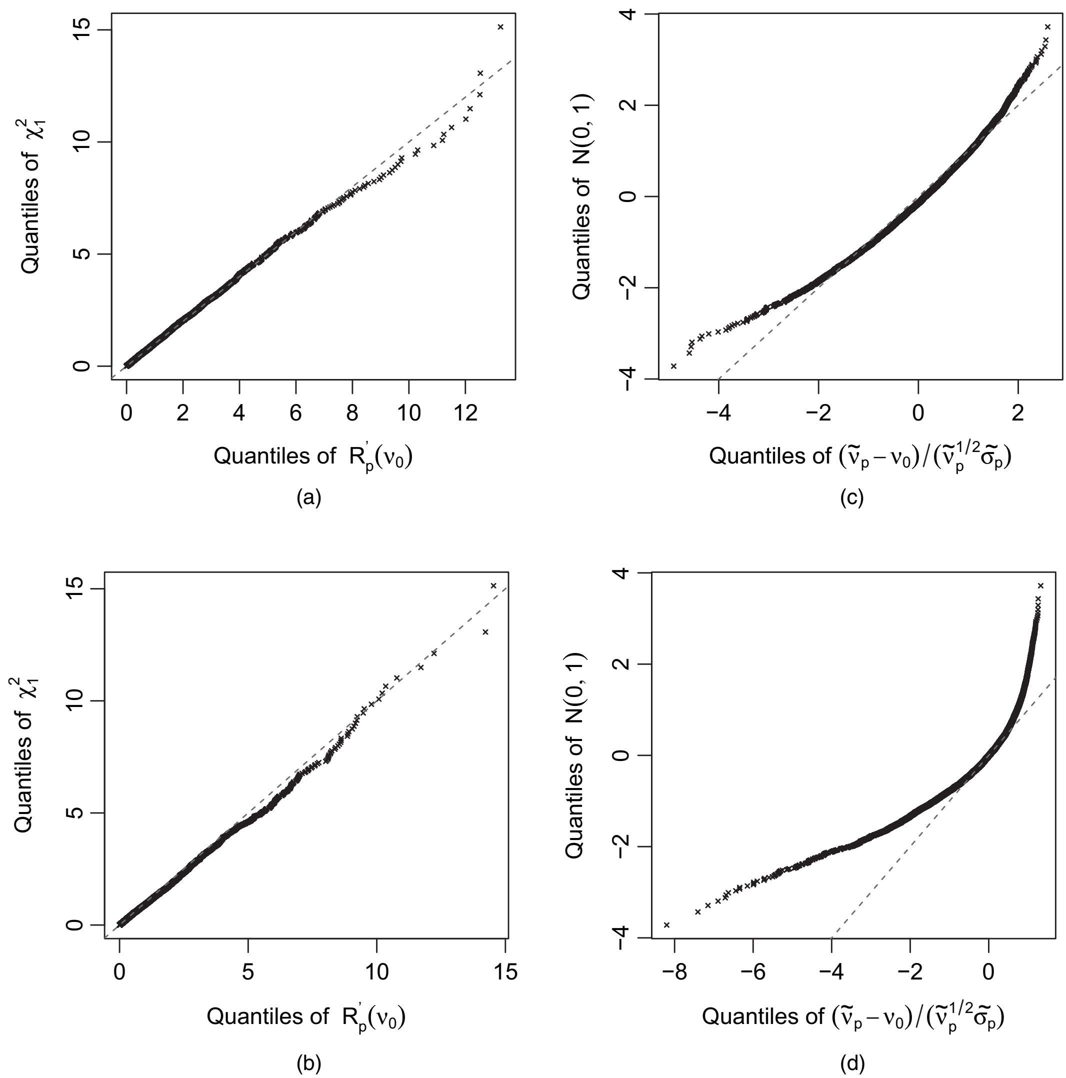

Fig. 1. (a), (b) *QQ*-plots of $R'_p(\nu_0)$ and (c), (d) $(\tilde{\nu}_p - \nu_0)/(\tilde{\nu}_p^{1/2}\tilde{\sigma}_p)$, under the parametric intensity model in (a), (c) scenario A and (b), (d) scenario B with $\nu_0 = 100$

