Brazilian Journal of Probability and Statistics
2016, Vol. 30, No. 1, 47–69
DOI: 10.1214/14-BJPS255

Likelihood-based inference for population size in a capture–recapture experiment with varying probabilities from occasion to occasion

Luis Ernesto Bueno Salasar[a], **José Galvão Leite**[a] and **Francisco Louzada**[b]

[a]*Universidade Federal de São Carlos*
[b]*Universidade de São Paulo*

Abstract. The estimation of the size of a population is, in general, performed using capture–recapture experiments. In this paper, we consider a closed population capture–recapture model in which individuals are captured independently and with the same probability in each sampling occasion, but the probabilities may vary from occasion to occasion. The unknown number of individuals is the parameter of interest, while the capture probabilities are the nuisance ones. Four likelihood functions free of nuisance parameters, namely the profile, conditional, uniform and Jeffrey's integrated likelihood functions are derived and procedures for point and interval estimation are discussed. The estimation of population size is illustrated on a real dataset. The frequentist properties of the estimators are evaluated by means of a simulation study. The Jeffrey's integrated likelihood achieved the best performance over all considered estimators for both point and interval estimation, particularly in situations with little information with small number of elements, small capture probabilities and small number of capture occasions.

1 Introduction

The so-called capture–recapture sampling process is frequently used in the estimation of the number of elements of a population. One of the first applications of such a method was made by Laplace (1783) in order to estimate the number of inhabitants of France. Later, Petersen (1896) applied the method to study fish population in Baltic sea and, independently, Lincoln (1930) applied the method to estimate the size of North American waterfowl population. The method applied by the later researchers was based on only two sampling occasions and the applied estimator became known as the "Lincoln–Petersen" estimator, which is obtained by equating the proportion of marked individuals in the second sample to the proportion of marked individuals in the population previous to the selection of the second sample. Since the 1950s several important scientific papers have been published on the subject. For instance, we may cite Chapman (1954), Darroch

Key words and phrases. Capture–recapture, conditional likelihood, Jeffrey's integrated likelihood, population size, profile likelihood, simulation, uniform integrated likelihood.
Received June 2013; accepted July 2014.

(1958, 1959), Seber (1965), Jolly (1965), Cormack (1968) and presently capture–recapture methods finds applications in various fields such as ecology (Otis et al., 1978; Seber, 1982), software reliability (Nayak, 1988; Basu and Ebrahimi, 2001), epidemiology (Seber, Huakau and Simmons, 2000; Lee et al., 2001; Chao et al., 2001; Lee, 2002), linguistics (Boender and Rinooy Kan, 1987; Thisted and Efron, 1987) among others. For a review of different models and applications see Seber (1982, 1986, 1992), Schwarz and Seber (1999), Pollock (2000), Chao (2001) and Amstrup et al. (2003).

Let N be the unknown number of elements of a closed population. Estimation of N is generally carried by maximum likelihood (Cormack, 1968, 1989; Darroch, 1958; Otis et al., 1978; Norris and Pollock, 1996; Pledger, 2000). However, maximum likelihood estimates is known to be biased and only an asymptotic estimate of bias and variance are available (Seber, 1982). Further, maximum likelihood estimates have been criticized because it may lead to some pathologies such as infinite estimates for the population size. These "likelihood failures" have been pointed out for different capture–recapture models (see, for instance, Seber and Whale (1970), Carle and Strub (1978), Leite, Oishi and Pereira (1988) and Fegatelli and Tardella (2013)).

Interval estimation for the population size are usually obtained by considering the Wald-type interval (Seber, 1982; Otis et al., 1978) and relies on asymptotic normality of the maximum likelihood estimator $\widehat{N}$, which may be highly skewed and the obtained confidence intervals can be misleading (Garthwaite and Buckland, 1990). Furthermore, this procedure has been criticized because the lower limit interval is not necessarily greater than the number of distinct elements captured and can even assume negative values. Sprott (1981) pointed out that variances are not very useful for finite samples when the likelihood function of N is skewed and suggests the consideration of an appropriate transformation which can achieve an approximate normal likelihood. Several transformations were considered in the literature to overcome this problem (Viveros and Sprott, 1986; McDonald and Palanacki, 1989; Chao, 1989). Another approach, discussed by Evans, Kim and O'Brien (1996) and Cormack (1992) is the consideration of the profile likelihood interval in the context of log-linear models for incomplete contingency tables.

Otis et al. (1978) described eight models for closed population which allows for three sources of variation on capture probabilities: time (t), behavior (b) and heterogeneity (h). Here, we shall focus on the model M_t, which allows for capture probabilities to vary from occasion to occasion. The model M_t has received much attention from both frequentist (Sanathanan, 1972b, 1973; Pickands and Raghavachari, 1974; Yip, 1991; Leite, Oishi and Pereira, 1987, 1988) and Bayesian perspectives (Castledine, 1981; Smith, 1988, 1991; George and Robert, 1992; Bolfarine, Leite and Rodrigues, 1992). From the Bayesian side, difficult relies on how to choose the prior distribution, since the resulting estimator for N may be very sensitive to this choice (Chao, 1989). Usually, researchers use informative priors when valuable information is available, otherwise noninformative priors may be considered. Wang, He and Sun (2007) give conditions for the existence of the posterior distributions and its mean for different noninformative prior distributions for model M_t and compare the frequentist performance of both point and interval Bayesian estimates under a simulation study. Our approach is closely related to the Bayesian procedure considered by Wang, He and Sun (2007) if one chooses the uniform prior for N and the loss function 0–1. They have considered the squared-error loss function and their estimation process need the calculation of the posterior marginal mean for N, which is carried by means of MCMC simulations. On the other hand, our method relies only on obtaining the maximum of the integrated likelihood function, which is much less computationally intensive.

Generally, the main interest resides in the estimation of the population size, so the capture probabilities are regarded as nuisance parameters. In this context, for the first time we study how methods for eliminating nuisance parameters (Cox, 1975; Berger, Liseo and Wolpert, 1999; Basu, 1977) can be applied for the so-called model M_t.

The paper is organized as follows. In Section 2, we present the probabilistic capture–recapture model. In Section 3, we present the profile, conditional, uniform integrated and Jeffreys integrated likelihood functions, as well as the method for obtaining the respective maximum likelihood estimates. In Section 4, the likelihood interval method is presented. In Section 5, the frequentist properties of the estimators are inspected by means of a simulation study. In Section 6, the methods are illustrated on a real data-set. In Section 7, some final comments are presented.

2 Capture–recapture model

In the following, we will consider the estimation of the number of elements of a *closed* population, that is, no recruitment (birth or immigration) or losses (deaths or emigration) can occur in the period of the experiment. Consider that the population is composed by N elements numbered 1 to N, which are sampled in k different occasions, $k \geq 2$. In the first sampling occasion, a number of elements are captured, marked, and then returned to the population. After allowing time for marked and unmarked elements to mix, a second sample is taken, the captured elements are marked with a mark distinct from the first occasion and then returned to the population. The process is repeated until k samples are obtained and provides the knowledge of the full capture–recapture history (trajectory) of each captured element.

To the ith element of the population, we associate the vector $\mathbf{X}_i = (X_{i,1}, \ldots, X_{i,k})$ with $X_{i,j} = 1$ if the ith element is captured in the jth occasion and $X_{i,j} = 0$ otherwise, $i = 1, \ldots, N$. Further, suppose the ith element is captured in the jth occasion, independently from the other elements and occasions, with probability p_j. Therefore, the $\mathbf{X}_i$'s are independent and identically distributed random vectors assuming the trajectory $\boldsymbol{\omega} = (\omega_1, \dots, \omega_k)$ with probability $\prod_{j=1}^{k} p_j^{\omega_j}(1-p_j)^{1-\omega_j}$, $\boldsymbol{\omega} \in \{0, 1\}^k$. In this way, the process may be regarded as N independent and identically distributed realizations with outcomes in the set $\{0, 1\}^k$ of all trajectories. The above assumptions corresponds to the model M_t discussed by Otis et al. (1978) and also to the Geiger and Werner model discussed by Sanathanan (1972b, 1973).

