Biometrika (2017), **104**, 3, *pp.* 527–543
Printed in Great Britain
doi: 10.1093/biomet/asx038
Advance Access publication 3 July 2017

Maximum empirical likelihood estimation for abundance in a closed population from capture-recapture data

By YUKUN LIU

School of Statistics, East China Normal University, Shanghai 200241, China
ykliu@sfs.ecnu.edu.cn

PENGFEI LI

Department of Statistics and Actuarial Science, University of Waterloo, 200 University Avenue West, Waterloo, Ontario N2L 3G1, Canada
pengfei.li@uwaterloo.ca

AND JING QIN

National Institute of Allergy and Infectious Diseases, National Institutes of Health, 6700B Rockledge Drive, Bethesda, Maryland 20892, U.S.A.
jingqin@niaid.nih.gov

Summary

Capture-recapture experiments are widely used to collect data needed for estimating the abundance of a closed population. To account for heterogeneity in the capture probabilities, Huggins (1989) and Alho (1990) proposed a semiparametric model in which the capture probabilities are modelled parametrically and the distribution of individual characteristics is left unspecified. A conditional likelihood method was then proposed to obtain point estimates and Wald-type confidence intervals for the abundance. Empirical studies show that the small-sample distribution of the maximum conditional likelihood estimator is strongly skewed to the right, which may produce Wald-type confidence intervals with lower limits that are less than the number of captured individuals or even are negative. In this paper, we propose a full empirical likelihood approach based on Huggins and Alho's model. We show that the null distribution of the empirical likelihood ratio for the abundance is asymptotically chi-squared with one degree of freedom, and that the maximum empirical likelihood estimator achieves semiparametric efficiency. Simulation studies show that the empirical likelihood-based method is superior to the conditional likelihood-based method: its confidence interval has much better coverage, and the maximum empirical likelihood estimator has a smaller mean square error. We analyse three datasets to illustrate the advantages of our empirical likelihood approach.

Some key words: Abundance estimation; Capture-recapture experiment; Dual system estimation; Empirical likelihood.

1. Introduction

In fields such as biology, ecology, demography, epidemiology and reliability, it is important to know the abundance of a species, the size of a closed population, or the number of defects in a system (Borchers et al., 2002, 2015). Mark-recapture or capture-recapture experiments are widely used for this purpose. In these experiments, individuals from the population of interest are captured, marked, and then released. At a later time, after the captured individuals have mixed with the others, another sample is taken. In general, taking more than two capture samples is common in biology and ecology, whereas taking just two is common in demography, epidemiology and reliability.

Mark-recapture or capture-recapture experiments are extensively used when it is not practicable to count all the individuals in the population. The method was originally developed for the estimation of animal abundance, but it has increasingly been applied to the estimation of population parameters for demographic events. For example, the U.S. Census Bureau uses the dual system estimation method, a slightly modified version of the capture-recapture method (Seber, 1982), to estimate the U.S. population (Hogan, 2000). This method produces valid population estimates under certain assumptions. In epidemiological studies, the capture-recapture method is used to estimate the completeness of disease registers. For example, Boden & Ozonoff (2008) used the capture-recapture method to estimate the level of reporting for the two most common U.S. sources of information about nonfatal injuries and illnesses: workers compensation data and the annual Survey of Occupational Injuries and Illnesses conducted by the Bureau of Labor Statistics. Tilling et al. (2001) applied the capture-recapture method with covariate adjustment to estimate the incidence of stroke in south London. In the past decade the method has also become widespread in noninvasive genetic sampling; see Lukacs & Burnham (2005) for a detailed review. It has been used in the context of software inspection (Barnard et al., 2003) to estimate the number of defects in an inspected artefact. This estimate can be used to decide whether the artefact requires reinspection to improve the phase containment of defects, which involves detecting faults in the current software phase rather than allowing them to escape into subsequent phases.

In this paper, we consider statistical inference for the abundance of a species based on capture-recapture data. We take k samples from a closed population. Let N be the abundance, and let $X_1, \ldots, X_N$ be the individuals' characteristics, which are independent and identically distributed and have cumulative distribution function $F(x)$ and probability density function $f(x)$. Let $D = (D_1, \ldots, D_k)^{\mathrm{T}}$ be the capture history of an individual, where $D_j = 1$ if the individual is captured on the jth occasion and $D_j = 0$ otherwise. There is observable population heterogeneity: individuals in different classes have different capture probabilities. To account for this, we adopt the semiparametric model proposed by Huggins (1989) and Alho (1990), in which the probability of capture on occasion j, $g_j(x) = \mathrm{pr}(D_j = 1 \mid X = x)$, is modelled parametrically and the distribution $F(x)$ is left unspecified. Moreover, the D_j are assumed to be independent conditionally on $X = x$. Suppose that n different individuals are observed and their characteristics are $x_1, \ldots, x_n$. Let $d_i = (d_{i1}, \ldots, d_{ik})^{\mathrm{T}}$ be the capture history of the ith observation and let $d_{i+} = \sum_{j=1}^{k} d_{ij}$ be the number of captures on the ith observation. Clearly, $d_{i+} > 0$ for the n observed individuals. We wish to make inference on the abundance N under the semiparametric model of Huggins (1989) and Alho (1990).

Fully parametric methods for estimating N, where the form of $F(x)$ is assumed to be known, have been extensively discussed. Borchers et al. (1998) developed a likelihood framework. Fewster & Jupp (2009) derived the asymptotic properties of the maximum likelihood estimator of N based on the full likelihood and those of the conditional maximum likelihood estimator of N based on the conditional distribution of $x_1, \ldots, x_n$ given n. Semiparametric methods, where $F(\cdot)$ is modelled as a functional parameter, are also available. Huggins (1989) and Alho (1990) proposed an estimator for N based on the conditional likelihood $\prod_{i=1}^{n} \mathrm{pr}(D = d_i \mid d_{i+} > 0, X = x_i)$ under the logistic regression model for $g_j(x)$. Their ideas have been borrowed and extended by many other researchers; see, for example, Borchers et al. (1998) and the references therein. More detailed developments of the parametric and semiparametric approaches can be found in Borchers et al. (2002), Marques & Buckland (2004) and Fewster & Jupp (2009), among others.

Most parametric and semiparametric asymptotic results concern asymptotic normality of the abundance estimator or log abundance estimator; these results are used to construct Wald-type confidence intervals for the abundance. However, even in the simplest case, the small-sample distribution of the maximum conditional likelihood abundance estimator is strongly skewed to the right (Evans et al., 1994). Moreover, in a numerical study Evans & Bonett (1994) found that the lower limit of the Wald-type confidence interval may be less than the number of individuals captured, or even negative. Similar observations have been made in our simulation studies and real-data analysis; see § 3 and 4. These undesirable properties motivate our work.

In this paper, we explore interval estimation for N based on the maximum full likelihood ratio under the semiparametric model of Huggins (1989) and Alho (1990). We propose to use the empirical likelihood, first introduced by Owen (1988, 1990) to mimic the parametric likelihood, since it has many nice properties. Empirical likelihood confidence regions are Bartlett-correctable (DiCiccio et al., 1991), range-preserving, and transformation-respecting (Hall & La Scala, 1990); they do not require estimation of the scale or skewness. Since the two seminal papers by Owen (1988, 1990), empirical likelihood has been applied to biomedical studies, survey sampling and economic research; see Owen (2001) and Newey & Smith (2004).

Although empirical likelihood has been used widely, as far as we know it has never been applied to abundance estimation under Huggins and Alho's semiparametric model. In our set-up, the semiparametric full likelihood contains three terms; see § 2·1. The first term involves the binomial likelihood for N, the second term is the conditional likelihood, and the third term is the marginal empirical likelihood of the covariate information. Hence, the conditional likelihood is only one component of the full likelihood. We plan to use the full likelihood, which combines all three terms, to construct confidence intervals for the abundance N based on the empirical likelihood ratio.

Developing the asymptotic properties of the empirical likelihood ratio for the abundance is very challenging. Standard methods and results from maximum empirical likelihood theory are not directly applicable because the support of n depends on the parameter N, which violates the regularity conditions. Furthermore, we have to deal with the binomial coefficient for the abundance parameter estimation in addition to selection-biased sampling. In Huggins and Alho's semiparametric set-up, we are able to show that the empirical likelihood ratio for the abundance N has an asymptotic chi-squared distribution with one degree of freedom. Simulations indicate that the empirical likelihood confidence interval for N has much better coverage than Wald-type confidence intervals based on the maximum conditional likelihood abundance estimator. Furthermore, we have found that the maximum empirical likelihood estimator of N has a smaller mean square error than the maximum conditional likelihood estimator of N. For convenience of presentation, all proofs are given in the Supplementary Material.