reason for this is that the quantiles of $(\tilde{\nu} - \nu_0)/(\tilde{\nu}^{1/2}\tilde{\sigma})$ are generally smaller than those of $N(0, 1)$ in both scenario A and scenario B. As the overall probability of being captured decreases from 87.5% to 49.0% or from scenario A to B, both the lower and the upper limits of $\mathcal{I}_2$ have worse coverage probabilities. This can be explained by Fig. 1: from scenario A to B, the finite sample distribution of $(\tilde{\nu} - \nu_0)/(\tilde{\nu}^{1/2}\tilde{\sigma})$ looks even further from its limiting distribution $N(0, 1)$.

Overall, the proposed estimator $\hat{\nu}$ is more reliable and more accurate than Chen's estimator $\tilde{\nu}$. The empirical likelihood confidence interval $\mathcal{I}_1$ always has more accurate coverage probabilities and more stable performance than the Wald-type confidence interval $\mathcal{I}_2$. The usual normality-based confidence interval $\mathcal{I}_2$ has severe two-sided undercoverage and unacceptable one-sided coverage when the probability of being captured is around $\frac{1}{2}$.

In the on-line supplementary document, we present additional simulation results for a small overall capture probability and a large ν_0. We also perform additional simulations to gain insight into when we can feel confident in the estimate proposed.

5. Real data analysis

We illustrate the proposed full likelihood estimation procedure by analysing illegal immigrant data for the Netherlands and *Prinia flaviventris* data from Hong Kong. For each data set, we choose to use the parametric model $\lambda(t|Z)=\exp(\beta_0+\beta_1^{\mathrm{T}}Z)$ and the semiparametric model $\lambda(t|Z)=\lambda_0(t)\exp(\beta_1^{\mathrm{T}}Z)$ to model the intensity function, where Z denotes a vector of covariates.

We make a remark on the inference under the two model assumptions. Let $l_{\mathrm{p}}(\nu,\alpha,\beta_0,\beta_1)$ and $l_{\mathrm{s}}(\nu,\alpha,\beta_1,\phi)$ be the profile log-likelihoods of the proposed method under the parametric and semiparametric models respectively. It can be verified that

$$l_{\mathrm{p}}(\nu,\alpha,\beta_0,\beta_1)=l_{\mathrm{s}}\{\nu,\alpha,\beta_1,\tau\exp(\beta_0)\}-\sum_{i=1}^{n}\log(\tau)N_i(\tau),$$

which implies that the empirical log-likelihood ratio functions of ν are the same under both models. The resulting maximum likelihood estimators $\hat{p}_i$ of p_i and the CL abundance estimators of ν are also equal to each other. Consequently, the point and interval estimators of the EL method and Chen's (2001) CL method coincide under the two intensity models.

5.1. Netherlands illegal immigrant data

We first consider the estimation of illegal immigrants in the Netherlands based on data (Heijden *et al.*, 2003) obtained from police records. These are count data for illegal immigrants who could not be effectively expelled from the country. The data record the number of times that the immigrant has been apprehended by the police, and they date back to 1995; Table 3. This is a real continuous time capture–recapture data set, and it has been analysed accordingly: see Schofield *et al.* (2017) and the references therein. The data set contains 1180 distinct illegal immigrants who have been apprehended by the police at least once. Let $\tau=1$ be the regularized period. We take either gender (1 for male and 0 for female) or age (1 if an individual is no more than 40 years old and 0 otherwise) as the covariate Z. The corresponding models are called the gender model and the age model.