Indexing each trajectory $\boldsymbol{\omega} = (\omega_1, \dots, \omega_k)$ by $l = \sum_{j=1}^{k} \omega_j 2^{j-1}$, we obtain the enumeration $\{\boldsymbol{\omega}^l = (\omega_1^l, \dots, \omega_k^l), l = 0, 1, \dots, L\}$ of the set $\{0, 1\}^k$, $L = 2^k - 1$. For each l, let Y_l be the number of elements of the population taking the trajectory $\boldsymbol{\omega}^l$, $l = 0, 1, \dots, L$. Then, the random vector $(Y_0, Y_1, \dots, Y_L)$ has a Multinomial distribution $M(N, (\theta_0, \dots, \theta_L))$, where $\theta_l = \prod_{j=1}^{k} p_j^{\omega_j^l}(1-p_j)^{1-\omega_j^l}$ is the probability of observing the trajectory $\boldsymbol{\omega}^l = (\omega_1^l, \dots, \omega_k^l)$, $l = 0, 1, \dots, 2^k - 1$. Since the trajectory $\boldsymbol{\omega}^0 = (0, \dots, 0)$ is not observed, the available data consists of the vector of counts $(Y_1, \dots, Y_L)$ and the likelihood of N and $\mathbf{p}$ is given by

$$
\begin{aligned}
L(N, \mathbf{p}) &= P\{Y_1 = y_1, \dots, Y_L = y_L | N, \mathbf{p}\} \\
&= \frac{N!}{(N-r)! y_1! \cdots y_L!} \theta_0^{N-r} \theta_1^{y_1} \cdots \theta_L^{y_L} \\
&= \frac{N!}{(N-r)! y_1! \cdots y_L!} \prod_{j=1}^{k} (1-p_j)^{N-r} \\
&\quad \times \prod_{l=1}^{L} \prod_{j=1}^{k} (p_j)^{y_l \omega_j^l} (1-p_j)^{y_l(1-\omega_j^l)} \\
&= \frac{N!}{(N-r)! y_1! \cdots y_L!} \prod_{j=1}^{k} (1-p_j)^{N-r} \prod_{j=1}^{k} p_j^{n_j} (1-p_j)^{r-n_j} \\
&= \frac{\Gamma(N+1)}{\Gamma(N-r+1) y_1! \cdots y_L!} \prod_{j=1}^{k} p_j^{n_j} (1-p_j)^{N-n_j},
\end{aligned}
\tag{1}
$$

$N \geq r, 0 < p_j < 1$, $j = 1, \dots, k$, where $\Gamma(\cdot)$ is the gamma function and $n_j = \sum_{l=1}^{L} y_j \omega_j^l$ is the number of captured elements in the jth occasion, $r = \sum_{l=1}^{L} y_l$ is the number of distinct elements captured at least once, and $y_1, \dots, y_L$ are nonnegative integers such that $y_1 + \cdots + y_L \leq N$.

From (2), it follows that the kernel of the likelihood function is given by

$$
K(N, \mathbf{p}) = \frac{\Gamma(N+1)}{\Gamma(N-r+1)} \prod_{j=1}^{k} p_j^{n_j} (1-p_j)^{N-n_j}, \tag{2}
$$

$N \geq r, 0 < p_j < 1$, $j = 1, \dots, k$. From the kernel (2), we conclude that $\mathbf{n} = (n_1, \dots, n_k)$ and r are jointly sufficient to estimate N and $\mathbf{p}$. The kernel of the log-likelihood is given by

$$\log\bigl(K(N,\mathbf{p})\bigr) = \log\bigl(\Gamma(N+1)\bigr) - \log\bigl(\Gamma(N-r+1)\bigr) \tag{3}$$

$$+ \sum_{j=1}^{k} n_j \log(p_j) + \sum_{j=1}^{k}(N-n_j)\log(1-p_j), \tag{4}$$

$N \geq r, 0 < p_j < 1,\ j = 1, \ldots, k.$

3 Likelihood functions

In this section, we present the expressions for the conditional, profile, Jeffreys and uniform integrated likelihoods. The profile likelihood is obtained by replacing the nuisance parameter $\mathbf{p}$ with its conditional maximum likelihood estimate. The conditional likelihood is obtained by conditioning the data on a appropriate statistic leading to a conditional distribution not depending on the nuisance parameter.

The integrated likelihoods are discussed in Berger, Liseo and Wolpert (1999) and consists in eliminate the nuisance parameter by integration with respect to a given density function. Thus, considering π a density function for $\mathbf{p}$, the integrated likelihood is

$$L^{\pi}(N) = \int_{(0,1)^k} L(N,\mathbf{p})\pi(\mathbf{p})\,d\mathbf{p}.$$

3.1 Conditional likelihood

Since captures occur independently from elements and occasions, the number n_j of captures at jth occasion, $j = 1, \ldots, k$, are independent binomial distributed variables with probability function

$$P\{n_1, n_2, \ldots, n_k | N, \mathbf{p}\} = \prod_{j=1}^{k} \binom{N}{n_j} p_j^{n_j}(1-p_j)^{N-n_j}, \tag{5}$$

and from (2) and (5), the conditional distribution of $(Y_1, \ldots, Y_L)$ given $(n_1, \ldots, n_k)$ is

$$P\{Y_1 = y_1, \ldots, Y_L = y_L | n_1, \ldots, n_k\} = \frac{N!}{(N-r)!y_1!\cdots y_L!\prod_{j=1}^{k}\binom{N}{n_j}}. \tag{6}$$

Since the conditional distribution (6) does not depend on the nuisance parameters $\mathbf{p}$, we take (6) as the conditional likelihood and its kernel is given by

$$K^{\mathcal{C}}(N) = \frac{\Gamma(N+1)}{\Gamma(N-r+1)} \prod_{j=1}^{k} \frac{\Gamma(N-n_j+1)}{\Gamma(N+1)}, \qquad N \geq r. \tag{7}$$

Thus, the kernel of the log-likelihood is

$$\log\bigl(K^{\mathcal{C}}(N)\bigr) = (1-k)\log\bigl(\Gamma(N+1)\bigr) - \log\bigl(\Gamma(N-r+1)\bigr) + \sum_{j=1}^{k}\log\bigl(\Gamma(N-n_j+1)\bigr)$$

for $N \geq r$.

Again, we treat N as a real variable and obtain the maximum $\widetilde{N}^{\mathcal{C}}$ of $K^{\mathcal{C}}$ solving

$$\frac{d}{dN}\log\bigl(K^{\mathcal{C}}(\widetilde{N}^{\mathcal{C}})\bigr) = (1-k)\psi(\widetilde{N}^{\mathcal{C}}+1) - \psi(\widetilde{N}^{\mathcal{C}}-r+1) + \sum_{j=1}^{k}\psi(\widetilde{N}^{\mathcal{C}}-n_j+1) = 0, \qquad \widetilde{N}^{\mathcal{C}} > r,$$

and the integer maximum conditional likelihood $\widehat{N}^{\mathcal{C}}$ will be either $[\widetilde{N}^{\mathcal{C}}]$ or $[\widetilde{N}^{\mathcal{C}}]+1$ according to which attains the greatest value of the conditional likelihood.

The conditional likelihood (7) can be also derived from a capture–recapture model with fixed sample sizes $n_1, \ldots, n_k$ considered in Leite, Oishi and Pereira (1988). The latter authors proved that a necessary and sufficient condition for a finite conditional maximum likelihood estimate is that $r < n_1 + n_2 + \cdots + n_k$. Another approach for deriving a conditional likelihood function is proposed by Sanathanan (1972a), which is based on a factorization of the likelihood function in two parts: a likelihood $L_1(N, \mathbf{p})$ derived from the conditional distribution of data given the number r of distinct elements captured and a likelihood $L_2(\mathbf{p})$ derived from the marginal distribution of r. The estimate of N is obtained via a two step procedure: (i) obtain $\widehat{\mathbf{p}}$ which maximizes $L_2(\mathbf{p})$; (ii) For $p = \widehat{\mathbf{p}}$, obtain $\widehat{N}^{\mathcal{C}}$ which maximizes $L_1(N, \mathbf{p})$. In this sense, our approach is simpler because the estimation process is performed in only one step.

3.2 Profile likelihood

The profile likelihood is defined by $L^{\mathcal{P}}(N) = L(N, \widehat{\mathbf{p}}(N))$, where $\widehat{\mathbf{p}}(N) = (\widehat{p}_1(N), \ldots, \widehat{p}_k(N))$ is the value of $\mathbf{p}$ which maximizes the kernel (2) for each fixed $N \geq r$. It is important to note that because of the discreteness of the parameter N, finding the maximum of the profile likelihood is the same of finding the maximum of the original likelihood (2).

From the kernel of the log-likelihood (3), we solve the equations

$$\frac{\partial \log(K(N, \widehat{\mathbf{p}}(N)))}{\partial p_j} = \frac{n_j}{\widehat{p}_j} - \frac{N-n_j}{1-\widehat{p}_j} = 0, \qquad j = 1, \ldots, k,$$

to obtain

$$\widehat{\mathbf{p}}(N) = \left(\frac{n_1}{N}, \ldots, \frac{n_k}{N}\right), \qquad N \geq r. \tag{8}$$

Replacing (8) into (3), the kernel of the profile likelihood is given by

$$K^{\mathcal{P}}(N) = \frac{\Gamma(N+1)}{\Gamma(N-r+1)} \prod_{j=1}^{k} \frac{(N-n_j)^{N-n_j}}{N^N}, \tag{9}$$

and

$$\begin{aligned} \log\bigl(K^{\mathcal{P}}(N)\bigr) = {} & \log\bigl(\Gamma(N+1)\bigr) - \log\bigl(\Gamma(N-r+1)\bigr) \\ & - kN\log(N) + \sum_{j=1}^{k}(N-n_j)\log(N-n_j) \end{aligned}$$

for $N \geq r$.

