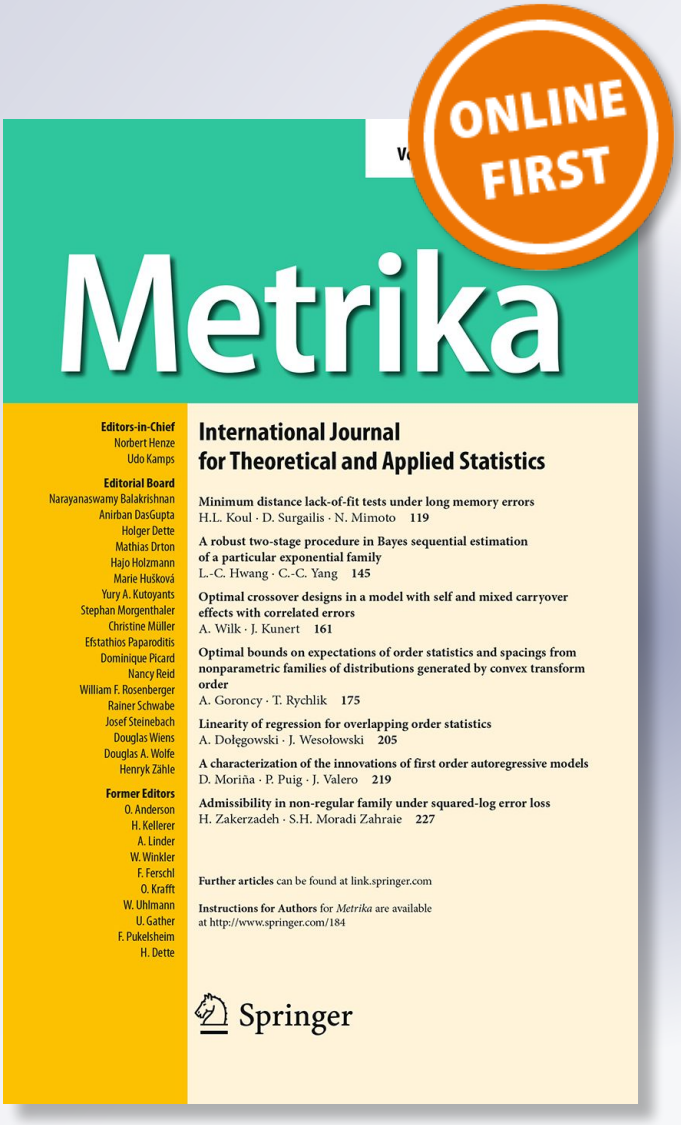
Likelihood ratio confidence interval for the abundance under binomial detectability models

Yang Liu, Yukun Liu, Yan Fan & Han Geng

Metrika
International Journal for Theoretical and Applied Statistics

ISSN 0026-1335

Metrika
DOI 10.1007/s00184-018-0655-2



Springer

Metrika
https://doi.org/10.1007/s00184-018-0655-2

Likelihood ratio confidence interval for the abundance under binomial detectability models

Yang Liu[1] · **Yukun Liu**[1] · **Yan Fan**[2] · **Han Geng**[1]

Received: 21 September 2017

Abstract Binomial detectability models are often used to estimate the size or abundance of a finite population in biology, epidemiology, demography and reliability. Special cases include incompletely observed multinomial models, capture–recapture models, and distance sampling models. The most commonly-used confidence interval for the abundance is the Wald-type confidence interval, which is based on the asymptotic normality of a reasonable point estimator of the abundance. However, the Wald-type confidence interval may have poor coverage accuracy and its lower limit may be less than the number of observations. In this paper, we rigorously establish that the likelihood ratio test statistic for the abundance under the binomial detectability models follows the chisquare limiting distribution with one degree of freedom. This provides a solid theoretical justification for the use of the proposed likelihood ratio confidence interval. Our simulations indicate that in comparison to the Wald-type confidence interval, the likelihood ratio confidence interval not only has more accurate coverage rate, but also exhibits more stable performance in a variety of binomial detectability models. The proposed interval is further illustrated through analyzing three real data-sets.