The analysis results are presented in Table 4. The EL abundance estimates are close to but slightly less than the CL abundance estimates. This coincides with the observations in our simulation study and again confirms the asymptotic equivalence results in theorems 2 and 4. The positive signs of the β_1-estimates for both gender and age indicate that younger individuals are more likely to be apprehended than older individuals, and males are more likely to be

Table 3. Illegal immigrant data for the Netherlands†

Covariate	*Covariate category*	f_1	f_2	f_3	f_4	f_5	f_6	*Total*
Age	>40 years	105	6					111
	<40 years	1540	177	37	13	1	1	1769
Gender	Female	366	24	6	1	1		398
	Male	1279	159	31	12		1	1482

†f_j: number of distinct individuals apprehended j times, $j=1,\ldots,6$.

Table 4. Point estimates of ν and (β_0, β_1), and 95% confidence intervals of ν under the gender and age models

Model	*Method*	*Estimate of* ν	*Confidence interval for* ν	*Estimate of* (β_0,β_1)
Age model	EL	7542.8	[6665.1, 9215.3]	(−2.2397, 1.1031)
	CL	7545.5	[6471.7, 8619.4]	(−2.2416, 1.1048)
Gender model	EL	7317.3	[6574.4, 8227.1]	(−1.5658, 0.4721)
	CL	7320.8	[6505.1, 8136.4]	(−1.5667, 0.4726)

apprehended than females. Since the age model produces a larger profile empirical log-likelihood (−655.22) than the gender model (−655.86), we recommend the age model on the basis of the Akaike information criterion. In addition, our simulation results imply that the EL point and interval estimators are more reliable than the CL estimators, so the number of illegal immigrants is likely to be 7543 with a 95% confidence interval [6665, 9215].

5.2. *Prinia flaviventris data in Hong Kong*

The second data set is a capture–recapture data set for the bird species *Prinia flaviventris* (Hwang and Huang, 2003) in Hong Kong in 1993, available in the R-package `PL.popN` (Stoklosa *et al.*, 2011). There are $n=164$ birds in total captured at least once over 17 weekly capture occasions ($\tau=17$). Although the data were obtained by a discrete time experiment, we follow Xu *et al.* (2007) and analyse them as if the birds were captured continuously. Wing length measurements (in millimetres) were collected for each individual; they are denoted by X. We take $Z=(X, X^2)$ as the covariate that is associated with the probability of being captured. In contrast with the gender and age variables in the first data set, this covariate takes continuous values.

5.2.1. *Estimation of* ν *and* $\beta=(\beta_0,\beta_1^{\mathrm{T}})^{\mathrm{T}}$

Table 5 presents point estimates for ν and β, for both the EL and the CL methods, and the EL and CL confidence intervals. We again see that the EL estimates for ν and β are almost the same as the corresponding CL estimates. However, the confidence intervals for ν based on the EL and CL are quite different. The lower limit 143.8 of $\mathcal{I}_2$ is lower than the number $n=164$ of individuals captured, which is clearly absurd. Hence, the more reliable confidence interval for ν is the EL confidence interval, which is [453, 1266] after rounding.

As observed by a referee, the capture probability or intensity function should be an increasing function of the wing length. However, the estimated intensity functions of our method and Chen's method both reach a maximum around 47, i.e. both functions are increasing when the wing length is less than 47 mm and are then decreasing. A possible reason for this is as follows. In general, a bird's wing length increases with increasing age. Juvenile and subadult birds have shorter wing lengths as well as a lower flight capacity, which may explain the lower intensity in

Table 5. Point estimates for ν and β, and 95% confidence intervals for ν

Method	*Estimate of* ν	*Confidence interval*	*Estimate of* β
EL	692.4	[453.2, 1266.0]	(−353.82, 14.944, −0.1592)
CL	709.8	[143.8, 1275.8]	(−363.81, 15.374, −0.1638)

shorter wing lengths. Adult birds with longer wings may be growing older and becoming less active. This may be the reason for the decreasing intensity in longer wings.

5.2.2. Estimation of cumulative hazards and covariate densities

The proposed EL procedure can produce reasonable estimators not only for the abundance and the underlying parameters but also for the hazard function $\lambda(t|z)$ and the marginal distribution function $F_{Z^*}(z)$, or the cumulative hazard function $\Lambda(t|z)=\int_0^t \lambda(s|z)\mathrm{d}s$ and the marginal density function $f_{Z^*}(z)$.