Treating N as a real variable, we obtain the maximum $\widetilde{N}^{\mathcal{P}}$ of $K^{\mathcal{P}}$ as the real solution of

$$\begin{aligned} \frac{d}{dN}\log\bigl(K^{\mathcal{P}}(\widetilde{N}^{\mathcal{P}})\bigr) = {} & \psi(\widetilde{N}^{\mathcal{P}}+1) - \psi(\widetilde{N}^{\mathcal{P}}-r+1) - k\log(\widetilde{N}^{\mathcal{P}}) \\ & + \sum_{j=1}^{k}\log(\widetilde{N}^{\mathcal{P}}-n_j) = 0 \end{aligned}$$

for $\widetilde{N}^{\mathcal{P}} > r$, where $\psi(\cdot)$ denotes the digamma function. Then, the integer maximum profile likelihood $\widehat{N}^{\mathcal{P}}$ will be either $[\widetilde{N}^{\mathcal{P}}]$ or $[\widetilde{N}^{\mathcal{P}}]+1$ according to which attains the greatest value of the profile likelihood, where $[z]$ denotes the greatest integer not greater than z, for real z.

The next theorem gives a necessary and sufficient condition for a finite maximum profile likelihood estimate.

Theorem 1. *The maximum profile likelihood estimator, $\widehat{N}^{\mathcal{P}}$, is no greater than the conditional maximum likelihood estimator $\widehat{N}^{\mathcal{C}}$. The estimator $\widehat{N}^{\mathcal{P}}$ is finite if, and only if, $r < n_1 + \cdots + n_k$.*

The proof of this theorem is placed in the Appendix.

3.3 Uniform integrated likelihood

Another way of eliminating the nuisance parameter $\mathbf{p}$, discussed in Berger, Liseo and Wolpert (1999), is integrating the likelihood with respect to a probability density function $\pi(\mathbf{p}|N)$. So, the integrated likelihood with respect to the p.d.f. π is defined by

$$L^{\pi}(N) = \int_{(0,1)^k} L(N,\mathbf{p})\,d\pi(\mathbf{p}|N).$$

Assuming that, given N, $p_1, \dots, p_k$ are independent and identically uniform distributed random variables in the interval $(0, 1)$, the uniform integrated likelihood is obtained from (2) as

$$
\begin{aligned}
L^{\mathcal{U}}(N) &= \frac{N!}{(N-r)!y_1!\cdots y_L!}\prod_{j=1}^{k}\int_0^1 p_j^{n_j}(1-p_j)^{N-n_j}\,dp_j \\
&= \frac{N!}{(N-r)!y_1!\cdots y_L!}\prod_{j=1}^{k}\frac{\Gamma(n_j+1)\Gamma(N-n_j+1)}{\Gamma(N+2)}, \qquad N \geq r,
\end{aligned}
$$

and its kernel is

$$
K^{\mathcal{U}}(N) = \frac{\Gamma(N+1)}{\Gamma(N-r+1)}\prod_{j=1}^{k}\frac{\Gamma(N-n_j+1)}{\Gamma(N+2)}, \qquad N \geq r. \tag{10}
$$

The kernel of the uniform integrated log-likelihood is

$$
\begin{aligned}
\log\big(K^{\mathcal{U}}(N)\big) &= (1-k)\log\big(\Gamma(N+1)\big) - \log\big(\Gamma(N-r+1)\big) \\
&\quad - k\log(N+1) + \sum_{j=1}^{k}\log\big(\Gamma(N-n_j+1)\big)
\end{aligned} \tag{11}
$$

for $N \geq r$ and the maximum $\widetilde{N}^{\mathcal{U}}$ of $K^{\mathcal{U}}$ is obtained by solving

$$
\begin{aligned}
\frac{d}{dN}\log\big(K^{\mathcal{U}}(\widetilde{N}^{\mathcal{U}})\big) &= (1-k)\psi(\widetilde{N}^{\mathcal{U}}+1) - \psi(\widetilde{N}^{\mathcal{U}}-r+1) \\
&\quad - \frac{k}{N+1} + \sum_{j=1}^{k}\psi(\widetilde{N}^{\mathcal{U}}-n_j+1) = 0
\end{aligned}
$$

for $\widetilde{N}^{\mathcal{U}} > r$ and the integer maximum uniform integrated likelihood $\widehat{N}^{\mathcal{U}}$ will be either $[\widetilde{N}^{\mathcal{U}}]$ or $[\widetilde{N}^{\mathcal{U}}]+1$ according to which attains the greatest value of the uniform integrated likelihood.

3.4 Jeffreys integrated likelihood

Now, we will assume that, given N, the random vector $\mathbf{p}$ is distributed according to the Jeffreys distribution $\pi^{\mathcal{J}}(\mathbf{p}|N)$ which has p.d.f. defined to be proportional to the square root of the determinant of Fisher information matrix $I_N(\mathbf{p})$ given by

$$
I_N(\mathbf{p}) = \left(\mathrm{E}\left[-\frac{\partial^2 \log L(N,\mathbf{p})}{\partial p_i\,\partial p_j}\right]\right)_{k\times k} \propto \left(\mathrm{E}\left[-\frac{\partial^2 \log K(N,\mathbf{p})}{\partial p_i\,\partial p_j}\right]\right)_{k\times k}.
$$

From (3), we have

$$
\frac{\partial^2 \log K(N,\mathbf{p})}{\partial p_i\,\partial p_j} = \begin{cases} 0, & \text{if } i \neq j, \\ \dfrac{-n_j + 2n_j p_j - Np_j^2}{p_j^2(1-p_j)^2}, & \text{if } i = j, \end{cases}
$$

$i, j = 1, \ldots, n$.

Since $n_j \sim \text{Binomial}(N, p_j)$, $j = 1, \ldots, k$, it follows that

$$\mathrm{E}\Big[-\frac{\partial^2 \log L(N, \mathbf{p})}{\partial p_i \, \partial p_j}\Big] = \begin{cases} 0, & \text{if } i \neq j, \\ \dfrac{N}{p_i(1-p_i)}, & \text{if } i = j. \end{cases}$$

Therefore, the Jeffreys p.d.f. for $\mathbf{p}$ is given by

$$\begin{aligned} \pi^{\mathcal{J}}(\mathbf{p}|N) &\propto \big[\det I_N(\mathbf{p})\big]^{1/2} \\ &\propto \prod_{j=1}^{k} \frac{1}{p_j^{1/2}(1-p_j)^{1/2}}, \end{aligned}$$

and the Jeffreys integrated likelihood is

$$\begin{aligned} L^{\mathcal{J}}(N) &= \int_{(0,1)^k} L(N, \mathbf{p}) \pi^{\mathcal{J}}(\mathbf{p}|N) \, d\mathbf{p} \\ &= \frac{\Gamma(N+1)}{\Gamma(N-r+1) y_1! \cdots y_L!} \prod_{j=1}^{k} \int_0^1 p_j^{n_j - 1/2}(1-p_j)^{N-n_j-1/2} \, dp_j \\ &= \frac{\Gamma(N+1)}{\Gamma(N-r+1) y_1! \cdots y_L!} \prod_{j=1}^{k} \frac{\Gamma(n_j + 1/2)\Gamma(N - n_j + 1/2)}{\Gamma(N+1)}, \\ & N \geq r. \end{aligned}$$

The kernel of the Jeffreys integrated likelihood is

$$K^{\mathcal{J}}(N) = \frac{\Gamma(N+1)}{\Gamma(N-r+1)} \prod_{j=1}^{k} \frac{\Gamma(N - n_j + 1/2)}{\Gamma(N+1)}, \qquad N \geq r,$$

and its logarithm is given by

$$\begin{aligned} \log\big(K^{\mathcal{J}}(N)\big) = (1-k)\log\big(\Gamma(N+1)\big) - \log\big(\Gamma(N-r+1)\big) \\ + \sum_{j=1}^{k} \log\big(\Gamma(N - n_j + 1/2)\big), \qquad N \geq r. \end{aligned}$$

The maximum $\widetilde{N}^{\mathcal{J}}$ of $K^{\mathcal{J}}$ is obtained by solving

$$\begin{aligned} \frac{d}{dN} \log\big(K^{\mathcal{J}}(\widetilde{N}^{\mathcal{J}})\big) = (1-k)\psi\big(\widetilde{N}^{\mathcal{J}} + 1\big) - \psi\big(\widetilde{N}^{\mathcal{J}} - r + 1\big) \\ + \sum_{j=1}^{k} \psi\big(\widetilde{N}^{\mathcal{J}} - n_j + 1/2\big) = 0, \qquad \widetilde{N}^{\mathcal{J}} > r, \end{aligned}$$

and the integer maximum Jeffreys integrated likelihood $\widehat{N}^{\mathcal{J}}$ will be either $[\widetilde{N}^{\mathcal{J}}]$ or $[\widetilde{N}^{\mathcal{J}}] + 1$ according to which attains the greatest value of the Jeffreys integrated likelihood.