Keywords Abundance · Binomial detectability models · Capture-recapture models · Confidence interval · Distance sampling models

✉ Yukun Liu
ykliu@sfs.ecnu.edu.cn

[1] School of Statistics, East China Normal University, Shanghai, China

[2] School of Statistics and Information, Shanghai University of International Business and Economics, Shanghai, China

Published online: 30 March 2018

1 Introduction

Knowing the size of a finite population is of great importance in many different areas. Examples include animal abundances in fisheries and wildlife biology, cf. Pollock (2000) and Borchers et al. (2002, 2015), frequencies of diseases in epidemiological studies, cf. Chao et al. (2001), population sizes in demography, cf. Hogan (2000), Chen and Lloyd (2002), and the number of faults in reliability system, cf. Barnard et al. (2003). Given that the study period is often short, it is reasonable to make a closed population assumption. Namely, there is no birth, death or migrations, and the population size remains unchanged throughout the study. The problem of interest is to make inference for the population size or abundance when data are available.

To this end, there have been a number of models developed in the literature, such as incompletely observed multinomial models, cf. Sanathanan (1972), mark-recapture models, cf. Otis et al. (1978), distance sampling models, cf. Buckland et al. (2001), and a unifying model for capture–recapture and distance sampling, cf. Borchers et al. (2015). The author in Fewster and Jupp (2009) pointed out that many of the aforementioned models can be unified to be binomial detectability models. Specifically, suppose the population of interest consists of N individuals and there are n observed individuals with covariate x_i's ($i = 1, 2, \ldots, n$). Examples for covariates include wing lengths of birds, body weights of animals, capture histories in a capture–recapture study, and distance of an individual from the observer in distance sampling. See Fewster and Jupp (2009) and Liu et al. (2017). A binomial detectability model implies that the joint distribution of all the observations has the form

$$\mathrm{pr}(n, x_1, \ldots, x_n; N, \theta) = \binom{N}{n} \{p(\theta)\}^n \{1 - p(\theta)\}^{N-n} \prod_{i=1}^{n} k(x_i; \theta), \qquad (1)$$

where $p(\theta)$ and $k(x; \theta)$ are pre-specific functions with $0 < p(\theta) < 1$ and an unknown parameter θ. A binomial distribution with success probability $p(\theta)$ is used to model n and a parametric model $k(x; \theta)$ is used to model the probability density of an ideal covariate given that the corresponding individual is detected. Following Fewster and Jupp (2009), we treat N as continuous and interpret $\binom{N}{n}$ as $\Gamma(N+1)/\{\Gamma(N-n+1)\Gamma(n+1)\}$.

Besides the binomial detectability models, an alternative modelling strategy is to put a parametric model on the probability of capture or being observed given covariates, cf. Alho (1990) and Huggins (1989). The resulting model (Alho–Huggins model for short) has attracted much attention so far, cf. Borchers et al. (1998, 2002), Marques and Buckland (2004) and Liu et al. (2017). It conditions on the covariates and does not need to model their distributions, while in the binomial detectability models, the conditional distribution of the covariates given detection is required. The Alho–Huggins model is mainly used for capture–recapture data, in which it is a common sense to have multiple capture occasions; there is no method available with a single capture occasion. By contrast, the binomial detectability models have much wider applications including not only capture–recapture data, but also incompletely observed multinomial data, cf. Sanathanan (1972) and distance sampling, cf. Buckland et al. (2001). Finally, the

goodness of $k(x;\theta)$ in the binomial detectability models can always be checked using data, however this is not often feasible for the capture probability function in the Alho–Huggins model.