Under the parametric intensity model $\lambda(t,z,\beta)$, the consistency of $\hat{\beta}_{\mathrm{p}}$ implies that $\hat{\Lambda}(t|z)=\int_0^t \lambda(s,z,\hat{\beta}_{\mathrm{p}})\,\mathrm{d}s$ is a consistent estimator of the cumulative hazard $\Lambda(t|z)=\int_0^t \lambda(s|z)\,\mathrm{d}s$. Under the semiparametric intensity model $\lambda(t|z)=\lambda_0(t)\exp(z^{\mathrm{T}}\beta)$, a reasonable estimator of $\Lambda_0(t)=\int_0^t \lambda_0(s)\mathrm{d}s$ is

$$\hat{\Lambda}_0(t)=\hat{\phi}_{\mathrm{s}}K^{-1}\sum_{k=1}^{K} I(t_k \leqslant t)=\hat{\phi}_{\mathrm{s}}\frac{(1/n)\sum_{i=1}^{n}\int_0^t \mathrm{d}N_i(s)}{(1/n)\sum_{i=1}^{n}\int_0^{\tau} \mathrm{d}N_i(s)},$$

where we have used the fact that $\Sigma_{k=1}^{K} I(t_k \leqslant t)=\Sigma_{i=1}^{n}\int_0^t \mathrm{d}N_i(s)$ for any t. We estimate $\Lambda(t|z)$ via $\hat{\Lambda}(t|z)=\hat{\Lambda}_0(t)\exp(z^{\mathrm{T}}\hat{\beta}_{\mathrm{s}})$. We can show that $\hat{\Lambda}(t|z)=\hat{\Lambda}_0(t)\exp(z^{\mathrm{T}}\hat{\beta}_{\mathrm{s}})$ is a consistent estimator of $\Lambda(t|z)$.

Given $\hat{\Lambda}(t|z)=\int_0^t \lambda(s,z,\hat{\beta}_{\mathrm{p}})\mathrm{d}s$ in the parametric case or $\hat{\Lambda}_0(t)\exp(z^{\mathrm{T}}\hat{\beta}_{\mathrm{s}})$ in the semiparametric case, let

$$\hat{p}_i=\frac{1}{n}\frac{1}{1+\hat{\xi}[\exp\{-\hat{\Lambda}(\tau|z_i)\}-\hat{\alpha}]}, \tag{12}$$

where $\hat{\xi}$ is the solution to

$$\sum_{i=1}^{n}\frac{\exp\{-\hat{\Lambda}(\tau|z_i)\}-\hat{\alpha}}{1+\xi[\exp\{-\hat{\Lambda}(\tau|z_i)\}-\hat{\alpha}]}=0,$$

and $\hat{\alpha}=\hat{\alpha}_{\mathrm{p}}$ (parametric) and $\hat{\alpha}_{\mathrm{s}}$ (semiparametric). We estimate the marginal distribution $F_{Z^*}(z)$ via

$$\hat{F}_{Z^*}(z)=\sum_{i=1}^{n}\hat{p}_i I(z_i \leqslant z).$$

The inequality $z_i \leqslant z$ holds elementwise for vector-valued z_i and z, and $I(\cdot)$ is the indicator function. For both $\hat{\Lambda}(t|z)=\hat{\Lambda}(t,z,\hat{\beta}_{\mathrm{p}})$ in the parametric case and $\hat{\Lambda}(t|z)=\hat{\Lambda}_0(t)\exp(z^{\mathrm{T}}\hat{\beta}_{\mathrm{s}})$ in the semiparametric case, we have $\hat{\xi}=-1/(1-\alpha_0)+o_{\mathrm{p}}(1)$; see the proofs of theorems 1 and 3 in the on-line supplementary document. This together with the consistency of $\hat{\alpha}$ implies that $\hat{F}_{Z^*}(z)$ is a consistent estimator of $F_{Z^*}(z)$.