4 Confidence intervals

In the following, we will denote $\mathcal{M}$ the method applied to eliminate the nuisance parameter, that is, $\mathcal{M}$ may be $\mathcal{P}$ (profile likelihood), $\mathcal{C}$ (conditional likelihood), $\mathcal{U}$ (uniform integrated likelihood) or $\mathcal{J}$ (Jeffreys integrated likelihood). Based on the likelihood $L^{\mathcal{M}}(N)$, free of nuisance parameters, we will consider the approximate $(1-\alpha)$ confidence interval $[\widehat{N}_L^{\mathcal{M}}, \widehat{N}_U^{\mathcal{M}}]$ for N defined by the set

$$\left\{N : \log\bigl(K^{\mathcal{M}}(N)\bigr) \geq \log\bigl(K^{\mathcal{M}}(\widehat{N}^{\mathcal{M}})\bigr) - \frac{1}{2}\chi_1^2(1-\alpha)\right\},$$

where $\chi_1^2(1-\alpha)$ denotes the $(1-\alpha)$th quantile of the chi-square distribution with 1 degree of freedom, $K^{\mathcal{M}}$ the kernel of the likelihood $L^{\mathcal{M}}$ and $\widehat{N}^{\mathcal{M}}$ the maximum likelihood estimate of N with respect to the likelihood $L^{\mathcal{M}}(N)$. The use of the likelihood free of nuisance parameters for construction of confidence intervals is suggested by Berger, Liseo and Wolpert (1999). Confidence intervals based on likelihood function have an intuitive appeal since it contains the most plausible values of N and have a close relationship with the HPD Bayesian intervals.

For sake of comparison, we shall also consider the popular Wald-type confidence interval for N, mostly used in applications, with lower and upper limits given by

$$\widehat{N}_L^{\mathcal{W}} = \widehat{N} - z_{(1-\alpha/2)}\sqrt{\widehat{\mathrm{Var}}(\widehat{N})} \quad \text{and} \quad \widehat{N}_U^{\mathcal{W}} = \widehat{N} + z_{(1-\alpha/2)}\sqrt{\widehat{\mathrm{Var}}(\widehat{N})},$$

where $\widehat{\mathrm{Var}}(\widehat{N})$ is an estimate of the variance of $\widehat{N}$ and z_γ is the γth quantile of a standard normal distribution. Here, we shall consider the asymptotic estimate given by Darroch (1958),

$$\widehat{\mathrm{Var}}(\widehat{N}) = \widehat{N}\left[\prod_{j=1}^{k}(1-\hat{p}_j)^{-1} + k - 1 - \sum_{j=1}^{k}(1-\hat{p}_j)^{-1}\right]^{-1},$$

where $\widehat{\mathbf{p}} = (\hat{p}_1, \hat{p}_2, \dots, \hat{p}_k)$ is the maximum likelihood estimate of $\mathbf{p}$ given by (8) replacing N with $\widehat{N}$.

5 Simulation study

In this section, we describe a simulation study designed to assess the frequentist properties of the proposed estimators. The population size was fixed at $N = 50, 100, 200, 400, 800$, the number of capture occasions at $k = 3, 4, \dots, 10$ and capture probabilities $\mathbf{p}$ were generated from the populations P_1, P_2, P_3, P_4, where $P_i = \mathrm{Beta}(\delta_i(\mu_i), \delta_i(1-\gamma_i))$, $i = 1, \dots, 4$. Note that if $p \sim \mathrm{Beta}(\delta\mu, \delta(1-\mu))$, then p has mean μ and variance $\mu(1-\mu)/(\delta+1)$, which allows us to interpret δ as a precision parameter. Table 1 displays the values of k_i, δ_i, which were chosen

Table 1 *Populations for capture probabilities*

Population	μ	δ	Variance
P_1	0.05	3.75	0.01
P_2	0.05	0.58	0.03
P_3	0.10	3.50	0.02
P_4	0.10	0.50	0.06

in order to give different mean and variance values for the generated capture probabilities. For each combination of population size, number of occasions and vector of capture probabilities, 1000 independent observations of the vector of counts $(Y_1, Y_2, \dots, Y_{2^k})$ were simulated from model (2).

5.1 Point estimation

For the sth simulated observation and each combination of fixed values of N, $\mathbf{p}$ and k, the maximum likelihood estimate $\widehat{N}^{\mathcal{P}}(s)$ (profile), $\widehat{N}^{\mathcal{C}}(s)$ (conditional), $\widehat{N}^{\mathcal{U}}(s)$ (uniform integrated) and $\widehat{N}^{\mathcal{J}}(s)$ (Jeffreys integrated) were determined, $s = 1, 2, \dots, 1000$. To inspect the frequentists properties of point estimators, we define the descriptive measures

$$\mathrm{RMSE}(\widehat{N}^{\mathcal{M}}) = \sqrt{\frac{\sum_{s=1}^{1000}(\widehat{N}^{\mathcal{M}}(s) - N)^2}{1000}},$$

$$\mathrm{RBIAS}(\widehat{N}^{\mathcal{M}}) = \frac{1}{1000}\sum_{s=1}^{1000}\frac{\widehat{N}^{\mathcal{M}}(s) - N}{N},$$

where RMSE is the square root of the mean squared-error, RBIAS is the relative bias of estimates and $\mathcal{M}$ assumes $\mathcal{P}$, $\mathcal{C}$, $\mathcal{U}$, $\mathcal{J}$. For some generated samples, the profile and conditional likelihood are strictly increasing and the maximum likelihood estimate is ∞, which occurs mainly in the case of low capture probabilities and small number of capture occasions. In such cases, we define both RMSE and RBIAS to be ∞. Figures 1, 2 and 3 display the above descriptive measures as function of k for the values of $N = 50, 200, 800$, respectively.

By the analysis of the graphs of RMSE and RBIAS as function of k, we conclude that: (i) increasing the number of capture occasions k and the capture probabilities $\mathbf{p}$ results in decreasing of both RMSE and RBIAS (in absolute value) for all estimators; (ii) the profile and conditional estimators assume very similar RMSE and RBIAS values; (iii) in all scenarios, the Jeffreys and uniform integrated estimator assume lower RMSE values in comparison with the profile and conditional estimators; (iv) the uniform and Jeffreys integrated estimators are negatively biased while the profile and conditional are slightly positively biased (except in the cases of low capture probabilities and small number of sampling occasion when it

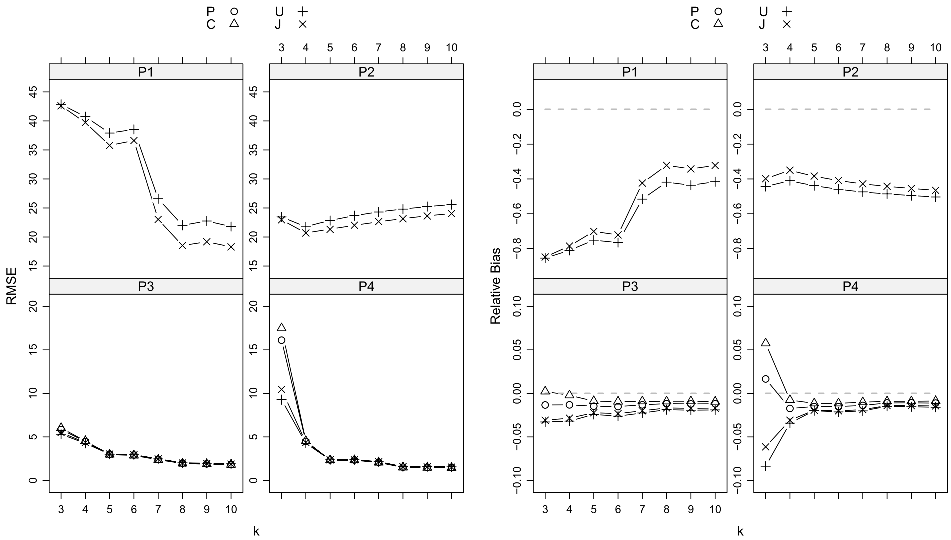

Figure 1 *Root mean squared-error* (*left*) *and relative bias* (*right*) *of profile* (*P*), *conditional* (*C*), *uniform integrated* (*U*) *and Jeffreys integrated* (*J*) *maximum likelihood estimates for* $N = 50$.