Regarding estimation in the binomial detectability models, existing interval estimators for the abundance N are generally Wald-type. See Fewster and Jupp (2009) for example. Their construction generally consists of three steps. First, one needs to find a reasonable point estimator, say $\breve{N}$. Second, the asymptotic normality of $\breve{N}$,

$$N^{-1/2}(\breve{N}-N)=N^{1/2}\{\log(\breve{N})-\log(N)\}+o_p(1)\xrightarrow{d}N(0,\sigma^2)$$

is established, where $\xrightarrow{d}$ means convergence in distribution, and a consistent estimator $\breve{\sigma}^2$ of σ^2 is constructed. Finally, the Wald-type confidence interval for N is constructed as $\breve{N}\pm z_{1-\alpha/2}\breve{\sigma}\sqrt{\breve{N}}$ or $\breve{N}\exp(\pm z_{1-\alpha/2}\breve{N}^{-1/2}\breve{\sigma})$, where $z_{1-\alpha/2}$ is the lower $1-\alpha/2$ quantile of the standard normal distribution. Although the two Wald-type confidence intervals have the same asymptotically correct coverage probability, the latter is generally superior to the former in terms of coverage probability, cf. Fewster and Jupp (2009). Hereafter we take the latter as the Wald-type confidence interval of N. However, even in the simplest case, the finite-sample distribution of the abundance estimator is far from normal and strongly skewed to the right; see Evans and Bonett (1994). This may lead to severe under-coverage or over-coverage of the corresponding Wald-type confidence intervals. We have similar observations in our simulation study. Moreover, the lower limit of the Wald-type confidence interval may be less than the number of individuals captured, which is clearly absurd. The necessary estimation of an asymptotic variance may further inflate the variation of the Wald-type confidence interval. The poor performance of the Wald-type confidence interval motivates us to investigate the problem and propose new confidence intervals for the abundance.

Instead of Wald-type interval estimators, the authors in Cormack (1992) and Evans et al. (1996) proposed the profile likelihood confidence interval for the abundance from capture–recapture data. This interval turns out to be the likelihood ratio confidence interval under the multinomial likelihood, and was shown to outperform the Wald-type interval. However, it can not incorporate covariates and may involve too many nuisance parameters when the number of capture occasions is large, leading to potential efficiency loss. In Liu et al. (2017) the authors proposed empirical likelihood ratio confidence intervals based on the Alho–Huggins model. They found that the empirical likelihood ratio confidence interval overcomes most shortcomings of the Wald-type confidence intervals. However, their confidence intervals are based on the Alho–Huggins model and do not apply directly to the binomial detectability models.

In Fewster and Jupp (2009) the authors mentioned the use of likelihood ratio confidence intervals for N under the binomial detectability models, but they did not provide a theoretical justification. In this paper, we rigorously establish the asymtptotic chisquare distribution of the likelihood ratio test statistic for the abundance N. This lays the foundation for the likelihood ratio confidence intervals, and also implies that the likelihood ratio confidence intervals have asymptotically correct coverage probabilities. Compared with the Wald-type interval, the likelihood ratio confidence interval is one-step and free of variance estimation; its lower limit is no less than the

sample size n. In addition, our simulations indicate that the likelihood ratio confidence interval has much more accurate coverage probabilities than the Wald-type interval, especially when the detection probability is low.

The rest of the paper is organized as follows. In Sect. 2, we review the maximum full and conditional likelihood estimators of (N, θ) under the binomial detectability models, and their asymptotic normality. We then formally establish the chisquare limiting distributions of both the likelihood ratio statistics for N and (N, θ). We report a simulation study in Sect. 3. In Sect. 4, we illustrate the proposed interval estimation method by applying it to three real data-sets. All the technical derivations are postponed to the "Appendix" for clarity.