For the *Prinia flaviventris* data, let X^* and X be the unbiased and biased wing lengths respectively. To estimate the density function $f_{X^*}(x)$ of the unbiased wing length, we consider the naive kernel density estimator,

$$\hat{f}_{\mathrm{u}}(x)=\frac{1}{n}\sum_{i=1}^{n}\frac{G\{(x_i-x)/h\}}{h},$$

where $G(\cdot)$ is a kernel function and h is a bandwidth. Here, we choose $G(\cdot)$ to be the standard normal density function, and we set $h=1.06\hat{\sigma}_x n^{-1/5}$ by rule of thumb, where $\hat{\sigma}_x^2$ is the sample

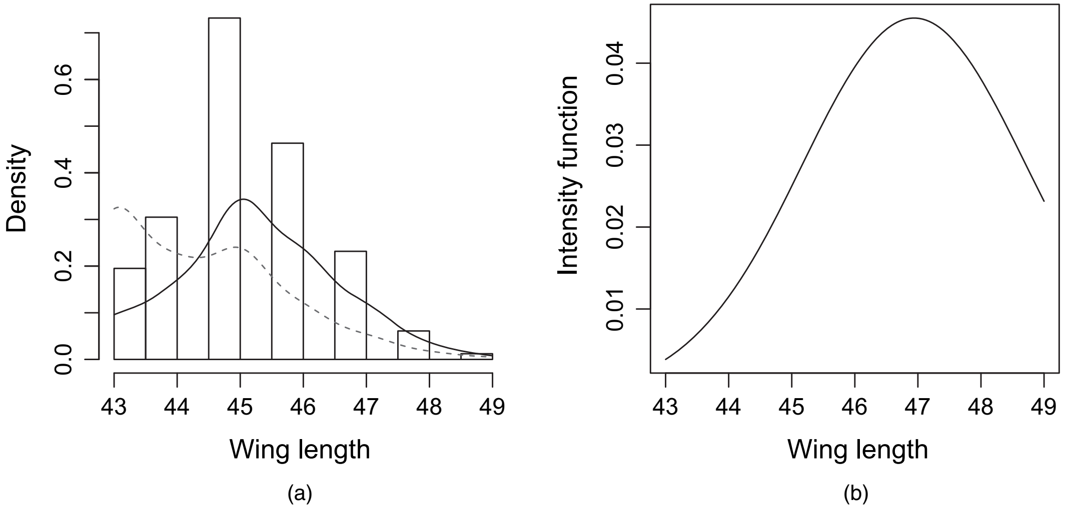

Fig. 2. (a) Histogram, naive kernel density estimate (———) and weighted kernel density estimate (– – –) of wing length under parametric model $\lambda(t|z) = \exp(\beta_0 + z^{\mathrm{T}}\beta_1)$ and (b) intensity function estimate under the parametric model

variance of the observed wing lengths $\{x_i : i = 1, 2, \ldots, n\}$. On the basis of the proposed EL, we estimate $f_{X^*}(x)$ via a weighted kernel density estimator

$$\hat{f}_{\mathrm{w}}(x) = \sum_{i=1}^{n} \hat{p}_i \frac{G\{(x_i - x)/h\}}{h},$$

where the $\hat{p}_i$s are defined in equation (12). We remark that, although the $\hat{p}_i$s are the estimated probability weights of the z_is, they are also those of the x_is because there is an invertible transformation $z_i = (x_i, x_i^2)^{\mathrm{T}}$ between them.

An advantage of the EL method proposed is that it can correct the sampling bias automatically. It can be shown that as $\nu_0 \to \infty$, if $h = o(1)$ and $(\nu_0 h^2)^{-1} = o(1)$, then

$$\hat{f}_{\mathrm{u}}(x) = \frac{1 - \exp\left\{ -\int_0^{\tau} \lambda(s|z)\,\mathrm{d}s \right\}}{1 - \alpha_0} f_{X^*}(x) + o_p(1),$$

$$\hat{f}_{\mathrm{w}}(x) = f_{X^*}(x) + o_p(1),$$

where $z = (x, x^2)^{\mathrm{T}}$. Hence, $\hat{f}_{\mathrm{u}}(\cdot)$ is asymptotically biased unless the intensity function $\lambda(t|z)$ is independent of z or the wing length x, whereas the proposed density estimator $\hat{f}_{\mathrm{w}}(x)$ is asymptotically unbiased.