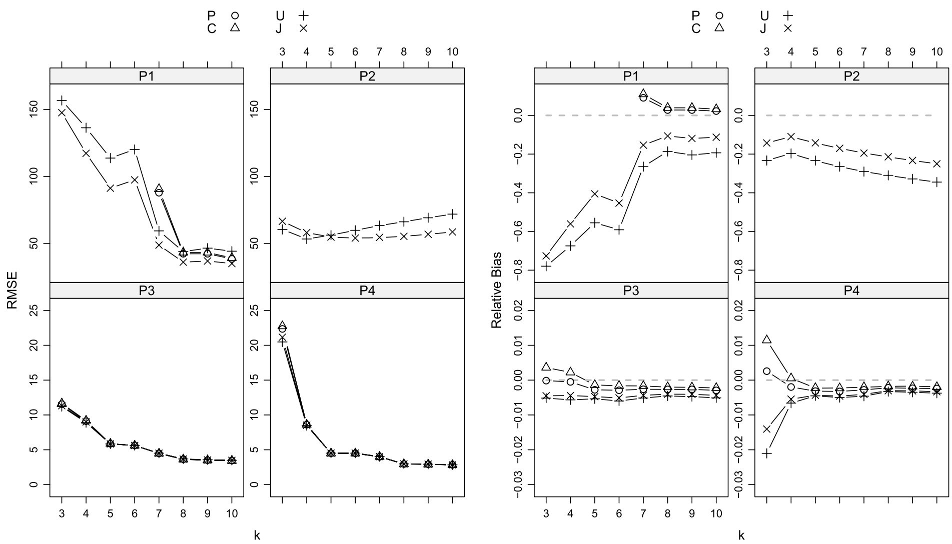

Figure 2 *Root mean squared-error* (*left*) *and relative bias* (*right*) *of profile* (*P*), *conditional* (*C*), *uniform integrated* (*U*) *and Jeffreys integrated* (*J*) *maximum likelihood estimates for* $N = 200$.

may result in ∞ estimate); (v) for low capture probabilities (P_1 and P_2), the profile and conditional estimators assumes greater RMSE values (∞ in many instances) than he uniform and Jeffreys integrated estimators, specially when heterogeneity

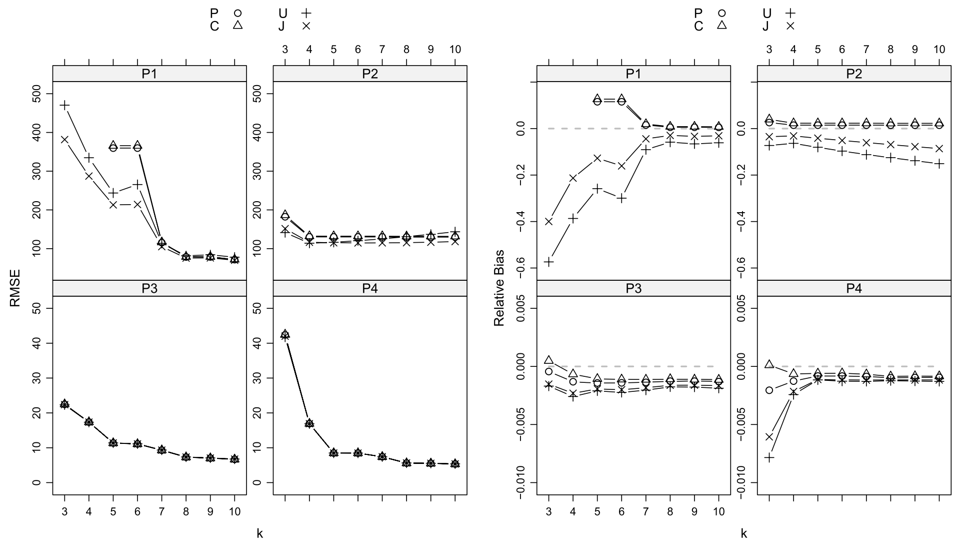

Figure 3 *Root mean squared-error* (*left*) *and relative bias* (*right*) *of profile* (*P*), *conditional* (*C*), *uniform integrated* (*U*) *and Jeffreys integrated* (*J*) *maximum likelihood estimates for* $N = 800$.

of capture probabilities are greater (P_2); (vi) for high capture probabilities (P_3 and P_4), the RMSE values for the estimators are quite similar when $N \geq 200$.

Based on the above considerations, we suggest the use of the profile or the conditional estimators when $N \geq 400$ and k is moderate to large ($k \geq 6$); when $N = 100$ or 200, the number of capture occasions is large ($k \geq 8$) and capture probabilities are not very heterogeneous (P_1 and P_2); and when $N = 50$, the number of capture occasions is large ($k \geq 8$) and the capture probabilities are in P_3 or P_4. In the other cases, when there is a combination of small number of elements N, small number of capture occasions k and low capture probability $\mathbf{p}$, there is not enough information to the profile and conditional estimator to have good performance and the Jeffreys estimator is preferable.

5.2 Confidence intervals

For the sth simulated observation and each combination of fixed values of N, $\mathbf{p}$ and k, it was determined the 95% confidence interval $[N_L^{\mathcal{P}}(s), N_U^{\mathcal{P}}(s)]$ (profile), $[N_L^{\mathcal{C}}(s), N_U^{\mathcal{C}}(s)]$ (conditional), $[N_L^{\mathcal{U}}(s), N_U^{\mathcal{U}}(s)]$ (uniform integrated) and $[N_L^{\mathcal{J}}(s), N_U^{\mathcal{J}}(s)]$ (Jeffreys integrated) and $[N_L^{\mathcal{W}}(s), N_U^{\mathcal{W}}(s)]$ (Wald), $s = 1, 2, \ldots, 1000$. To inspect the frequentists properties of the confidence intervals, we define the following descriptive measures

$$\mathrm{CP}(\mathcal{M}) = \frac{1}{1000} \sum_{k=1}^{S} \mathbb{1}_{[N_L^{\mathcal{M}}(s), N_U^{\mathcal{M}}(s)]}(N),$$

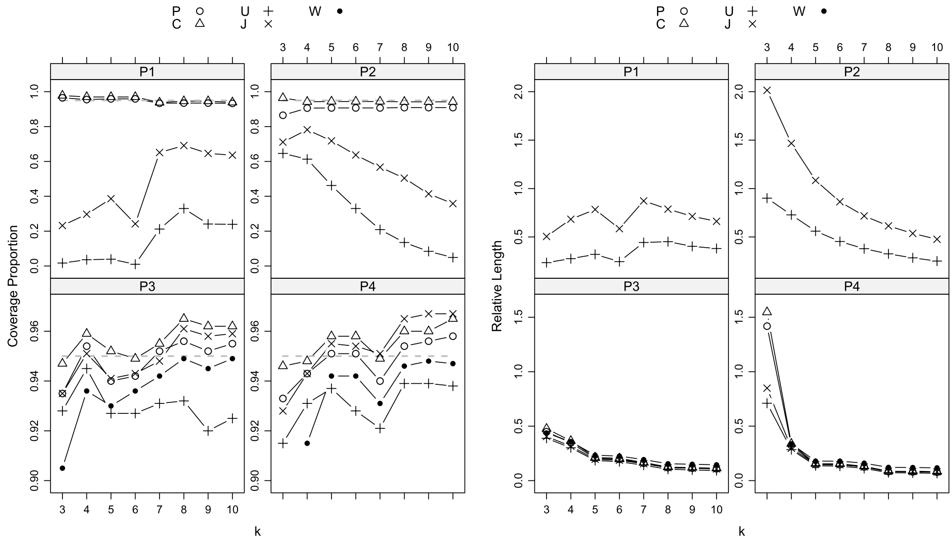

Figure 4 *Coverage proportion (left) and relative length (right) of profile (P), conditional (C), uniform integrated (U), Jeffreys integrated (J) and Wald (W) confidence intervals for* $N = 50$.

$$\mathrm{RLEN}(\mathcal{M}) = \frac{1}{1000} \sum_{k=1}^{S} \frac{N_U^{\mathcal{M}}(s) - N_L^{\mathcal{M}}(s)}{N},$$

where $\mathbb{1}_A$ denotes the indicator function for the set A, CP is the coverage proportion, RLEN is the relative mean length of intervals and $\mathcal{M}$ assumes $\mathcal{P}$, $\mathcal{C}$, $\mathcal{U}$, $\mathcal{J}$ and $\mathcal{W}$. Figures 4, 5 and 6 display the above descriptive measures as function of k for the values of $N = 50, 200, 800$, respectively.

By the analysis of the graphs of CP and RLEN as function of k, we can observe that: (i) increasing the number of capture occasions k and the number of elements N results in smaller values of RLEN and CP closer to the nominal 0.95 confidence level; (ii) increasing the capture probabilities results in CP closer to the 0.95 confidence level and smaller values of RLEN; (iii) the intervals can be increasingly ordered by RLEN as: uniform, Jeffeys, Wald, profile and conditional; (iv) the profile, conditional have very similar CP and RLEN values; (v) the uniform interval has very poor CP performance compared to the others, particularly for capture probabilities in P_1 and P_2; (vi) for $N \geq 400$, all the intervals have similar performance when the capture probabilities are in P_3 and P_4, Jeffreys interval have superior performance when the capture probabilities are in P_1 and P_2; (vii) for $N = 50, 100, 200$, the profile, conditional and Wald intervals have CP slightly closer to the nominal level in comparison to the Jeffreys interval, but have greater RLEN values, in many instances ∞ (or not defined in the case of Wald interval) when $\mathbf{p}$ is in P_1 and P_2.