2 Likelihood ratio interval estimation

With the notation previously defined, the log-likelihood of (N, θ) up to a constant is

$$\begin{aligned}\ell(N,\theta) &= \log\{\Gamma(N+1)\} - \log\{\Gamma(N-n+1)\} \\ &\quad + (N-n)\log\{1-p(\theta)\} + n\log\{p(\theta)\} + \ell_c(\theta)\end{aligned}$$

with $\ell_c(\theta) = \sum_{i=1}^{n} \log\{k(x_i;\theta)\}$. This log-likelihood is exactly the same as Eq. (4) in Fewster and Jupp (2009). Under the binomial detectability model, the two most popular point estimators of (N, θ) are the maximum full likelihood estimator

$$(\hat{N}, \hat{\theta}) = \arg_{N,\theta} \max \ell(N, \theta) \tag{2}$$

and the maximum conditional likelihood estimator $(\tilde{N}, \tilde{\theta})$, where

$$\tilde{N} = n/p(\tilde{\theta}), \quad \tilde{\theta} = \arg_{\theta} \max \ell_c(\theta). \tag{3}$$

The references Sanathanan (1972) and Fewster and Jupp (2009) disclosed that these two estimators follow the same asymptotic normal distribution.

The following lemma, from Fewster and Jupp (2009), establishes the equivalence and the common limiting distribution of the two maximum likelihood estimators. Let (N_0, θ_0) be the true value of (N, θ) and $I(\theta) = \mathbb{E}\{\partial \log\{k(X;\theta)\}/\partial\theta\}^{\otimes 2}$, where $X \sim k(x;\theta)$ and $A^{\otimes 2} = AA^{\top}$ for a vector or matrix A. Define $p_1(\theta) = \partial \log p(\theta)/\partial\theta$ and

$$\Sigma = \frac{p(\theta_0)}{1-p(\theta_0)} \begin{pmatrix} 1 & p_1^{\top}(\theta_0) \\ p_1(\theta_0) & \{p_1(\theta_0)\}^{\otimes 2} + \{1-p(\theta_0)\} I(\theta_0) \end{pmatrix}.$$

Lemma 1 (See Fewster and Jupp 2009) *Suppose the conditional density $k(x;\theta)$ satisfies the regularity conditions in* "Appendix A"*, and $p(\theta)$ satisfies $0 < p(\theta) < 1$ for all θ and has a continuous second derivative. If the matrix Σ is positive definite, then as $N_0 \to \infty$,*

(i) $\hat{N} - \tilde{N} = O_p(1)$;

(ii) *for* $(\breve{N}, \breve{\theta}) = (\hat{N}, \hat{\theta})$ *or* $(\tilde{N}, \tilde{\theta})$,

$$N_0^{1/2} \begin{pmatrix} \log(\breve{N}) - \log(N_0) \\ \breve{\theta} - \theta_0 \end{pmatrix} = \begin{pmatrix} N_0^{-1/2}(\breve{N} - N_0) \\ N_0^{1/2}(\breve{\theta} - \theta_0) \end{pmatrix} + o_p(1) \xrightarrow{d} N(0, \Sigma^{-1});$$

(iii) *for* $\breve{N} = \hat{N}$ *or* $\tilde{N}$,

$$N_0^{1/2}\{\log(\breve{N}) - \log(N_0)\} = N_0^{-1/2}(\breve{N} - N_0) + o_p(1) \xrightarrow{d} N(0, \sigma^2),$$

where $\sigma^2 = \{p_1(\theta_0)^\top I(\theta_0)^{-1} p_1(\theta_0) + 1 - p(\theta_0)\}/\{p(\theta_0)\}$.

Wald-type confidence intervals can hence be constructed according to Lemma 1:

$$\mathcal{I}_{\text{Wald}} = \left[\tilde{N} \exp\left(-z_{1-\alpha/2}\tilde{N}^{-1/2}\tilde{\sigma}\right),\ \tilde{N} \exp\left(z_{1-\alpha/2}\tilde{N}^{-1/2}\tilde{\sigma}\right)\right], \tag{4}$$

where

$$\tilde{\sigma}^2 = \{p_1(\tilde{\theta})^\top \tilde{I}(\tilde{\theta})^{-1} p_1(\tilde{\theta}) + 1 - p(\tilde{\theta})\}/\{p(\tilde{\theta})\}$$

and

$$\tilde{I}(\theta) = (1/n) \sum_{i=1}^{n} \{\partial \log\{k(x_i; \theta)\}/\partial\theta\}^{\otimes 2}.$$