Fig. 2 shows the histogram and kernel density estimates of the wing length (Fig. 2(a)) and the estimated hazard function (Fig. 2(b)) under the parametric model. The full curve is the usual kernel density estimate. It appears close to the histogram in shape since both reflect the observed wing lengths. The broken curve is the weighted kernel density estimate under the parametric model. The weighted kernel density estimate under the semiparametric model is exactly the same because the $\hat{p}_i$s are unchanged when $\lambda(t|z)$ is changed from the parametric model $\exp(\beta_0 + z^{\mathrm{T}}\beta_1)$ to the semiparametric model $\lambda_0(t)\exp(z^{\mathrm{T}}\beta_1)$. We observe that the naive kernel density estimate and the weighted kernel density estimate are very different. Together with Fig. 2(b), this finding indicates that the EL method proposed succeeds in correcting the sampling bias: it puts more probability weight on the observations with less intensity, and less weight on the observations with more intensity.

6. Discussion

In this paper, we have explored the strength of the EL (Owen, 1988) in capture–recapture studies. Our approach is a new development of the EL that solves non-regular statistical problems. It has potential applications to the estimation of abundance in many other fields where sampling bias occurs, such as truncation problems and meta-analyses of publication bias.

Overall, the point estimators proposed are quite close to the maximum CL estimators. However, the likelihood ratio confidence interval proposed has two clear advantages over the Wald-type confidence interval based on the CL: it usually has a more accurate coverage probability, and its lower limit is never less than the number of individuals that are captured. A further advantage is that it provides a reasonable estimator for the hazard function and the covariates' marginal distribution function, as illustrated in Section 5.2. Although the first real data application did not show a significant difference between the full and CL methods, our simulations show the clear advantage of the full likelihood approach.

Our approach assumes that the covariate is time independent, e.g. gender. Other covariates such as wing length may vary over time, especially when the sampling period is long. It would be interesting to study full EL inference for abundance with time-dependent covariates. One benefit of time-independent covariates is that the proposed point and the interval estimator for the abundance are independent of the time points where captures and recaptures occur. This benefit disappears when the covariate is time dependent.

Our work is based on model (1), which is an M_{th} model (Seber, 1982; Borchers *et al.*, 2002) because the intensity varies not only from individual to individual but also from capture occasion to capture occasion. It would be possible to extend our work to an M_{tbh} model by also taking behavioural effects into account (Farcomeni and Scacciatelli, 2013). We leave this to future work.

Acknowledgements

We are grateful to the Joint Editor, Associate Editor and two referees for their constructive comments that significantly improved the paper.

Dr Yukun Liu was supported by the National Natural Science Foundation of China under grants 11771144, 11371142, 11501354 and 11501208, Chinese Ministry of Education 111 project under grant B14019 and the programme of the Shanghai Subject Chief Scientist under grant 14XD1401600. Dr Pengfei Li was supported in part by Natural Sciences and Engineering Research Council of Canada grant RGPIN-2015-06592.

The first two authors contributed equally to the paper.

References

Alho, J. M. (1990) Logistic regression in capture-recapture models. *Biometrics*, **46**, 623–635.

Andersen, P. K., Borgan, Ø., Gill, R. D. and Keiding, N. (1993) *Statistical Models based on Counting Processes*. New York: Springer.

Andersen, P. K. and Gill, R. D. (1982) Cox's regression model for counting processes: a large sample study. *Ann. Statist.*, **10**, 1100–1120.

Becker, N. G. (1984) Estimating population size from capture-recapture experiments in continuous time. *Aust. J. Statist.*, **26**, 1–7.