2. Empirical likelihood inference

2·1. *Model set-up and empirical likelihood*

Following Huggins (1989) and Alho (1990), we model the probability of capture on occasion j $(j = 1, \dots, k)$ by the logistic regression model $g_j(x) = g(x, \beta_j)$, where

$$g(x, \beta_j) = \frac{\exp\{\beta_j^{\mathrm{T}} q(x)\}}{1 + \exp\{\beta_j^{\mathrm{T}} q(x)\}}, \tag{1}$$

with $q(x)$ being a prespecified b-variate function whose first component is 1. For example, when x is a scalar, we may choose $q(x)$ to be $(1,x)^{\mathrm{T}}$ or $(1,x,x^2)^{\mathrm{T}}$. Model (1) is an M_{th} model (Otis et al., 1978; Seber, 1982; Borchers et al., 2002) because the capture probability varies not only from individual to individual but also from capture occasion to capture occasion.

Let $\beta^{\mathrm{T}} = (\beta_1^{\mathrm{T}}, \ldots, \beta_k^{\mathrm{T}})$ and define $\phi(x,\beta) = \prod_{j=1}^{k}\{1 - g(x,\beta_j)\}$, which is the probability that an ideal observation X is not observed on any of the k occasions given $X = x$. Then $\alpha = \int \phi(x,\beta)\,\mathrm{d}F(x)$ is the probability that an ideal observation is not observed on any of the k occasions. To ensure the identifiability of (N,β,α), the conditions that $k \geqslant 2$ and the components of $q(x)$ are linearly independent are assumed throughout the paper.

We now develop the full likelihood of (N,β,α,F), which is the product of three components: the likelihood from n, the likelihood from $d_1,\ldots,d_n$ conditional on $x_1,\ldots,x_n$ and given that the n individuals have been captured at least once, and the likelihood from $x_1,\ldots,x_n$ given that the n individuals have been captured at least once.

First, note that $n \sim \mathrm{Bi}(N, 1-\alpha)$. Therefore its contribution to the likelihood is

$$\binom{N}{n}(1-\alpha)^n\alpha^{N-n} = \frac{\Gamma(N+1)}{\Gamma(n+1)\Gamma(N-n+1)}(1-\alpha)^n\alpha^{N-n}, \tag{2}$$

where $\Gamma(\cdot)$ is the gamma function. Second, given that the ith individual has been captured at least once and has covariate x_i, the conditional probability of observing the capture history of the ith individual is

$$\begin{aligned}\mathrm{pr}(D = d_i \mid d_{i+} > 0, X = x_i) &= \frac{\mathrm{pr}(D = d_i, d_{i+} > 0 \mid X = x_i)}{\mathrm{pr}(d_{i+} > 0 \mid X = x_i)}\\ &= \frac{\mathrm{pr}(D = d_i \mid X = x_i)}{\mathrm{pr}(d_{i+} > 0 \mid X = x_i)}\\ &= \frac{\prod_{j=1}^{k}\{1 - g(x_i,\beta_j)\}^{1-d_{ij}}\{g(x_i,\beta_j)\}^{d_{ij}}}{1 - \phi(x_i,\beta)}.\end{aligned}$$

Hence the likelihood, known as the conditional likelihood (Alho, 1990; Huggins, 1989), from $d_1,\ldots,d_n$ conditional on $x_1,\ldots,x_n$ and given that the n individuals have been captured at least once is

$$L_c(\beta) = \prod_{i=1}^{n}\frac{\prod_{j=1}^{k}\{1 - g(x_i,\beta_j)\}^{1-d_{ij}}\{g(x_i,\beta_j)\}^{d_{ij}}}{1 - \phi(x_i,\beta)}. \tag{3}$$

Third, given that the ith individual has been captured at least once, the conditional probability of observing x_i is

$$\mathrm{pr}(X = x_i \mid d_{i+} > 0) = \frac{\mathrm{pr}(d_{i+} > 0 \mid X = x_i)\,\mathrm{pr}(X = x_i)}{\mathrm{pr}(d_{i+} > 0)} = \frac{\{1 - \phi(x_i,\beta)\}\,\mathrm{d}F(x_i)}{1 - \alpha}.$$

Therefore, the likelihood from $x_1,\ldots,x_n$ given that the n individuals have been captured at least once is

$$\prod_{i=1}^{n}\frac{\{1 - \phi(x_i,\beta)\}\,\mathrm{d}F(x_i)}{1 - \alpha}. \tag{4}$$

Upon combining (2)–(4), the full likelihood function of (N, β, α, F) is

$$\frac{\Gamma(N+1)}{\Gamma(n+1)\Gamma(N-n+1)}\alpha^{N-n} \times \prod_{i=1}^{n}\left[\mathrm{d}F(x_i)\prod_{j=1}^{k}\{1-g(x_i,\beta_j)\}^{1-d_{ij}}\{g(x_i,\beta_j)\}^{d_{ij}}\right]. \quad (5)$$

As pointed out by Fewster & Jupp (2009), although N is necessarily a positive integer, the likelihood function (5) makes sense for any positive N, and there is negligible error in treating N as continuous for the asymptotics reported in this paper. Hence we will treat N as continuous.

Let $p_i = \mathrm{d}F(x_i)$. The empirical loglikelihood (Owen, 2001), up to a constant not dependent on the unknown parameters, is

$$\log\left\{\frac{\Gamma(N+1)}{\Gamma(N-n+1)}\right\} + (N-n)\log\alpha + \sum_{i=1}^{n}\log p_i$$
$$+\sum_{i=1}^{n}\sum_{j=1}^{k}\left[d_{ij}\log g(x_i,\beta_j) + (1-d_{ij})\log\{1-g(x_i,\beta_j)\}\right],$$

where the feasible p_i $(i = 1, \ldots, n)$ satisfy

$$p_i \geqslant 0, \quad \sum_{i=1}^{n} p_i = 1, \quad \sum_{i=1}^{n} p_i\{\phi(x_i,\beta)-\alpha\} = 0.$$

The above formulation ignores ties in $x_1, \ldots, x_n$. If ties occur, we should interpret p_i as $\mathrm{d}F(x_i)/m_i$, where m_i is the number of times that x_i appears in $x_1, \ldots, x_n$. As discussed in Owen (2001, § 2.3), the resulting probability weights (6) and profile empirical loglikelihood (7) do not change.

Given (β, α), in general the empirical loglikelihood achieves its maximum when

$$p_i = \frac{1}{n}\frac{1}{1+\lambda\{\phi(x_i,\beta)-\alpha\}}, \quad (6)$$

where λ satisfies

$$\sum_{i=1}^{n}\frac{\phi(x_i,\beta)-\alpha}{1+\lambda\{\phi(x_i,\beta)-\alpha\}} = 0.$$

When we profile out the p_i, the profile empirical loglikelihood of (N, β, α) is

$$\ell(N,\beta,\alpha) = \log\left\{\frac{\Gamma(N+1)}{\Gamma(N-n+1)}\right\} + (N-n)\log\alpha - \sum_{i=1}^{n}\log\left[1+\lambda\{\phi(x_i,\beta)-\alpha\}\right]$$
$$+\sum_{i=1}^{n}\sum_{j=1}^{k}\left[d_{ij}\log g(x_i,\beta_j) + (1-d_{ij})\log\{1-g(x_i,\beta_j)\}\right]. \quad (7)$$

The maximum empirical likelihood estimators of (N, β, α) are

$$(\hat{N}, \hat{\beta}, \hat{\alpha}) = \arg\max_{N,\beta,\alpha} \ell(N,\beta,\alpha). \quad (8)$$

The empirical likelihood ratio functions of (N, β, α) and N are

$$R(N,\beta,\alpha) = 2\left\{\sup_{N,\beta,\alpha} \ell(N,\beta,\alpha) - \ell(N,\beta,\alpha)\right\} = 2\{\ell(\hat{N},\hat{\beta},\hat{\alpha}) - \ell(N,\beta,\alpha)\}, \tag{9}$$

$$R'(N) = 2\left\{\sup_{N,\beta,\alpha} \ell(N,\beta,\alpha) - \sup_{\beta,\alpha} \ell(N,\beta,\alpha)\right\} = 2\{\ell(\hat{N},\hat{\beta},\hat{\alpha}) - \ell(N,\hat{\beta}_N,\hat{\alpha}_N)\}, \tag{10}$$

where $(\hat{\beta}_N, \hat{\alpha}_N) = \arg\max_{\beta,\alpha} \ell(N,\beta,\alpha)$ given N.