By the consideration of a combination of the criteria CP closer to 0.95 and smaller RLEN, Jeffreys interval had the best performance when $N \leq 200$, particu-

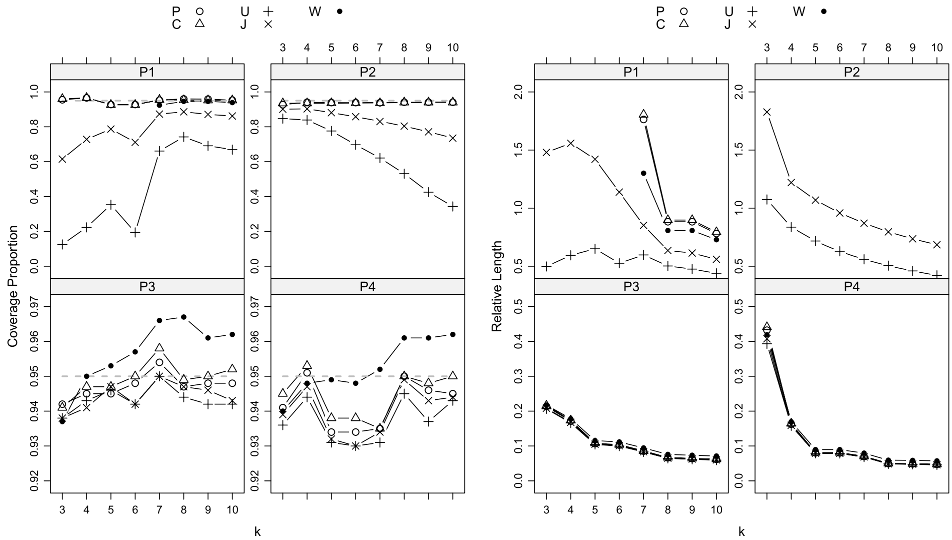

Figure 5 *Coverage proportion* (*left*) *and relative length* (*right*) *of profile* (*P*), *conditional* (*C*), *uniform integrated* (*U*), *Jeffreys integrated* (*J*) *and Wald* (*W*) *confidence intervals for* $N = 200$.

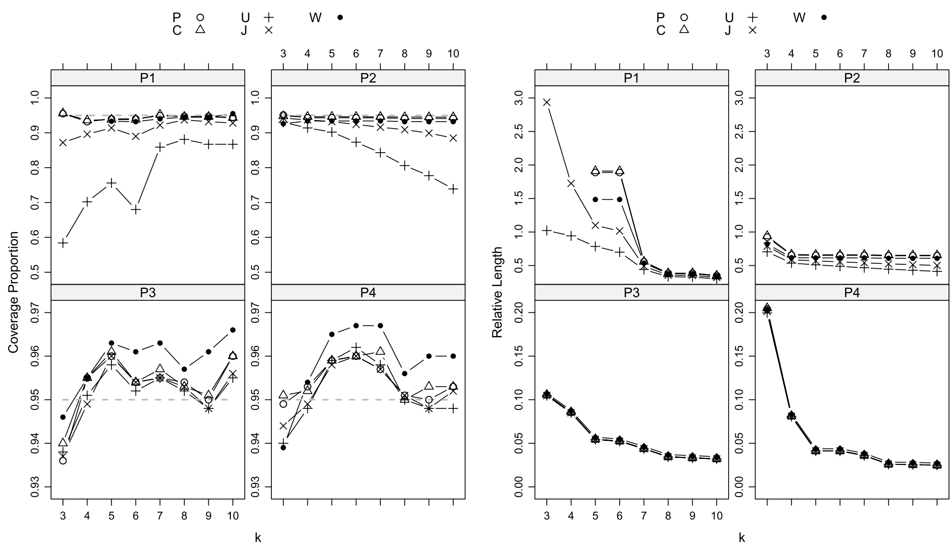

Figure 6 *Coverage proportion* (*left*) *and relative length* (*right*) *of profile* (*P*), *conditional* (*C*), *uniform integrated* (*U*), *Jeffreys integrated* (*J*) *and Wald* (*W*) *confidence intervals for* $N = 800$.

larly with capture probabilities in P_1 and P_2. For $N \geq 400$, the profile, conditional, Jeffreys and Wald intervals have very similar performance. The uniform interval had the worst performance over all considered scenarios.

6 Numerical example

In this section, we illustrate the proposed method on a real data-set obtained from a capture–recapture experiment of "Redear Sunfish" in Gordy Lake, Indiana (USA) (Gerking, 1953). In this experiment, $r = 138$ distinct fishes were captured and the number of captures in each of the $k = 14$ occasions are shown in Table 2. This example was used by Ricker (1975) to illustrate frequentist methods and by Castledine (1981), Smith (1988, 1991), George and Robert (1992) to illustrate Bayesian methods.

For these data, the graphics of the profile, conditional, uniform and Jeffreys integrated relative log-likelihoods (divided by the maximum absolute value of the log-likelihood) are presented in Figure 7. We can observe that the profile and conditional log-likelihoods decrease slower (after the maximum value) than the uniform and Jeffreys integrated log-likelihoods. This characteristic will imply in narrower confidence intervals obtained from the uniform and Jeffreys integrated likelihoods relatives to those obtained from the profile and conditional likelihoods.

For this data-set, the maximum likelihood estimates of $(\widehat{N})$ and the 95% confidence interval for each discussed likelihood are displayed in Table 3. Using a Bayesian approach, Smith (1991) considers the Jeffreys prior $(\pi(N) \propto 1/N)$ for

Table 2 *Number of captures in each occasion*

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n_i	10	27	17	7	1	5	6	15	9	18	16	5	7	19

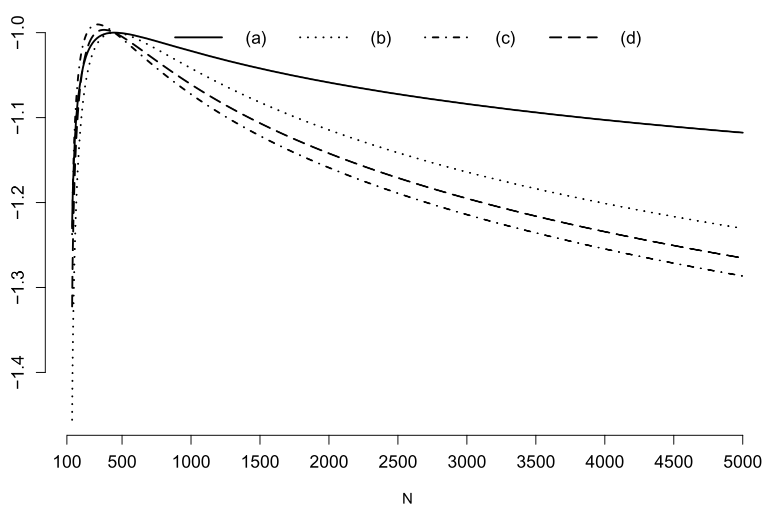

Figure 7 (a) *Profile*, (b) *conditional*, (c) *uniform Integrated and* (d) *Jeffreys integrated relative log-likelihood for the Gordy Lake data.*

Table 3 *Maximum likelihood and 95% confidence interval estimates for the Gordy Lake data*

Method	$\widehat{N}$	95 % CI
Wald	–	[290, 600]
Profile	445	[325, 654]
Conditional	448	[326, 658]
Uniform integrated	322	[254, 428]
Jeffreys integrated	369	[282, 510]

N and Beta(a, b) independent prior distributions for the capture probabilities $\mathbf{p}$, obtaining posterior estimates for N choosing different values of a and b. In the cases of uniform ($a = 1$, $b = 1$) and Jeffreys ($a = 0.5$, $b = 0.5$) prior distribution for $\mathbf{p}$, the 95% confidence intervals $[255, 434]$ and $[281, 520]$ for N are reported, respectively. From Table 3, we conclude that the intervals obtained using the uniform and Jeffreys integrated likelihoods are similar to those obtained by Smith (1991).

7 Concluding remarks

In this paper, we have discussed some likelihood-based approaches for inference on the number of elements of a population considering different methods for eliminating the nuisance parameters. The profile, conditional, uniform and Jeffreys integrated likelihood functions are presented and estimation based on each likelihood function is discussed. The integrated likelihood functions are obtained by means of eliminating the nuisance parameter by integration with respect to a noninformative density function for the capture probabilities. In situations with little information, that is, small number of elements, small capture probabilities and small number of capture occasions, the profile and conditional likelihoods performs poorly and the integrated likelihoods provide an alternative method of estimation with better frequentist performance. In such situations, Jeffreys integrated likelihood achieved the best performance over all considered estimators for both point and interval estimation, the profile and conditional estimators may not exist, and the popular Wald interval may not be defined. Furthermore, the lower limit of all likelihood-based intervals considered are no less than r, condition not guaranteed to hold for the Wald confidence interval.

Appendix: Proof of Theorem 1

In the proof of Theorem 1, we will use the following lemma.

Lemma 1. *For all $y > 0$ and positive integer t, the following inequality holds*:

$$\log(y+t) - \log(y) < \frac{1}{2(y+t)} + \sum_{i=1}^{t} \frac{1}{y+t-i} - \frac{1}{2y}$$
$$= \frac{1}{2(y+t)} + \sum_{i=1}^{t-1} \frac{1}{y+t-i} + \frac{1}{2y}.$$

Proof. Let $y > 0$ be fixed.