Becker, N. G. and Heyde, C. C. (1990) Estimating population size from multiple recapture experiments. *Stoch. Processes Appl.*, **36**, 77–83.

Borchers, D. L., Buckland, S. T. and Zucchini, W. (2002) *Estimating Animal Abundance: Closed Populations*. London: Springer.

Borchers, D. L., Stevenson, B. C., Kidney, D., Thomas, L. and Marques, T. A. (2015) A unifying model for capture-recapture and distance sampling surveys of wildlife populations. *J. Am. Statist. Ass.*, **110**, 195–204.

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.

Chao, A. (1987) Estimating the population size for capture-recapture data with unequal catchability. *Biometrics*, **43**, 783–791.

Chao, A. and Lee, S. M. (1993) Estimating population size for continuous-time capture-recapture models via sample coverage. *Biometr. J.*, **35**, 29–45.

Chao, A., Tsay, P. K., Lin, S. H., Shau, W. Y. and Chao, D. Y. (2001) The applications of capture-recapture models to epidemiological data. *Statist. Med.*, **20**, 3123–3157.

Chen, K. (2001) Parametric and semiparametric models for recapture and removal studies: a likelihood approach. *J. R. Statist. Soc.* B, **63**, 607–619.

Chen, S. X. and Lloyd, C. J. (2000) A non-parametric approach to the analysis of two stage mark-recapture experiments. *Biometrika*, **87**, 633–649.

Chen, S. X. and Lloyd, C. J. (2002) Estimation of population size from biased samples using non-parametric binary regression. *Statist. Sin.*, **12**, 505–518.

Cox, D. R. (1972) Regression models and life-tables (with discussion). *J. R. Statist. Soc.* B, **34**, 187–220.

Craig, C. C. (1953) On the utilization of marked specimens in estimating populations of flying insects. *Biometrika*, **40**, 170–176.

Darroch, J. N. (1958) The multiple recapture census: I, Estimation of a closed population. *Biometrika*, **45**, 343–359.

DiCiccio, T. J., Hall, P. and Romano, J. P. (1991) Empirical likelihood is Bartlett-correctable. *Ann. Statist.*, **19**, 1053–1061.

Evans, M. A. and Bonett, D. G. (1994) Bias reduction for multiple recapture estimators of closed population size. *Biometrics*, **50**, 388–395.

Farcomeni, A. and Scacciatelli, D. (2013) Heterogeneity and behavioral response in continuous time capture-recapture, with application to street cannabis use in Italy. *Ann. Appl. Statist.*, **7**, 2293–2314.

Farcomeni, A. and Tardella, L. (2012) Identifiability and inferential issues in capture-recapture experiments with heterogeneous detection probabilities. *Electron. J. Statist.*, **6**, 2602–2626.

Fewster, R. M. and Jupp, P. E. (2009) Inference on population size in binomial detectability models. *Biometrika*, **96**, 805–820.

Hall, P. and La Scala, B. (1990) Methodology and algorithms of empirical likelihood. *Int. Statist. Rev.*, **58**, 109–127.

van der Heijden, P. G. M., Bustami, R., Cruyff, M. J. L. F., Engbersen, G. and van Houwelingen, H. C. (2003) Point and interval estimation of the population size using the truncated Poisson regression model. *Statist. Modllng*, **3**, 305–322.

Holzmann, H., Munk, A. and Zucchini, W. (2006) On identifiability in capture-recapture models. *Biometrics*, **62**, 934–936.

Horvitz, D. G. and Thompson, D. J. (1952) A generalization of sampling without replacement from a finite universe. *J. Am. Statist. Ass.*, **47**, 663–685.

Huggins, R. M. (1989) On the statistical analysis of capture experiments. *Biometrika*, **76**, 133–140.

Huggins, R. and Hwang, W. H. (2011) A review of the use of conditional likelihood in capture-recapture experiments. *Int. Statist. Rev.*, **79**, 385–400.

Hwang, W. H. and Chao, A. (2002) Continuous-time capture-recapture models with covariates. *Statist. Sin.*, **12**, 1115–1131.