2·2. *Asymptotic properties*: *general case*

In this section, we establish the limiting behaviour of the maximum empirical likelihood estimators and the empirical likelihood ratios when no constraints are imposed on the β_j.

We begin by defining some notation. Let N_0, $\beta_0 = (\beta_{10}^{\mathrm{T}}, \ldots, \beta_{k0}^{\mathrm{T}})^{\mathrm{T}}$ and α_0 be the true values of N, β and α, respectively. Write $G_1(x) = \{g(x,\beta_{10}), \ldots, g(x,\beta_{k0})\}^{\mathrm{T}}$, $G_2(x) = \mathrm{diag}\{G_1(x)\}$ and $\phi_* = E[\{1-\phi(X,\beta_0)\}^{-1}]$. We use $\otimes$ to denote the Kronecker product operator. The following matrix W is closely related to the asymptotic variance matrix of the maximum empirical likelihood estimators:

$$W = \begin{pmatrix} -V_{11} & 0 & -V_{13} \\ 0 & -V_{22} + V_{24}V_{44}^{-1}V_{42} & -V_{23} + V_{24}V_{44}^{-1}V_{43} \\ -V_{31} & -V_{32} + V_{34}V_{44}^{-1}V_{42} & -V_{33} + V_{34}V_{44}^{-1}V_{43} \end{pmatrix}, \tag{11}$$

where

$$V_{11} = 1 - \alpha_0^{-1}, \quad V_{13} = \alpha_0^{-1},$$

$$V_{22} = E\left[\left\{\frac{\phi(X,\beta_0)}{1-\phi(X,\beta_0)} G_1(X)G_1^{\mathrm{T}}(X) + G_2^2(X) - G_2(X)\right\} \otimes \{q(X)q(X)^{\mathrm{T}}\}\right],$$

$$V_{23} = V_{32}^{\mathrm{T}} = E\left\{\frac{\phi(X,\beta_0)}{1-\phi(X,\beta_0)} G_1(X) \otimes q(X)\right\}, \quad V_{24} = V_{42}^{\mathrm{T}} = (1-\alpha_0)^2 V_{23},$$

$$V_{33} = \phi_* - \alpha_0^{-1}, \quad V_{34} = V_{43} = (1-\alpha_0)^2\phi_*, \quad V_{44} = (1-\alpha_0)^4\phi_* - (1-\alpha_0)^3.$$

We refer to the Supplementary Material for more discussion on V_{ij}.

THEOREM 1. *Assume that the support of X is compact, the capture probability function $g_j(x)$ is $g(x,\beta_j)$ as defined in* (1)*, and the vector-valued function $q(x)$ is b-variate with linearly independent components. Let (N_0, β_0, α_0) be the true value of (N, β, α) with $\alpha_0 \in (0,1)$. If W defined in* (11) *is nonsingular, then as $N_0 \to \infty$:*

(i) ${N_0}^{1/2}\{\log(\hat{N}/N_0),\ \hat{\beta}^{\mathrm{T}} - \beta_0^{\mathrm{T}},\ \hat{\alpha} - \alpha_0\}^{\mathrm{T}} \to N(0, W^{-1})$ *in distribution;*
(ii) $R(N_0, \beta_0, \alpha_0) \to \chi^2_{bk+2}$ *in distribution and* $R'(N_0) \to \chi^2_1$ *in distribution, where k is the number of capture occasions.*

Based on the limiting chi-squared distribution of the empirical likelihood ratio in Theorem 1, we may construct a confidence interval for N at level $1-a$ as

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

where $\chi^2_{1,1-a}$ is the $(1-a)$th quantile of the χ^2_1 distribution. Theorem 1 guarantees that $\mathcal{I}_1$ has asymptotically correct coverage probability.

While empirical likelihood estimation in this setting is new, maximum conditional likelihood estimation has been investigated in the literature (Huggins, 1989; Alho, 1990). Denote by $\ell_c(\beta) = \log L_c(\beta)$ the conditional loglikelihood given the observed data, where $L_c(\beta)$ defined in (3) is the conditional likelihood. The maximum conditional likelihood estimator of N is defined as

$$\tilde{N} = \sum_{i=1}^{n} \frac{1}{1-\phi(x_i,\tilde{\beta})},$$

where $\tilde{\beta} = \arg\max_\beta \ell_c(\beta)$.

THEOREM 2. *Under the assumptions in Theorem 1, as $N_0 \to \infty$:*

(i) $\hat{N} - \tilde{N} = O_p(1)$;
(ii) $(\hat{N} - N_0)/N_0^{1/2}$, $(\tilde{N} - N_0)/N_0^{1/2}$, $N_0^{1/2}\log(\hat{N}/N_0)$ *and* $N_0^{1/2}\log(\tilde{N}/N_0)$ *all converge in distribution to* $N(0,\sigma^2)$, *where* $\sigma^2 = \phi_* - 1 - V_{32}V_{22}^{-1}V_{23}$.

Theorem 2 is analogous to Theorems 1 and 2 in Fewster & Jupp (2009); see their equations (A10) and (A17). It shows a close relationship between the maximum empirical likelihood estimator $\hat{N}$ and the maximum conditional likelihood estimator $\tilde{N}$ under Huggins and Alho's semiparametric model. Fewster & Jupp (2009) presented similar results under fully parametric models. In the Supplementary Material, we further show that the maximum empirical likelihood estimator $\hat{N}$ is semiparametric efficient in the sense that its asymptotic variance σ^2 is the supremum of the asymptotic variances of the maximum parametric likelihood estimator of N under all parametric submodels.

Fewster & Jupp (2013) proposed three types of confidence intervals for N: the likelihood ratio, score and Wald intervals under fully parametric models. The empirical likelihood ratio-based interval $\mathcal{I}_1$ for N has been discussed above. Based on the profile empirical loglikelihood, we can construct a score test-based confidence interval for N. However, the profile empirical loglikelihood for N does not have a closed form. We do not currently have a simple way to implement the score test-based confidence interval for N based on the profile empirical loglikelihood, so we do not consider it in our numerical study. A Wald-type confidence interval based on $\hat{N}$ is not needed since it requires an additional variance estimate compared with $\mathcal{I}_1$. For the conditional maximum likelihood method, the conditional loglikelihood under Huggins and Alho's semiparametric model does not involve N; hence it cannot be directly used to construct likelihood ratio-based or score test-based confidence intervals. Wald-type confidence intervals based on $\tilde{N}$ are the only option in this case.

Wald-type interval estimators of N need a consistent estimator of σ^2. Based on the form of σ^2 in Theorem 2, an estimator of σ^2 is

$$\hat{\sigma}^2 = \hat{\phi}_* - 1 - \hat{V}_{32}\hat{V}_{22}^{-1}\hat{V}_{23}, \tag{12}$$

where $\hat{\phi}_* = \tilde{N}^{-1}\sum_{i=1}^{n}\{1-\phi(x_i,\tilde{\beta})\}^{-2}$ and

$$\hat{V}_{23} = \hat{V}_{32}^{\mathrm{T}} = \tilde{N}^{-1}\sum_{i=1}^{n} \frac{\phi(x_i,\tilde{\beta})}{\{1-\phi(x_i,\tilde{\beta})\}^2} G_1(x_i,\tilde{\beta}) \otimes q(x_i),$$

$$\hat{V}_{22} = -\tilde{N}^{-1} \sum_{i=1}^{n} \left[\left\{ d_i - \frac{G_1(x_i, \tilde{\beta})}{1 - \phi(x_i, \tilde{\beta})} \right\} \left\{ d_i - \frac{G_1(x_i, \tilde{\beta})}{1 - \phi(x_i, \tilde{\beta})} \right\}^{\mathrm{T}} \right] \otimes \{q(x_i) q(x_i)^{\mathrm{T}}\}.$$

In the Supplementary Material, we show that $\hat{\sigma}^2$ is a root-N_0-consistent estimator of σ^2. Note that $\hat{\phi}_*$, $\hat{V}_{23}$ and $\hat{V}_{22}$ are used to construct the Wald-type interval estimators of N based on $\tilde{N}$, but not for the proposed $\mathcal{I}_1$. Hence, we use $(\tilde{\beta}, \tilde{N})$ rather than $(\hat{\beta}, \hat{N})$ in $\hat{\phi}_*$, $\hat{V}_{23}$ and $\hat{V}_{22}$.