First, let us prove the assertion for the case $t = 1$, i.e.,

$$\log(y+1) - \log(y) < \frac{1}{2}\left(\frac{1}{y+1} + \frac{1}{y}\right).$$

Let $f(x) = 1/x$ be defined for $x > 0$. Since f is convex, the line segment joining the points $(y, 1/y)$ and $(y+1, 1/(y+1))$ lies above the graph of f in the interval $(y, y+1)$, that is,

$$\frac{1}{y+\xi} < \frac{1}{y} + \left(\frac{1}{y+1} - \frac{1}{y}\right)\xi, \qquad 0 < \xi < 1.$$

Thus,

$$\begin{aligned}\log(y+1) - \log(y) &= \int_0^1 \frac{1}{y+\xi}\, d\xi \\ &< \int_0^1 \left[\frac{1}{y} + \left(\frac{1}{y+1} - \frac{1}{y}\right)\xi\right] d\xi \qquad (12) \\ &= \frac{1}{2}\left(\frac{1}{y+1} + \frac{1}{y}\right).\end{aligned}$$

For the case $t \geq 2$, we see that

$$\log(y+t) - \log(y) = \sum_{i=1}^{t} \left[\log(y+t+1-i) - \log(y+t-i)\right].$$

From (12), it follows that

$$\begin{aligned}\log(y+t) - \log(y) &< \sum_{i=1}^{t} \left[\frac{1}{2}\left(\frac{1}{y+t-i+1} + \frac{1}{y+t-i}\right)\right] \\ &= \frac{1}{2(y+t)} + \sum_{i=1}^{t-1} \frac{1}{y+t-i} + \frac{1}{2y} \\ &= \frac{1}{2(y+t)} + \sum_{i=1}^{t} \frac{1}{y+t-i} - \frac{1}{2y}.\end{aligned}$$

□

Theorem 1. *The maximum profile likelihood estimator, $\widehat{N}^{\mathcal{P}}$, is no greater than the conditional maximum likelihood estimator $\widehat{N}^{\mathcal{C}}$. The estimator $\widehat{N}^{\mathcal{P}}$ is finite if, and only if, $r < n_1 + \cdots + n_k$.*

Proof. Observe that if $\widehat{N}^{\mathcal{C}} = \infty$, then $\widehat{N}^{\mathcal{P}} \leq \widehat{N}^{\mathcal{C}}$ is obviously true. Thus, suppose that $\widehat{N}^{\mathcal{C}} < \infty$ and consider the ratios $R^{\mathcal{C}}(N) = K^{\mathcal{C}}(N+1)/K^{\mathcal{C}}(N)$ and $R^{\mathcal{P}}(N) = K^{\mathcal{P}}(N+1)/K^{\mathcal{P}}(N)$, which are given by the expressions

$$R^{\mathcal{C}}(N) = \left[1 - \frac{r}{N+1}\right]^{-1} \prod_{j=1}^{k} \left(1 - \frac{n_j}{N+1}\right), \tag{13}$$

$$\begin{aligned} R^{\mathcal{P}}(N) = {} & \left[1 - \frac{r}{N+1}\right]^{-1} \prod_{j=1}^{k} \left(1 - \frac{n_j}{N+1}\right) \\ & \times \prod_{j=1}^{k} \left\{ \left(1 + \frac{1}{N - n_j}\right)^{N - n_j} \frac{N^N}{(N+1)^{N+1}} \right\} \end{aligned} \tag{14}$$

for $N \geq r$.

The ratios are useful to determine the behavior of the likelihood function. Note that if the ratio in N is less than 1, the likelihood function in $N+1$ is less than that in N; if it is equal to 1, the likelihood function in $N+1$ is equal to that in N; and, if is greater than 1, the likelihood in $N+1$ is greater than that in N.

Leite, Oishi and Pereira (1988) have proved that: (i) if $r < n_1 + \cdots + n_k$, then the maximum conditional likelihood estimate $\widehat{N}^{\mathcal{C}}$ is finite and $R^{\mathcal{C}}(N) < 1$ for all $N \geq \widehat{N}^{\mathcal{C}}$; (ii) if $r = n_1 + \cdots + n_k$, then $R^{\mathcal{C}}(N) > 1$ for all $N \geq r$, which implies that $\widehat{N}^{\mathcal{C}} = +\infty$.

Since $(1 + 1/k)^k$ is an increasing sequence in k, it follows that

$$\prod_{j=1}^{k} \left\{ \left(1 + \frac{1}{N - n_j}\right)^{N - n_j} \frac{N^N}{(N+1)^{N+1}} \right\} < 1,$$

which implies from (13) and (14) that $R^{\mathcal{P}}(N) < R^{\mathcal{C}}(N)$ for all $N \geq r$. Therefore, the maximum profile likelihood estimate is always less than or equal to the conditional one. As an immediate consequence, it follows that $\widehat{N}^{\mathcal{P}}$ is finite if $r < n_1 + \cdots + n_k$.

To prove the reciprocal, it suffices to show that if $r = n_1 + \cdots + n_k$ then $K^{\mathcal{P}}$ is an strictly increasing function. Considering $r = n_1 + \cdots + n_k$, let f be an extension to real interval $[r, \infty)$ of the function $K^{\mathcal{P}}(N)$ given by

$$f(x) = \sum_{i=0}^{r-1} \log(x - i) + \sum_{i=1}^{k} \big[(x - n_i)\log(x - n_i) - x\log(x)\big], \qquad x \geq r, \tag{15}$$

where we assume that $0 \log(0) = 0$.

Thus, we shall prove that the $f'(x) > 0$ for all $x > r$, that is,

$$f'(x) = \sum_{i=0}^{n-1} \frac{1}{x-i} + \sum_{i=1}^{k} \big[\log(x-n_i) - \log(x)\big] > 0, \qquad x > r,$$

which is the same as proving that

$$\sum_{i=1}^{k} \big[\log(x) - \log(x-n_i)\big] < \sum_{i=0}^{n_1+n_2+\cdots+n_k-1} \frac{1}{x-i} \tag{16}$$

for any choice of values $n_1, n_2, \ldots, n_k \in \mathbb{N}^* = \{1, 2, \ldots\}$ and any real number $x > n_1 + n_2 + \cdots + n_k$. The relation (16) will be proved by finite induction on k.

From Lemma 1, it follows that

$$\log(x) - \log(x-n_i) < \frac{1}{2x} + \sum_{j=1}^{n_i} \frac{1}{x-j} - \frac{1}{2(x-n_i)} \tag{17}$$

$$= \frac{1}{2x} + \sum_{j=1}^{n_i-1} \frac{1}{x-j} + \frac{1}{2(x-n_i)} \tag{18}$$

for $x > n_i$, $i = 1, 2, \ldots, k$.

Suppose that $k = 2$ and, without loss of generality, that $n_1 \geq n_2$, it follows from (17) and (18) that

$$\begin{aligned}
&\sum_{i=1}^{2} \big[\log(x) - \log(x-n_i)\big] \\
&\quad < \frac{1}{2x} + \sum_{j=1}^{n_1} \frac{1}{x-j} - \frac{1}{2(x-n_1)} \\
&\qquad + \frac{1}{2x} + \sum_{j=1}^{n_2-1} \frac{1}{x-j} + \frac{1}{2(x-n_2)} \\
&\quad = \frac{1}{x} + \sum_{j=1}^{n_1} \frac{1}{x-j} + \sum_{j=1}^{n_2-1} \frac{1}{x-j} + \frac{1}{2}\left(\frac{1}{x-n_2} - \frac{1}{x-n_1}\right) \\
&\quad < \frac{1}{x} + \sum_{j=1}^{n_1} \frac{1}{x-j} + \sum_{j=1}^{n_2-1} \frac{1}{x-n_1-j} \\
&\quad = \sum_{j=0}^{n_1+n_2-1} \frac{1}{x-j},
\end{aligned}$$

$x > n_1 + n_2$, which proves (16) for $k = 2$.

Let us assume that assertion (16) holds for $k = k'$, $k' \geq 2$. Under this hypothesis, we will prove the assertion (16) for $k = k' + 1$. Let $n_1, \ldots, n_{k'}, n_{k'+1}$ in $\mathbb{N}^* = \{1, 2, \ldots\}$ and $x > n_1 + \cdots + n_{k'} + n_{k'+1}$. Since $x > n_1 + \cdots + n_{k'}$, it follows from induction hypothesis that

$$\sum_{i=1}^{k'} [\log(x) - \log(x - n_i)] < \sum_{j=0}^{n_1+n_2+\cdots+n_{k'}-1} \frac{1}{x - j},$$

and using Lemma 1, it follows that

$$\begin{aligned}
&\sum_{i=1}^{k'+1} [\log(x) - \log(x - n_i)] \\
&\quad = \sum_{i=1}^{k'} [\log(x) - \log(x - n_i)] + \log(x) - \log(x - n_{k'+1}) \\
&\quad < \sum_{j=0}^{n_1+n_2+\cdots+n_{k'}-1} \frac{1}{x - j} + \sum_{j=1}^{n_{k'+1}} \frac{1}{x - j} + \frac{1}{2}\left(\frac{1}{x} - \frac{1}{x - n_{k'+1}}\right) \\
&\quad < \sum_{j=0}^{n_1+n_2+\cdots+n_{k'}-1} \frac{1}{x - j} + \sum_{j=1}^{n_{k'+1}} \frac{1}{(x - n_1 - \cdots - n_{k'} + 1) - j} \\
&\quad = \sum_{j=0}^{n_1+n_2+\cdots+n_{k'+1}-1} \frac{1}{x - j},
\end{aligned}$$

which implies (16) for $k = k' + 1$. □

References

Amstrup, S. C., McDonald, T. L. and Manly, B. F. J., eds. (2003). *Handbook of Capture–Recapture Analysis*. London: John Wiley.