Hwang, W. H. and Huang, S. Y. H. (2003) Estimation in capture-recapture models when covariates are subject to measurement errors. *Biometrics*, **59**, 1113–1122.

Lin, D. Y. and Yip, P. S. F. (1999) Parametric regression models for continuous time removal and recapture studies. *J. R. Statist. Soc.* B, **61**, 401–411.

Link, W. A. (2003) Nonidentifiability of population size from capture-recapture data with heterogeneous detection probabilities. *Biometrics*, **59**, 1123–1130.

Liu, Y., Li, P. and Qin, J. (2017) Maximum empirical likelihood estimation for abundance in a closed population from capture–recapture data. *Biometrika*, **104**, 527–543.

Newey, W. K. and Smith, R. J. (2004) High order properties of GMM and generalized empirical likelihood estimators. *Econometrica*, **72**, 219–255.

Otis, D. L., Burnham, K. P., White, G. C. and Anderson, D. R. (1978) Statistical inference from capture data on closed animal populations. *Wldlif. Monog.*, **62**, 3–135.

Owen, A. B. (1988) Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, **75**, 237–249.

Owen, A. B. (1990) Empirical likelihood ratio confidence regions. *Ann. Statist.*, **18**, 90–120.

Owen, A. B. (2001) *Empirical Likelihood*. New York: Chapman and Hall.

Pollock, K. H. (1991) Modeling capture, recapture, and removal statistics for estimation of demographic parameters for fish and wildlife populations: past, present, and future. *J. Am. Statist. Ass.*, **86**, 225–238.

Pollock, K. H. (2000) Capture-recapture models. *J. Am. Statist. Ass.*, **95**, 293–296.

Schofield, M. R., Barker, R. J. and Gelling, N. (2017) Continuous-time capture–recapture in closed populations. *Biometrics*, **74**, 626–635.
Seber, G. A. F. (1982) *The Estimation of Animal Abundance and Related Parameters*, 2nd edn. London: Arnold.
Stoklosa, J., Hwang, W. H., Wu, S. H. and Huggins, R. (2011) Heterogeneous capture-recapture models with covariates: a partial likelihood approach for closed populations. *Biometrics*, **67**, 1659–1665.
Wilson, K. R. and Anderson, D. R. (1995) Continuous-time capture-recapture population estimation when capture probabilities vary over time. *Environ. Ecol. Statist.*, **2**, 55–69.
Xi, L., Yip, P. S. F. and Watson, R. (2007) A unified likelihood-based approach for estimating population size in continuous-time capture-recapture experiments with frailty. *Biometrics*, **63**, 228–236.
Xu, Y., Liu, L., You, N., Pan, H. and Yip, P. (2007) Estimating population size for a continuous time frailty model with covariates in a capture-recapture study. *Biometrics*, **63**, 917–921.
Yip, P. S. F. and Chao, A. (1996) Estimating population size from capture-recapture studies via sample coverage and estimating functions. *Communs Statist. Stoch. Modls*, **12**, 17–35.
Yip, P. S. F., Fong, D. Y. T. and Wilson, K. (1993) Estimating population size by recapture sampling via estimating function. *Stoch. Modls*, **9**, 179–193.
Yip, P. S. F., Huggins, R. M. and Lin, D. Y. (1996) Inference for capture-recapture experiments in continuous time with variable rates. *Biometrika*, **83**, 477–483.
Yip, P. S. F., Lin, H. Z. and Xi, L. (2005) A semiparametric method for estimating population size for capture-recapture experiments with random covariates in continuous time. *Biometrics*, **61**, 1085–1092.
Yip, P. S. F., Xi, L., Chao, A. and Hwang, W. H. (2000) Estimating the population size with a behavioral response in capture-recapture experiment. *Environ. Ecol. Statist.*, **7**, 405–414.

Supporting information
Additional ‘supporting information’ may be found in the on-line version of this article:

‘Supplementary material for “Full likelihood inference for abundance from continuous-time capture-recapture data”’.