Because of the asymptotic normality in Theorem 2 and the consistency of $\hat{\sigma}^2$, both $(\tilde{N} - N_0)/(\tilde{N}^{1/2}\hat{\sigma})$ and $\tilde{N}^{1/2} \log(\tilde{N}/N_0)/\hat{\sigma}$ are asymptotically pivotal, which leads to two Wald-type confidence intervals for N based on the conditional likelihood:

$$\mathcal{I}_2 = [\tilde{N} - z_{1-a/2}\tilde{N}^{1/2}\hat{\sigma},\ \tilde{N} + z_{1-a/2}\tilde{N}^{1/2}\hat{\sigma}],$$
$$\mathcal{I}_3 = \left[\exp\{\log(\tilde{N}) - z_{1-a/2}\tilde{N}^{-1/2}\hat{\sigma}\},\ \exp\{\log(\tilde{N}) + z_{1-a/2}\tilde{N}^{-1/2}\hat{\sigma}\}\right],$$

where $z_{1-a/2}$ is the $(1 - a/2)$th quantile of the standard normal distribution.

An alternative confidence interval for N uses the transformation $\log(\tilde{N} - n)$, which was attributed to Burnham by Chao (1987). Using the results in Theorem 2, we can show that

$$C(N_0; \tilde{N}) = \frac{\log(\tilde{N} - n) - \log(N_0 - n)}{[\log\{1 + \tilde{N}\hat{\sigma}^2/(\tilde{N} - n)^2\}]^{1/2}} \tag{13}$$

is asymptotically distributed as $N(0, 1)$. Hence, the third Wald-type confidence interval for N based on the conditional likelihood is $\mathcal{I}_4 = \{N : |C(N; \tilde{N})| \leqslant z_{1-a/2}\}$. An advantage of $\mathcal{I}_4$ is that its lower limit is guaranteed to be greater than the number of captured individuals n. In § 3 we will use simulation to compare the performance of $\mathcal{I}_1, \ldots, \mathcal{I}_4$.

2·3. *Asymptotic properties: special case where the β_j are all equal*

When the β_j are all equal, $\phi(x, \beta)$ reduces to $\phi_s(x, \beta_s) = \{1 - g(x, \beta_s)\}^k$, where β_s denotes the common value of the β_j. This model is called the M_{h} model; see, for example, Borchers et al. (2002) and Stoklosa et al. (2011). In this situation, the profile empirical loglikelihood $\ell_s(N, \beta_s, \alpha)$ can be directly obtained from the profile empirical loglikelihood in (7):

$$\ell_s(N, \beta_s, \alpha) = \log\left\{\frac{\Gamma(N+1)}{\Gamma(N-n+1)}\right\} + (N-n)\log\alpha - \sum_{i=1}^{n} \log\left[1 + \lambda\{\phi_s(x_i, \beta_s) - \alpha\}\right]$$
$$+ \sum_{i=1}^{n} \left[d_{i+} \log g(x_i, \beta_s) + (k - d_{i+}) \log\{1 - g(x_i, \beta_s)\}\right],$$

where λ is the solution to

$$\sum_{i=1}^{n} \frac{\phi_s(x_i, \beta_s) - \alpha}{1 + \lambda\{\phi_s(x_i, \beta_s) - \alpha\}} = 0. \tag{14}$$

With the profile empirical loglikelihood $\ell_s(N, \beta_s, \alpha)$, we define the maximum empirical likelihood estimators $(\hat{N}_s, \hat{\beta}_s, \hat{\alpha}_s)$ of (N, β_s, α), the empirical likelihood ratio $R_s(N, \beta_s, \alpha)$ for (N, β_s, α) and the empirical likelihood ratio $R'_s(N)$ for N similarly to the definitions of $(\hat{N}, \hat{\beta}, \hat{\alpha})$, $R(N, \beta, \alpha)$

and $R'(N)$ in (8), (9) and (10). To present the asymptotics, we define a new W matrix, namely W_s, which is W in (11) with ϕ_*, V_{23}, V_{24} and V_{22} replaced by $\phi_{s*} = E[\{1-\phi_s(X,\beta_{s0})\}^{-1}]$,

$$V_{23s} = E\left\{\frac{\phi_s(X,\beta_{s0})}{1-\phi_s(X,\beta_{s0})}\, kg(X,\beta_{s0})q(X)\right\}, \quad V_{24s} = (1-\alpha_0)^2 V_{23s}$$

and

$$V_{22s} = E\left[\left\{\frac{\phi_s(X,\beta_{s0})}{1-\phi_s(X,\beta_{s0})}\, k^2 g^2(X,\beta_0) + kg^2(X,\beta_0) - kg(X,\beta_0)\right\} q(X)q(X)^{\mathrm{T}}\right].$$

Here $(N_0, \beta_{s0}, \alpha_0)$ is the true value of (N, β_s, α).

Corollary 1. *Assume that the support of X is compact and the capture probability function is $g_j(x) = g(x,\beta_s)$ with $q(x)$ as in Theorem 1. Let $(N_0, \beta_{s0}, \alpha_0)$ be the true value of (N, β_s, α). If W_s defined above is nonsingular, then as $N_0 \to \infty$:*

(i) ${N_0}^{1/2}\{\log(\hat{N}_s/N_0),\ \hat{\beta}_s^{\mathrm{T}} - \beta_{s0}^{\mathrm{T}},\ \hat{\alpha}_s - \alpha_0\}^{\mathrm{T}} \to N(0, W_s^{-1})$ *in distribution;*
(ii) $R_s(N_0, \beta_{s0}, \alpha_0) \to \chi^2_{b+2}$ *in distribution and* $R'_s(N_0) \to \chi^2_1$ *in distribution.*

Given the observations, the conditional loglikelihood is

$$\ell_{cs}(\beta_s) = \sum_{i=1}^{n} \big[d_{i+} \log g(x_i,\beta_s) + (k-d_{i+}) \log\{1-g(x_i,\beta_s)\}\big] - \sum_{i=1}^{n} \log\{1-\phi_s(x_i,\beta_s)\}.$$

Similarly to Huggins (1989) and Alho (1990), we define the maximum conditional likelihood estimator of N as

$$\tilde{N}_s = \sum_{i=1}^{n} \frac{1}{1-\phi_s(x_i,\tilde{\beta}_s)},$$

where $\tilde{\beta}_s = \arg\max_{\beta_s} \ell_{cs}(\beta_s)$. The following corollary is equivalent to Theorem 2 when the β_j are all equal.

Corollary 2. *Under the assumptions in Corollary 1, as $N_0 \to \infty$:*

(i) $\hat{N}_s - \tilde{N}_s = O_{\mathrm{p}}(1)$*;*
(ii) $(\hat{N}_s - N_0)/N_0^{1/2}$, $(\tilde{N}_s - N_0)/N_0^{1/2}$, ${N_0}^{1/2}\log(\hat{N}_s/N_0)$ *and* ${N_0}^{1/2}\log(\tilde{N}_s/N_0)$ *all converge in distribution to* $N(0,\sigma_s^2)$*, where* $\sigma_s^2 = \phi_{s*} - 1 - V_{32s}V_{22s}^{-1}V_{23s}$.

Similarly to $\hat{\sigma}^2$ in (12), a consistent estimator of σ_s^2 can be constructed as

$$\hat{\sigma}_s^2 = \hat{\phi}_{s*} - 1 - \hat{V}_{32s}\hat{V}_{22s}^{-1}\hat{V}_{32s}^{\mathrm{T}}, \tag{15}$$

where $\hat{\phi}_{s*} = \tilde{N}_s^{-1}\sum_{i=1}^{n}\{1-\phi_s(x_i,\tilde{\beta}_s)\}^{-2}$ and

$$\hat{V}_{23s} = \hat{V}_{32s}^{\mathrm{T}} = \tilde{N}_s^{-1}\sum_{i=1}^{n} \frac{\phi_s(x_i,\tilde{\beta}_s)}{\{1-\phi_s(x_i,\tilde{\beta}_s)\}^2}\, kg(x_i,\tilde{\beta}_s)q(x_i),$$

$$\hat{V}_{22s} = -\tilde{N}_s^{-1}\sum_{i=1}^{n}\left\{d_{i+} - \frac{kg(x_i,\tilde{\beta}_s)}{1-\phi_s(x_i,\tilde{\beta}_s)}\right\}^2 q(x_i)q(x_i)^{\mathrm{T}}.$$

It can be shown that $\hat{\sigma}_s^2$ is a root-N_0-consistent estimator of σ_s^2.