However such intervals may suffer from variance estimation, unreasonable lower limits and severe undercoverage. Given the nice performance of likelihood ratio confidence intervals for N from capture–recapture data, cf. Cormack (1992), Evans et al. (1996) and Liu et al. (2017), we may expect that those under the binomial detectability models would also perform well. Define the likelihood ratio functions of (N, θ) and N as

$$R(N, \theta) = 2\{\ell(\hat{N}, \hat{\theta}) - \ell(N, \theta)\} \quad \text{and} \quad R'(N) = 2\{\ell(\hat{N}, \hat{\theta}) - \ell(N, \hat{\theta}_N)\},$$

where $(\hat{N}, \hat{\theta})$ is defined in Eq. (2) and $\hat{\theta}_N = \arg\max_{\theta_*} \ell(N, \theta_*)$ given N.

Theorem 1 *Suppose the conditions of Lemma 1 are satisfied. As* $N_0 \to \infty$, $R(N_0, \theta_0) \xrightarrow{d} \chi^2_{d+1}$ *and* $R'(N_0) \xrightarrow{d} \chi^2_1$, *where* d *is the dimension of* θ.

According to Theorem 1, we recommend the use of a likelihood ratio confidence interval for N_0 at level $1 - \alpha$:

$$\mathcal{I}_{\text{LR}} = \left\{N : R'(N) \le \chi^2_{1,1-\alpha}\right\}, \tag{5}$$

where $\chi^2_{1,1-\alpha}$ is the $(1 - \alpha)$ quantile of the chisquare distribution with one degree of freedom. Theorem 1 indicates that $\mathcal{I}_{\text{LR}}$ has asymptotically correct coverage probabilities when N_0 is large enough. The authors in Fewster and Jupp (2009) also mentioned

the use of $\mathcal{I}_{\mathrm{LR}}$, but they did not give a theoretical justification. In comparison with $\mathcal{I}_{\mathrm{Wald}}$, there is no need of variance estimation in $\mathcal{I}_{\mathrm{LR}}$. Since the likelihood function makes sense only for $N \geq n$, the lower limit of $\mathcal{I}_{\mathrm{LR}}$ must be no less than n. There is no such a guarantee for $\mathcal{I}_{\mathrm{Wald}}$. Lemma 1 implies that $\mathcal{I}_{\mathrm{Wald}}$ also has asymptotically correct coverage probabilities when N_0 is large enough.

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 Lemma 1, we can show that

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

is asymptotically distributed as $N(0, 1)$. Hence, another Wald-type confidence interval for N based on the conditional likelihood is

$$\mathcal{I}_{\mathrm{Chao}} = \left\{N : |C(N; \tilde{N})| \leq z_{1-\alpha/2}\right\}.$$

An advantage of $\mathcal{I}_{\mathrm{Chao}}$ is that its lower limit is ensured to be larger than n, the number of individuals been captured. In the next section, we shall compare by simulation the finite-sample performance of $\mathcal{I}_{\mathrm{LR}}$, $\mathcal{I}_{\mathrm{Wald}}$ and $\mathcal{I}_{\mathrm{Chao}}$.

3 Simulation

3.1 Simulation set-up

We generate data from four binomial detectability models; the first three are capture–recapture models and the last is a distance sampling model. In capture–recapture experiments, the capture probability of an individual may be affected by time (t; also called capture occasion), behavioral response (b), or heterogeneity between individuals (h). All combinations of these factors produce M_0, M_t, M_b, M_h, M_{th}, M_{bh}, M_{bt}, and M_{tbh} models, see Otis et al. (1978). In all the four models, we choose $N_0 = 100$ as small populations are common in practical situations.