Basu, D. (1977). On the elimination of nuisance parameters. *Journal of the American Statistical Association* **72**, 355–366. MR0451477

Basu, S. and Ebrahimi, N. (2001). Bayesian capture–recapture methods for error detection and estimation of population size: Heterogeneity and dependence. *Biometrika* **88**, 269–279. MR1841274

Berger, J. O., Liseo, B. and Wolpert, L. (1999). Integrated likelihood methods for eliminating nuisance parameters. *Statistical Science* **14**, 1–28. MR1702200

Boender, C. G. E. and Rinooy Kan, A. H. G. (1987). A multinomial Bayesian approach to the estimation of population and vocabulary size. *Biometrika* **74**, 849–856. MR0919853

Bolfarine, H., Leite, J. G. and Rodrigues, J. (1992). On the estimation of the size of a finite and closed population. *Biometrical Journal* **34**, 577–593. MR1179478

Carle, F. L. and Strub, M. R. (1978). A new method for estimating population size from removal data. *Biometrics* **34**, 621–630.

Castledine, B. J. (1981). A Bayesian analysis of multiple-recapture sampling for a closed population. *Biometrika* **67**, 197–210. MR0614956

Chao, A. (1989). Estimating population size for sparse data in capture–recapture data with unequal catchability. *Biometrics* **43**, 783–791. MR0920467

Chao, A. (2001). An overview of closed capture–recapture models. *Journal of Agricultural, Biological, and Environmental Statistics* **6**, 158–175.

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. *Statistics in Medicine* **20**, 3123–3157.

Chapman, D. G. (1954). The estimation of biological populations. *The Annals of Mathematical Statistics* **25**, 1–15. MR0061336

Cormack, R. M. (1968). The statistics of capture–recapture methods. *Oceanography and Marine Biology* **6**, 455–506.

Cormack, R. M. (1989). Log-linear models for capture–recapture. *Biometrics* **45**, 395–413.

Cormack, R. M. (1992). Interval estimation for mark–recapture studies of closed populations. *Biometrics* **48**, 567–576. MR1173495

Cox, D. R. (1975). Partial likelihood. *Biometrika* **62**, 269–276. MR0400509

Darroch, J. N. (1958). The multiple-recapture census: I. Estimation of a closed population. *Biometrika* **45**, 343–359. MR0119360

Darroch, J. N. (1959). The multiple-recapture census: II. Estimation when there is immigration or death. *Biometrika* **46**, 336–351. MR0119361

Evans, M. A., Kim, H. and O'Brien, T. E. (1996). An application of profile-likelihood based confidence interval to capture–recapture estimators. *Journal of Agricultural, Biological and Environmental Statistics* **1**, 131–140. MR1807771

Fegatelli, D. A. and Tardella, L. (2013). Improved inference on capture recapture models with behavioural effects. *Statistical Methods and Applications* **22**, 45–66. MR3038181

Garthwaite, P. H. and Buckland, S. T. (1990). Analysis of a multiple-recapture census by computing conditional probabilities. *Biometrics* **46**, 231–238. MR1059113

George, E. I. and Robert, C. P. (1992). Capture–recapture estimation via Gibbs sampling. *Biometrika* **79**, 677–683. MR1209469

Gerking, S. D. (1953). Vital statistics of the fish population of Gordy Lake, Indiana. *Transactions of the American Fisheries Society* **82**, 48–67.

Jolly, G. M. (1965). Explicit estimates from capture–recapture data with both death and immigration-stochastic model. *Biometrika* **52**, 225–247. MR0210227

Laplace, P. S. (1783). Sur les naissances, les mariages, et les morts. Paris.

Lee, A. J. (2002). Effect of list errors on the estimation of population size. *Biometrics* **58**, 185–191. MR1891378

Lee, A. J., Seber, G. A. F., Holden, J. K. and Huakau, J. T. (2001). Capture–recapture, epidemiology, and list mismatches: Several lists. *Biometrics* **57**, 707–713. MR1859807

Leite, J. G., Oishi, J. and Pereira, C. A. B. (1987). Exact maximum likelihood estimate of a finite population size. *Probability in the Engineering and Informational Sciences* **1**, 225–236.

Leite, J. G., Oishi, J. and Pereira, C. A. B. (1988). A note on the exact maximum likelihood estimation of the size of a finite and closed population. *Biometrika* **75**, 178–180. MR0932837

Lincoln, F. C. (1930). Calculating waterfowl abundance on the basis of banding returns. Technical report, U.S. Department of Agriculture Circular.

McDonald, J. F. and Palanacki, D. (1989). Interval estimation of the size of a small population from a mark–recapture experiment. *Biometrics* **45**, 1223–1231. MR1040635

Nayak, T. K. (1988). Estimating population size by recapture sampling. *Biometrika* **75**, 113–120. MR0932824

Norris, J. L. and Pollock, K. H. (1996). Nonparametric MLE under two closed capture–recapture models with heterogeneity. *Biometrics* **52**, 639–649.

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

Petersen, C. G. (1896). The yearly immigration of young plaice into the Limfjord from the German Sea. Technical report, Report of the Danish Biological Station.

Pickands, J. and Raghavachari, M. (1974). Exact and asymptotic inference for the size of a population. *Biometrika* **74**, 355–363. MR0903136

Pledger, S. (2000). Unified maximum likelihood estimates for closed capture–recapture models using mixtures. *Biometrics* **56**, 434–442.

Pollock, K. H. (2000). Capture–recapture models. *Journal of the American Statistical Association* **95**, 293–296.

Ricker, W. E. (1975). Computation and interpretation of biological statistics of fish populations. Technical report, Department of the Environment, Fisheries and Marine Service.

Sanathanan, L. (1972a). Estimating the size of a multinomial population. *The Annals of Mathematical Statistics* **43**, 142–152. MR0298815

Sanathanan, L. (1972b). Models and estimation methods in visual scanning experiments. *Technometrics* **14**, 813–829.

Sanathanan, L. (1973). A comparison of some models in visual scanning experiments. *Technometrics* **15**, 67–78.

Schwarz, C. J. and Seber, G. A. F. (1999). Estimating animal abundance: Review III. *Statistical Science* **14**, 427–456.

Seber, G. A. F. (1965). A note on multiple recapture census. *Biometrika* **52**, 249–259. MR0210228

Seber, G. A. F. (1982). *The Estimation of Animal Abundance and Related Parameters*. London: John Wiley. MR0686755

Seber, G. A. F. (1986). A review of estimating animal abundance. *Biometrics* **42**, 267–292.

Seber, G. A. F. (1992). A review of estimating animal abundance II. *International Statistical Review* **60**, 129–166.

Seber, G. A. F., Huakau, J. T. and Simmons, D. (2000). Capture–recapture, epidemiology and lists mismatches: Two lists. *Biometrics* **56**, 1227–1232. MR1815625

Seber, G. A. F. and Whale, J. F. (1970). The removal method for two and three samples. *Biometrika* **26**, 393–400.

Smith, P. J. (1988). Bayesian methods for capture–recapture surveys. *Biometrics* **44**, 1177–1189.

Smith, P. J. (1991). Bayesian analyses for a multiple capture–recapture model. *Biometrika* **78**, 399–407. MR1131174

Sprott, D. A. (1981). Maximum likelihood applied to a capture and recapture model. *Biometrics* **37**, 371–375. MR0673044

Thisted, R. and Efron, B. (1987). Did Shakespeare write a newly-discovery poem? *Biometrika* **74**, 445–455. MR0909350

Viveros, R. and Sprott, D. A. (1986). Conditional inference and maximum likelihood in a capture–recapture model. *Communications in Statistics* **15**, 1035–1046.

Wang, X., He, C. Z. and Sun, D. (2007). Bayesian population estimation for small sample capture–recapture data using noninformative prior. *Journal of Statistical Planning and Inference* **137**, 1099–1118. MR2301466

Yip, P. (1991). A martingale estimating equation for capture–recapture experiment in discrete time. *Biometrics* **47**, 1081–1088. MR1141953

L. E. B. Salasar
J. G. Leite
UFSCar—Universidade Federal de São Carlos
São Carlos, SP
Brazil
E-mail: luis.salasar@gmail.com
leite@ufscar.br

F. Louzada
ICMC—Universidade de São Paulo
São Carlos, SP
Brazil
E-mail: louzada@icmc.usp.br