The results in Corollaries 1 and 2 suggest four confidence intervals for N, which are similar to $\mathcal{I}_1, \ldots, \mathcal{I}_4$:

$$\mathcal{I}_{1s} = \{N : R_s'(N) \leqslant \chi^2_{1,1-a}\},$$
$$\mathcal{I}_{2s} = \big[\tilde{N}_s - z_{1-a/2}\tilde{N}_s^{1/2}\hat{\sigma}_s,\ \tilde{N}_s + \tilde{z}_{1-a/2}N_s^{1/2}\hat{\sigma}_s\big],$$
$$\mathcal{I}_{3s} = \big[\exp\{\log(\tilde{N}_s) - z_{1-a/2}\tilde{N}_s^{-1/2}\hat{\sigma}_s\},\ \exp\{\log(\tilde{N}_s) + z_{1-a/2}\tilde{N}_s^{-1/2}\hat{\sigma}_s\}\big],$$
$$\mathcal{I}_{4s} = \{N : |C_s(N; \tilde{N}_s)| \leqslant z_{1-a/2}\},$$

where $C_s(N; \tilde{N}_s)$ is $C(N; \tilde{N}_s)$ in (13) with $\hat{\sigma}^2$ replaced by $\hat{\sigma}_s^2$.

3. SIMULATION STUDY

In this section we investigate three aspects of the finite-sample performance of the proposed empirical likelihood inference method. We study whether the χ_1^2 distribution provides a good approximation to the finite-sample distribution of the empirical likelihood ratio statistic for N and whether normal distributions provide good approximations to the finite-sample distributions of the maximum conditional likelihood estimator of N. We compare the maximum empirical likelihood estimator and the maximum conditional likelihood estimator of N. We compare four confidence intervals for N, based on the empirical likelihood ratio calibrated by the limiting χ_1^2 distribution, $\mathcal{I}_1$ or $\mathcal{I}_{1s}$, and the three Wald-type confidence intervals $\mathcal{I}_2, \mathcal{I}_3, \mathcal{I}_4$ or $\mathcal{I}_{2s}, \mathcal{I}_{3s}, \mathcal{I}_{4s}$ based on the maximum conditional likelihood estimator of N. We calculate two mean square errors to evaluate the goodness of a generic estimator $\breve{N}$ of N,

$$\text{MSE}_1(\breve{N}) = (\breve{N} - N_0)^2/N_0, \quad \text{MSE}_2(\breve{N}) = N_0\{\log(\breve{N}/N_0)\}^2.$$

We perform simulations for both the general case and the special case where the β_j are all equal. The numerical procedure for implementing the empirical likelihood-based methods is discussed in the Supplementary Material.

For all our simulations, the number of repetitions is 2000. We fix the population size at $N_0 = 200$ or 400 in both cases. Results for $N_0 = 100$ and 150 are presented in the Supplementary Material. For the interval estimation of N, we present only the two-sided coverage probability at the nominal level 95%. The one-tailed coverage probabilities of the signed square root of the empirical likelihood ratio-based confidence interval and the three Wald-type confidence intervals are presented in the Supplementary Material.

We first consider the general case. We set the number of capture occasions to $k = 2$ or 3 and generate data from the following two scenarios.

Scenario 1. The covariate X is univariate and follows the standard normal distribution. The capture probability function on the jth occasion is $g(x, \beta_j)$ in (1) with the true $q(x)$ being $q_{01}(x) = (1, x)^{\mathrm{T}}$. When $k = 3$, we set the true value of β to $\beta_0 = (0, -3, -1, -2, -2, 1)^{\mathrm{T}}$, and the first four components of β_0 are taken as the true value of β for $k = 2$.

Scenario 2. The covariate $X = (X_1, X_2)^{\mathrm{T}}$ is bivariate, where X_1 follows the standard normal distribution and X_2 follows the Bernoulli distribution with success probability 0·5, and the capture probability function on the jth occasion is $g(x, \beta_j)$ with the true $q(x)$ being $q_{02}(x) = (1, x_1, x_2)^{\mathrm{T}}$. We choose a binary X_2 to mimic a discrete characteristic,

Table 1. *Averages $\bar{n}$ of sample sizes, two types of mean square errors of $\hat{N}$ and $\tilde{N}$, and coverage probabilities* (%) *of* $\mathcal{I}_1, \ldots, \mathcal{I}_4$ *at nominal level* 95% *under Scenarios* 1 *and* 2

				MSE_1		MSE_2		Level: 95%			
Scenario	N_0	k	$\bar{n}$	$\hat{N}$	$\tilde{N}$	$\hat{N}$	$\tilde{N}$	$\mathcal{I}_1$	$\mathcal{I}_2$	$\mathcal{I}_3$	$\mathcal{I}_4$
1	200	2	115	275	331	37	42	92·7	86·2	88·8	90·9
	200	3	136	13	17	8	9	93·2	91·5	92·8	94·2
	400	2	229	151	171	32	35	92·1	87·7	89·1	91·3
	400	3	271	8	9	7	7	93·2	92·1	93·2	94·2
2	200	2	111	277	329	40	44	92·7	86·7	89·4	91·9
	200	3	134	9	11	6	7	94·8	93·5	94·4	95·5
	400	2	222	155	179	37	40	93·0	89·9	91·6	92·7
	400	3	268	6	7	5	6	95·7	94·2	94·8	95·8

such as the sex, of an individual. When $k = 3$, we set the true value of β to $\beta_0 = (0{\cdot}1, -2{\cdot}5, -0{\cdot}15, -1{\cdot}5, -1{\cdot}5, -0{\cdot}2, -0{\cdot}5, -0{\cdot}8, -0{\cdot}1)^{\mathrm{T}}$, and the first six components of this vector are taken as the true value of β for $k = 2$.

Under Scenario 1, the probability of overall capture is $1 - \alpha_0 = 0{\cdot}573$ when $k = 2$ and $0{\cdot}676$ when $k = 3$. Under Scenario 2, these probabilities are $0{\cdot}556$ and $0{\cdot}670$ when $k = 2$ and 3, respectively. Recall that α_0 represents the overall probability of noncapture rather than capture. To implement our method and the conditional likelihood method, we set $q(x)$ in $g(x, \beta_j)$ to $q_{01}(x)$ for Scenario 1 and $q_{02}(x)$ for Scenario 2. Table 1 gives the averages $\bar{n}$ of the sample sizes, the MSE_1 and MSE_2 values for both the proposed maximum empirical likelihood estimator $\hat{N}$ and the maximum conditional likelihood estimator $\tilde{N}$, and the simulated coverage probabilities of $\mathcal{I}_1, \ldots, \mathcal{I}_4$ for the abundance N at the nominal level 95% under Scenarios 1 and 2.

As expected, $\bar{n}$ is very close to $N_0(1 - \alpha_0)$ in every case. The proposed maximum empirical likelihood estimator $\hat{N}$ has smaller mean square errors than the maximum conditional likelihood estimator $\tilde{N}$. As N_0 increases from 200 to 400 or as k varies from 2 to 3, both $\hat{N}$ and $\tilde{N}$ become more accurate. In terms of the coverage precision, the empirical likelihood ratio-based confidence interval $\mathcal{I}_1$ has a clear advantage over the Wald-type confidence intervals $\mathcal{I}_2$ and $\mathcal{I}_3$, and it has a moderate advantage over $\mathcal{I}_4$ in Scenario 1 with $N_0 = 200$ and $k = 2$. The gains of $\mathcal{I}_1$ in coverage probability range from 2% to 6%. We have similar findings for Scenario 2 with $N_0 = 200$ and $k = 2$. As N_0 varies from 200 to 400 or as k varies from 2 to 3, $\mathcal{I}_1$ has quite stable coverage probabilities, while the coverage probabilities of $\mathcal{I}_2$, $\mathcal{I}_3$ and $\mathcal{I}_4$ increase. In terms of coverage accuracy, $\mathcal{I}_2$ is uniformly worse than $\mathcal{I}_3$, and $\mathcal{I}_4$ is uniformly better than $\mathcal{I}_3$. This indicates that the log transformation on $\tilde{N}$ increases the coverage probabilities of the Wald-type confidence intervals so that they become close to the nominal levels, while the log transformation on $\tilde{N} - n$ brings the coverage probabilities of the Wald-type confidence intervals closer to the nominal level.