(A) Suppose there are m ($m \geq 2$) capture occasions in a capture–recapture study, and let x denote the total number of captures of a generic subject. We consider the capture–recapture model M_0, which implies that all subjects have the same capture probability, say θ, on each of the m capture occasions. If we take x as a covariate, then the $p(\theta)$ and $k(x; \theta)$ functions corresponding to model M_0 are

$$p(\theta) = 1 - (1 - \theta)^m \quad \text{and} \quad k(x; \theta) = \frac{\binom{m}{x}\theta^x(1 - \theta)^{m-x}}{p(\theta)}, \quad x \in \{1, 2, \ldots, m\}.$$

We choose $m = 5$ and appropriate θ_0 values such that $p(\theta_0) = 0.5$ and 0.8.

(B) We consider the capture–recapture model M_{tb}, in which both capture occasion and behavioral response affect the capture probability. Suppose there are m capture occasions. Let p_j be the capture probability of an individual at its first capture if its first capture is at the jth capture occasion. And let $c_j = \phi p_j$ be the probability of an individual at the jth capture if it was captured previously, we refer to Chao et al. (2000), where $\phi > 0$ is an unknown scalar. The overall underlying parameters constitute $\theta = (p_1, \ldots, p_m, \phi)$. Denote $x = (x_{(1)}, \ldots, x_{(m)})$ a generic history where $x_{(j)} \in \{0, 1\}$, and $x_{(j)} = 1$ means that the individual was captured at the jth occasion and 0 otherwise. If we take the capture history x as a covariate, then the $p(\theta)$ and $k(x; \theta)$ functions corresponding to model M_{tb} are $p(\theta) = 1 - \prod_{j=1}^{m}(1 - p_j)$ and

$$k(x;\theta) = \frac{\prod_{j=1}^{\delta(x)} \left\{ p_j^{x_{(j)}} (1-p_j)^{1-x_{(j)}} \right\} \cdot \prod_{j=\delta(x)+1}^{m} \left\{ (\phi p_j)^{x_{(j)}} (1-\phi p_j)^{1-x_{(j)}} \right\}}{1-\prod_{j=1}^{m}(1-p_j)},$$

where $\delta(x) = \min\{j : x_{(j)} > 0\}$. We choose $m = 10$, $(p_1, p_2, p_3, p_4, p_5, p_6, p_7, p_8, p_9, p_{10}) = (0.2, 0.2, 0.15, 0.24, 0.24, 0.2, 0.2, 0.15, 0.24, 0.24)$, and $\phi_0 = 0.8$ or 1.5.

(C) The third capture–recapture model we consider is model M_h with m capture occasions. Let x denote a characteristic of a generic individual and suppose it follows the exponential distribution with rate θ. Namely, its probability density function is $f(x; \theta) = \theta e^{-\theta x}$. Suppose in each occasion, the probability of a generic individual being captured given its covariate x is $1 - e^{-x}$. If we take x as a covariate, then the $p(\theta)$ and $k(x; \theta)$ functions corresponding to model M_h are $p(\theta) = m/(m + \theta)$ and

$$k(x;\theta) = \theta(m+\theta)(1-e^{-mx})e^{-\theta x}/m.$$

We choose $\theta_0 = 18$ and $m = 2$ or 4.

(D) We consider a distance sampling model. Let $A = 1$ be the area of the region to be sampled and $w = 0.5$ be the half-width of the strip around a single line transect with length 1, where animals might be detected. Then $a = 2w = 1$ is the area covered by this strip. Let x be the perpendicular distance of an individual from the line, which follows a uniform distribution between 0 and 0.5. We choose the commonly used half-normal detection function $\exp(-\theta x^2)$ to model the probability of detecting an animal that is at vertical distance x from the line. If we take x as a covariate, then the $p(\theta)$ and $k(x; \theta)$ functions corresponding to the distance sampling model are

$$p(\theta) = 2(\pi/\theta)^{1/2}\left[\Phi\left\{0.5(2\theta)^{1/2}\right\} - 0.5\right] \quad \text{and}$$

$$k(x;\theta) = \frac{\exp\left(-\theta x^2\right)}{(\pi/\theta)^{1/2}\left[\Phi\left\{0.5(2\theta)^{1/2}\right\} - 0.5\right]}, \quad 0 \le x \le 0.5,$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal. We choose appropriate θ_0 values such that $p(\theta_0) = 0.1$ and 0.25.