In the Supplementary Material we present additional simulation results under Scenarios 1 and 2. We summarize our findings as follows. First, the results indicate that the distribution of the empirical likelihood ratio is quite close to χ^2_1, and the distributions of $(\tilde{N} - N_0)/(\tilde{N}^{1/2}\hat{\sigma})$ and $\tilde{N}^{1/2} \log(\tilde{N}/N_0)/\hat{\sigma}$ are not close to normal. The results also show that the distribution of $C(N_0; \tilde{N})$ is quite close to normal. These observations may explain why the empirical likelihood ratio-based confidence intervals $\mathcal{I}_1$ always have more accurate coverage probabilities than the Wald-type confidence intervals $\mathcal{I}_2$ and $\mathcal{I}_3$, but have only a slight advantage over $\mathcal{I}_4$. Second, we observe that $\mathcal{I}_1$ has slightly longer length than I_2 and I_3 but much better coverage accuracy. Further, $\mathcal{I}_1$ in general has shorter length than $\mathcal{I}_4$ but better or comparable coverage accuracy.

Table 2. *Averages $\bar{n}$ of sample sizes, two types of mean square errors of $\hat{N}$ and $\tilde{N}$, and coverage probabilities* (%) *of $\mathcal{I}_{1s}, \ldots, \mathcal{I}_{4s}$ at nominal level* 95% *under Scenarios* 3 *and* 4

				MSE$_1$		MSE$_2$		Level: 95%			
Scenario	N_0	k	$\bar{n}$	$\hat{N}_s$	$\tilde{N}_s$	$\hat{N}_s$	$\tilde{N}_s$	$\mathcal{I}_{1s}$	$\mathcal{I}_{2s}$	$\mathcal{I}_{3s}$	$\mathcal{I}_{4s}$
3	200	2	96	989	1506	73	100	93·7	84·0	87·0	87·3
	200	8	146	53	69	10	13	91·4	84·8	86·6	91·7
	400	2	192	1364	1829	120	149	92·6	84·4	87·4	87·5
	400	8	293	10	14	7	9	92·8	86·8	88·7	93·4
4	200	2	122	304	369	34	40	92·6	86·7	89·6	90·9
	200	8	161	3	4	2	3	90·3	86·6	87·8	92·3
	400	2	243	88	109	29	34	93·3	89·6	91·0	92·4
	400	8	321	3	4	3	3	90·8	88·4	89·1	91·5

Although I_4 is easy to implement, the theoretical results for I_4 remain open even in the no-covariate case of Chao (1987). Finally, the two abundance estimators $\tilde{N}$ and $\hat{N}$ are indeed quite close, although $\tilde{N}$ is slightly larger than $\hat{N}$ in general.

We next study the special case where all the β_j are equal. The population size is still $N_0 = 200$ or 400, and the number of capture occasions is $k = 2$ or 8. We chose $k = 8$ because it is comparable to the number of occasions, 5, 14 and 17, in the three real datasets analysed in § 4. We generated data from another two scenarios.

Scenario 3. The covariate X is the same as in Scenario 1, and the capture probability function is $g(x, \beta_s)$ with the true $q(x)$ function being $q_{03}(x) = (1, x, x^2)^{\mathrm{T}}$ and $\beta_{s0} = (-1, 2, 0{\cdot}2)^{\mathrm{T}}$.

Scenario 4. The covariate $X = (X_1, X_2)^{\mathrm{T}}$ is the same as in Scenario 2. The capture probability function is $g(x, \beta_s)$ with the true q function being $q_{04}(x) = (1, x_1, x_2)^{\mathrm{T}}$ and $\beta_{s0} = (0{\cdot}1, -2{\cdot}5, -0{\cdot}15)^{\mathrm{T}}$.

Under Scenario 3, the probabilities of overall capture are $1 - \alpha_0 = 0{\cdot}493$ and $0{\cdot}762$ when $k = 2$ and 8. Under Scenario 4, the probabilities of overall capture are 0·616 and 0·803 when $k = 2$ and 8. When implementing our method and the conditional likelihood method, we set $q(x)$ to $q_{03}(x)$ and $q_{04}(x)$ in Scenarios 3 and 4, respectively. The simulation results are summarized in Table 2.

Again $\bar{n}$ is close to $N_0(1 - \alpha_0)$ in every case. The maximum empirical likelihood estimator is still uniformly more accurate than the maximum conditional likelihood estimator in terms of MSE$_1$ and MSE$_2$. As k increases from 2 to 8, both point estimators become noticeably more accurate. In terms of coverage precision, the empirical likelihood ratio-based confidence interval $\mathcal{I}_{1s}$ is much better than $\mathcal{I}_{4s}$ in Scenarios 3 and 4 with $k = 2$ and $N_0 = 200$ and in Scenario 3 with $k = 2$ and $N_0 = 400$, although they are comparable in the other settings. The gain in coverage probability of $\mathcal{I}_{1s}$ compared to $\mathcal{I}_{4s}$ can be as large as 6% in Scenario 3 with $N_0 = 200$ and $k = 2$. Both $\mathcal{I}_{1s}$ and $\mathcal{I}_{4s}$ are uniformly more accurate than $\mathcal{I}_{2s}$ and $\mathcal{I}_{3s}$. In general, the transformation $\log(\tilde{N} - n)$ indeed improves the coverage of the Wald-type confidence intervals. The empirical likelihood ratio-based confidence intervals $\mathcal{I}_{1s}$ have reduced coverage probabilities as k increases. A possible interpretation is that for fixed N_0, the approximation of the limiting χ_1^2 distribution to the finite-sample distribution of the empirical likelihood ratio worsens as k increases. Nevertheless, the empirical likelihood ratio-based confidence intervals still perform better than $\mathcal{I}_{2s}$ and $\mathcal{I}_{3s}$ and comparably to $\mathcal{I}_{4s}$ as k increases.

In the Supplementary Material, we report the results of more simulations for small N_0 and large N_0. Further discussion of our observations can also be found in the Supplementary Material.

Moreover, we propose a bootstrap procedure to improve the performance of the empirical likelihood ratio-based confidence interval; the details are given in the Supplementary Material.

4. Real-data analysis

We illustrate the application of the proposed empirical likelihood method by analysing three real datasets: possum data (Heinze et al., 2004; Huggins & Hwang, 2007), mouse data (Stoklosa et al., 2011), and bird data (Hwang & Huang, 2003; Huggins & Hwang, 2010). The possum data, concerning captures of the mountain pygmy possum, were collected at Mount Hotham in the snowfields of Victoria, Australia, over five consecutive nights in November 2003. The body mass, in grams, of each captured animal was measured. For this dataset, $n = 43$ possums were captured at least once over $k = 5$ occasions. The mouse dataset records captures of the harvest mouse conducted at Wulin Recreation Area in Shei-Pa National Park, Taiwan, in the summer of 2008, over $k = 14$ occasions. Each captured individual was weighed and then released. In total, $n = 142$ mice were captured at least once. The bird data contain the captures and wing lengths of the bird species *Prinia flaviventris*; the data were collected at the Mai Po Bird Sanctuary of Hong Kong in 1993 over 17 weekly capture occasions. For this dataset, $n = 164$ birds were captured at least once over $k = 17$ occasions. All three datasets are available in the supplementary material of Stoklosa et al. (2011).

In the data analysis, we use X to denote the body mass for the possum and mouse data and the wing length for the bird data. We use the $M_{\rm h}$ model for all three datasets, as suggested by Stoklosa et al. (2011); that is, for each dataset, we assume that all the β_j are equal to a common value β_s. We choose $q(x) = (1, x, x^2)^{\rm T}$ as used by Stoklosa et al. (2011). Table 3 gives the point estimates $\hat{N}_s$ and $\tilde{N}_s$ and the 95% confidence intervals $\mathcal{I}_{1s}, \ldots, \mathcal{I}_{4s}$. For all three datasets, $\hat{N}_s$ and $\tilde{N}_s$ are quite close to each other, and this is in accordance with the results of our simulation studies. The confidence intervals are, however, quite different. For all three datasets, the empirical likelihood ratio-based interval $\mathcal{I}_{1s}$ has reliable performance and produces reasonable results. In contrast, the two Wald-type intervals $\mathcal{I}_{2s}$ and $\mathcal{I}_{3s}$ are unstable and may produce unsatisfactory results. For the mouse data, $\mathcal{I}_{2s}$ and $\mathcal{I}_{3s}$ are comparable to $\mathcal{I}_{1s}$. However, for the possum data the lower limits, 33 and 38, of $\mathcal{I}_{2s}$ and $\mathcal{I}_{3s}$ are below the number of observations, $n = 43$. This is also the case for the bird data, where the lower limit of $\mathcal{I}_{2s}$ is 92 and $n = 164$. The confidence interval $\mathcal{I}_{4s}$, which is also preferable to $\mathcal{I}_{2s}$ and $\mathcal{I}_{3s}$, seems close to $\mathcal{I}_{1s}$.

Table 3 also gives the maximum empirical likelihood estimates $(\hat{\beta}_s, \hat{\alpha}_s)$, the maximum conditional likelihood estimate $\tilde{\beta}_s$, and $\hat{\lambda}_s$, which is the solution to (14) with $(\hat{\beta}_s, \hat{\alpha}_s)$ in place of (β_s, α). We observe that $\hat{\lambda}_s \approx -1/(1 - \hat{\alpha}_s)$ for all three datasets, which is quite reasonable since we showed in our theoretical analysis that $\hat{\alpha}_s = \alpha_0 + o_{\rm p}(1)$ and $\hat{\lambda}_s = -1/(1 - \alpha_0) + o_{\rm p}(1)$ for some $\alpha_0 \in (0, 1)$. The estimates $\hat{\beta}_s$ and $\tilde{\beta}_s$ are also close to each other for all three datasets, as are the corresponding estimated capture probability functions. Figure 1 shows the estimated capture probability functions based on $\hat{\beta}_s$. It also displays histograms of the covariates and the usual kernel density estimates, which are defined as

$$\hat{f}_u(x) = \sum_{i=1}^{n} (nh)^{-1} K\{(x_i - x)h^{-1}\},$$

where h is a bandwidth and $K(x)$ is a kernel function, usually chosen to be the standard normal density function. We choose the bandwidth h by rule of thumb: $h = 1{\cdot}06\, \hat{\sigma}_x n^{-1/5}$ where $\hat{\sigma}_x^2$ is the sample variance of the covariates x_i.

Table 3. *Analysis results for the three real datasets*

Dataset	Point estimate	95% confidence interval	Estimates of β_s and σ_s^2
Possum	$\hat{N}_s = 55$	$\mathcal{I}_{1s} = [45, 127]$	$\hat{\beta}_s = (-41{\cdot}51, 2{\cdot}14, -0{\cdot}03)$
$n = 43$	$\tilde{N}_s = 59$	$\mathcal{I}_{2s} = [33, 84]$	$\tilde{\beta}_s = (-45{\cdot}36, 2{\cdot}34, -0{\cdot}03)$
		$\mathcal{I}_{3s} = [38, 91]$	$\hat{\sigma}_s^2 = 2{\cdot}95$
		$\mathcal{I}_{4s} = [47, 109]$	$\hat{\alpha}_s = 0{\cdot}23,\ \hat{\lambda}_s = -1{\cdot}24$
Mouse	$\hat{N}_s = 175$	$\mathcal{I}_{1s} = [159, 200]$	$\hat{\beta}_s = (-4{\cdot}19, 0{\cdot}29, -0{\cdot}001)$
$n = 142$	$\tilde{N}_s = 176$	$\mathcal{I}_{2s} = [158, 195]$	$\tilde{\beta}_s = (-4{\cdot}25, 0{\cdot}30, -0{\cdot}002)$
		$\mathcal{I}_{3s} = [159, 197]$	$\hat{\sigma}_s^2 = 0{\cdot}53$
		$\mathcal{I}_{4s} = [162, 201]$	$\hat{\alpha}_s = 0{\cdot}19,\ \hat{\lambda}_s = -1{\cdot}22$
Bird	$\hat{N}_s = 657$	$\mathcal{I}_{1s} = [394, 2360]$	$\hat{\beta}_s = (-357{\cdot}81, 15{\cdot}12, -0{\cdot}16)$
$n = 164$	$\tilde{N}_s = 675$	$\mathcal{I}_{2s} = [92, 1257]$	$\tilde{\beta}_s = (-368{\cdot}44, 15{\cdot}57, -0{\cdot}17)$
		$\mathcal{I}_{3s} = [284, 1600]$	$\hat{\sigma}_s^2 = 131{\cdot}00$
		$\mathcal{I}_{4s} = [341, 1636]$	$\hat{\alpha}_s = 0{\cdot}75,\ \hat{\lambda}_s = -4{\cdot}01$

n, sample size; $(\hat{N}_s, \hat{\beta}_s, \hat{\alpha}_s)$, the maximum empirical likelihood estimate of (N, β_s, α); $\hat{\lambda}_s$, the solution to (14) with $(\hat{\beta}_s, \hat{\alpha}_s)$ in place of (β_s, α); $(\tilde{N}_s, \tilde{\beta}_s)$, the maximum conditional likelihood estimate of (N, β_s); $\mathcal{I}_{1s}$, the empirical likelihood ratio-based confidence interval for N; $\mathcal{I}_{2s}, \mathcal{I}_{3s}$ and $\mathcal{I}_{4s}$, Wald-type confidence intervals for N.

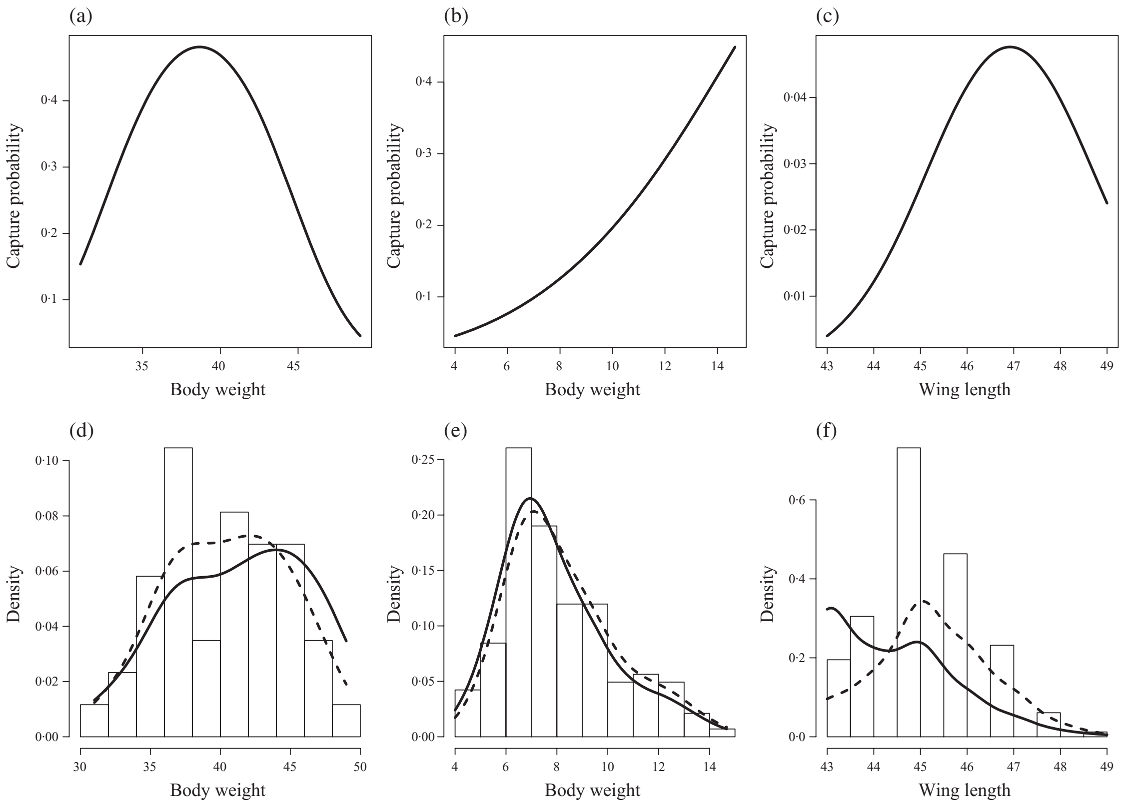

Fig. 1. Capture probability functions and kernel density estimates for the covariates of the three real datasets: panels (a)–(c) show the estimated capture probability functions of the possum, mouse and bird data, respectively; panels (d)–(f) plot the histogram, the usual kernel density estimates $\hat{f}_u(x)$ (dotted), and the weighted estimates $\hat{f}_w(x)$ (solid) of the possum body weights, mouse body weights, and bird wing lengths.