3.2 Simulation results

The confidence intervals previously mentioned are two-sided, say $[N_l, N_u]$. From each of them, we can construct two one-sided confidence intervals, $[N_l, \infty]$ (lower limit) and $[n, N_u]$ (upper limit). Their coverage probabilities are nearly $1-\alpha/2$ if that for $[N_l, N_u]$ is $1-\alpha$. In this simulation study, we conducted 2000 simulations for each of the simulation settings and consider three confidence levels 90, 95 and 99%. Tables 1 and 2 report the simulated coverage probabilities of $\mathcal{I}_{\text{LR}}$, $\mathcal{I}_{\text{Wald}}$, $\mathcal{I}_{\text{Chao}}$, and the corresponding one-sided confidence intervals.

To evaluate the performance of a two-sided interval, we also consider the average interval score criterion, cf. Gneiting and Raftery (2007). The interval score of $[N_l, N_u]$ at level $1-\alpha$ is defined to be

$$(N_u - N_l) + \frac{2}{\alpha}(N_0 - N_u)I(N_0 > N_u) + \frac{2}{\alpha}(N_l - N_0)I(N_0 < N_l).$$

The smaller the interval score, the better the confidence interval. The average interval scores of the three two-sided intervals are given in the paratheses in Tables 1 and 2.

It is impressive to see that in scenarios C and D, the proposed interval $\mathcal{I}_{\text{LR}}$ always produces the most accurate coverage probabilities for both one- and two-sided interval estimation and the smallest interval scores for two-sided interval estimation at all the three confidence levels. In scenario A, $\mathcal{I}_{\text{LR}}$ and $\mathcal{I}_{\text{wald}}$ have nearly the same ideal performance in term of coverage accuracy, while $\mathcal{I}_{\text{Chao}}$ often produces undercoverage. In scenario A with $p(\theta_0) = 0.5$, although $\mathcal{I}_{\text{LR}}$ has larger interval scores than $\mathcal{I}_{\text{Wald}}$ and $\mathcal{I}_{\text{Chao}}$, its one- and two-sided coverage probabilities are relatively the most accurate. In scenario B, $\mathcal{I}_{\text{Chao}}$ is the winner in terms of interval score and lower-limit coverage accuracy, while its upper limit often produces undercoverage compared with that of $\mathcal{I}_{\text{LR}}$. Although $\mathcal{I}_{\text{Chao}}$ improves $\mathcal{I}_{\text{Wald}}$ by a large margin in scenario B, in scenario A with $p(\theta_0) = 0.8$ its two-sided intervals and upper-limits have even larger undercoverage and its lower-limits have even larger overcoverage than the latter. In scenarios C and D, both $\mathcal{I}_{\text{Wald}}$ and $\mathcal{I}_{\text{Chao}}$ have severe two-sided and lower-limit overcoverage and severe upper-limit undercoverage, while $\mathcal{I}_{\text{LR}}$ almost always has the most desirable performance. For example, at the 90% level, lower-limit overcoverage of $\mathcal{I}_{\text{Wald}}$ and $\mathcal{I}_{\text{Chao}}$ are at least 2.7% in scenario C and at least 3.4% in scenario D. In summary, $\mathcal{I}_{\text{Chao}}$ does not always improve $\mathcal{I}_{\text{Wald}}$; Although beaten by $\mathcal{I}_{\text{Chao}}$ in scenario B, $\mathcal{I}_{\text{LR}}$ has the best and most stable overall performance.