Since the observed covariates from $F(x)$ are subject to selection bias, the naive kernel density estimator $\hat{f}_u(x)$ is a biased estimator of $f(x)$. Hence, neither the histogram nor $\hat{f}_u(x)$ reflects the underlying true distribution of X. The selection bias can be corrected by the proposed empirical likelihood method. Given the maximum empirical likelihood estimators $\hat{\beta}_s$ and $\hat{\alpha}$, we obtain the maximum empirical likelihood estimators of the covariate distribution $F(x)$ as $\hat{F}(x) = \sum_{i=1}^{n} \hat{p}_{si} I(x_i \leqslant x)$, where the maximum empirical likelihood estimators of the probability weights are

$$\hat{p}_{si} = \frac{1}{n} \frac{1}{1 + \hat{\lambda}_s\{\phi_s(x_i, \hat{\beta}_s) - \hat{\alpha}_s\}},$$

with $\hat{\lambda}_s$ being the solution to

$$\sum_{i=1}^{n} \frac{\phi_s(x_i, \hat{\beta}_s) - \hat{\alpha}_s}{1 + \lambda\{\phi_s(x_i, \hat{\beta}_s) - \hat{\alpha}_s\}} = 0.$$

Using these probability weights, we construct a weighted kernel estimator of the covariate density function,

$$\hat{f}_w(x) = \sum_{i=1}^{n} \hat{p}_{si} K\{(x_i - x)h^{-1}\}h^{-1},$$

where the bandwidth $h = 1{\cdot}06\,\hat{\sigma}_x n^{-1/5}$ is the same as in $\hat{f}_u(x)$.

PROPOSITION 1. *Assume that the conditions of Corollary 1 hold and that $K(x)$ is a bounded, symmetric and continuous density function. Further, assume $f(x) > 0$ for the given x. As N_0 goes to infinity, if $h = o(1)$ and $N_0 h^2 \to \infty$, then*

$$\hat{f}_w(x) = f(x) + o_{\mathrm{p}}(1), \quad \hat{f}_u(x) = (1 - \alpha_0)^{-1}\{1 - \phi_s(x, \beta_0)\} f(x) + o_{\mathrm{p}}(1).$$

Proposition 1 indicates that as estimators of $f(x)$, the weighted kernel density estimator $\hat{f}_w(x)$ is consistent while the usual kernel density estimator $\hat{f}_u(x)$ is inconsistent unless $g(x, \beta_s)$ is independent of the covariate x. The weighted kernel density estimates are also plotted in Fig. 1. The bias correction can be observed in the figure. Compared with the usual kernel density estimate, the weighted estimate places more probability at x where the capture probability is small and less probability at x where the capture probability is large. This agrees with our intuition: observations with higher capture probabilities are more easily observed than those with lower capture probabilities. Our empirical likelihood method succeeds in correcting this bias.

When comparing the estimated covariate density function with the empirical one in the second row of Fig. 1, we observe that they are close to each other for the possum and mouse datasets but not for the bird dataset. A possible reason is that a majority of the animals were caught in the first two datasets, i.e., $n = 43$ out of $\hat{N} = 55$ for the possums and $n = 142$ out of $\hat{N} = 175$ for the mice. In contrast, the bird data have $n = 164$ versus $\hat{N} = 657$, i.e., only a small proportion of the birds were captured.

Acknowledgement

The authors thank the editor, associate editor, and two referees for constructive comments and suggestions that have led to significant improvements in the paper. Liu was supported by the National Natural Science Foundation of China, the Chinese 111 Project, and the Program of Shanghai Subject Chief Scientist. Li was supported in part by the Natural Sciences and Engineering Research Council of Canada.

Supplementary material

Supplementary Material available at *Biometrika* online contains proofs of Theorems 1 and 2, Corollaries 1 and 2, the semiparametric efficiency of $\hat{N}$, Proposition 1 and the consistency of $\hat{\sigma}^2$ in (12) and $\hat{\sigma}_s^2$ in (15) as well as numerical procedures and more simulation results.

References

Alho, J. M. (1990). Logistic regression in capture-recapture models. *Biometrics* **46**, 623–35.

Barnard, J., Emam, K. & Zubrow, D. (2003). Using capture-recapture models for the reinspection decision. *Software Qual. Professional* **5**, issue 2, March.

Boden, L. I. & Ozonoff, A. (2008). Capture-recapture estimates of nonfatal workplace injuries and illnesses. *Ann. Epidemiol.* **18**, 500–6.

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

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

Borchers, D. L., Zucchini, W. & Fewster, R. M. (1998). Mark-recapture models for line transect surveys. *Biometrics* **54**, 1207–20.

Chao, A. (1987). Estimating the population size for capture-recapture data with unequal catchability. *Biometrics* **43**, 783–91.

DiCiccio, T. J., Hall, P. & Romano, J. P. (1991). Empirical likelihood is Bartlett-correctable. *Ann. Statist.* **19**, 1053–61.

Evans, M. A. & Bonett, D. G. (1994). Bias reduction for multiple recapture estimators of closed population size. *Biometrics* **50**, 388–95.

Evans, M. A., Bonett, D. G. & McDonald, L. L. (1994). A general theory for modeling capture-recapture data from a closed population. *Biometrics* **50**, 396–405.

Fewster, R. M. & Jupp, P. E. (2009). Inference on population size in binomial detectability models. *Biometrika* **96**, 805–20.

Fewster, R. M. & Jupp, P. E. (2013). Information on parameters of interest decreases under transformations. *J. Mult. Anal.* **96**, 34–9.

Hall, P. & La Scala, B. (1990). Methodology and algorithms of empirical likelihood. *Int. Statist. Rev.* **58**, 109–27.

Heinze, D., Broome, L. & Mansergh, I. (2004). A review of the ecology and conservation of the mountain pygmy-possum *Burramys parvus*. In *The Biology of Australian Possums and Gliders*, R. L. Goldingay and S. M. Jackson (eds). Chipping Norton, Australia: Surrey Beatty & Sons, pp. 254–67.

Hogan, H. (2000). *Accuracy and Coverage Evaluation 2000: Decomposition of Dual System Estimate Components*. DSSD Census 2000 Procedures and Operation Memorandum Series B-8. Washington, DC: U.S. Census Bureau.

Huggins, R. M. (1989). On the statistical analysis of capture experiments. *Biometrika* **76**, 133–40.

Huggins, R. M. & Hwang, W. H. (2007). Non-parametric estimation of population size from capture-recapture data when the capture probability depends on a covariate. *Appl. Statist.* **56**, 429–43.

Huggins, R. M. & Hwang, W. H. (2010). A measurement error model for heterogeneous capture probabilities in mark-recapture experiments: An estimating equation approach. *J. Agric. Biol. Envir. Statist.* **15**, 198–208.

Hwang, W. H. & Huang, S. Y. H. (2003). Estimation in capture-recapture models when covariates are subject to measurement errors. *Biometrics* **59**, 1113–22.

Lukacs, P. M. & Burnham, K. P. (2005). Review of capture-recapture methods applicable to noninvasive genetic sampling. *Molec. Ecol.* **14**, 3909–19.

Marques, F. F. C. & Buckland, S. T. (2004). Covariate models for the detection function. In *Advanced Distance Sampling: Estimating Abundance of Biological Populations*, S. T. Buckland, D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers and L. Thomas (eds). Oxford: Oxford University Press, pp. 31–47.

NEWEY, W. K. & SMITH, R. J. (2004). High order properties of GMM and generalized empirical likelihood estimators. *Econometrica* **72**, 219–55.

OTIS, D. L., BURNHAM, K. P., WHITE, G. C. & ANDERSON, D. R. (1978). Statistical inference for capture data on closed animal populations. *Wildl. Monog.* **62**, 1–135.

OWEN, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75**, 237–49.

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

OWEN, A. B. (2001). *Empirical Likelihood*. New York: Chapman & Hall/CRC.

SEBER, G. A. F. (1982). *Estimation of Animal Abundance and Related Parameters*. New York: Hafner, 2nd ed.

STOKLOSA, J., HWANG, W. H., WU, S. H. & HUGGINS, R. M. (2011). Heterogeneous capture-recapture models with covariates: A partial likelihood approach for closed populations. *Biometrics* **67**, 1659–65.

TILLING, K., STERNE, J. A. & WOLFE, C. D. (2001). Multilevel growth curve models with covariate effects: Application to recovery after stroke. *Statist. Med.* **20**, 685–704.

[*Received on* 27 *January* 2016. *Editorial decision on* 4 *June* 2017]