To get some insight into the above simulation results, we display in Figs. 1 and 2 the box-plots of the interval lengths of $\mathcal{I}_{\text{LR}}$, $\mathcal{I}_{\text{Wald}}$ and $\mathcal{I}_{\text{Chao}}$, and the QQ-plots of $R'(N_0)$ versus χ^2_1, $\sqrt{\tilde{N}}\{\log(\tilde{N}) - \log(N_0)\}/\tilde{\sigma}$ versus $N(0,1)$, and $C(N_0; \tilde{N})$ versus $N(0,1)$. In scenarios C, and D, our previous observation is that the coverage probabilities of $\mathcal{I}_{\text{LR}}$ are always almost identical to the nominal levels. This is probably because the finite-sample distribution of $R'(N_0)$ is very close to its limiting distribution, χ^2_1, as

Table 1 Simulated coverage probabilities (%) and interval scores of (in parentheses) $\mathcal{I}_{\mathrm{LR}}$, $\mathcal{I}_{\mathrm{Wald}}$ and $\mathcal{I}_{\mathrm{Chao}}$ in Scenarios A and B

	Scenario A		Scenario B	
	$p(\theta_0) = 0.5$	0.8	$\phi_0 = 0.8$	1.5
Two-sided				
Level = 90%				
$\mathcal{I}_{\mathrm{LR}}$	90.05 (97.74)	89.55 (29.70)	86.14 (60.61)	86.02 (44.91)
$\mathcal{I}_{\mathrm{Wald}}$	90.40 (92.55)	90.00 (29.50)	85.53 (55.68)	86.95 (40.33)
$\mathcal{I}_{\mathrm{Chao}}$	89.05 (97.60)	88.05 (30.88)	87.03 (53.97)	86.20 (41.60)
Level = 95%				
$\mathcal{I}_{\mathrm{LR}}$	94.90 (113.71)	94.60 (33.93)	92.21 (74.92)	92.90 (60.12)
$\mathcal{I}_{\mathrm{Wald}}$	95.85 (106.07)	94.95 (33.93)	88.63 (73.24)	89.40 (51.18)
$\mathcal{I}_{\mathrm{Chao}}$	94.20 (112.92)	93.30 (35.69)	93.74 (67.71)	92.75 (50.52)
Level = 99%				
$\mathcal{I}_{\mathrm{LR}}$	99.15 (151.30)	99.05 (41.19)	97.85 (138.56)	97.87 (103.68)
$\mathcal{I}_{\mathrm{Wald}}$	98.70 (138.28)	98.55 (43.78)	92.54 (139.37)	93.80 (90.26)
$\mathcal{I}_{\mathrm{Chao}}$	99.15 (145.76)	98.15 (47.21)	98.75 (107.95)	98.65 (74.26)
Lower limit				
Level = 90%				
$\mathcal{I}_{\mathrm{LR}}$	90.30	88.15	79.10	82.02
$\mathcal{I}_{\mathrm{Wald}}$	90.05	88.75	81.00	83.15
$\mathcal{I}_{\mathrm{Chao}}$	91.65	92.50	87.95	88.05
Level = 95%				
$\mathcal{I}_{\mathrm{LR}}$	95.15	94.20	87.85	88.41
$\mathcal{I}_{\mathrm{Wald}}$	94.15	93.80	85.50	86.95
$\mathcal{I}_{\mathrm{Chao}}$	96.20	96.75	94.05	94.35
Level = 99%				
$\mathcal{I}_{\mathrm{LR}}$	99.15	98.35	96.45	96.93
$\mathcal{I}_{\mathrm{Wald}}$	98.15	98.05	91.20	92.25
$\mathcal{I}_{\mathrm{Chao}}$	99.55	99.80	99.25	99.50
Upper limit				
Level = 90%				
$\mathcal{I}_{\mathrm{LR}}$	89.30	90.40	95.50	95.47
$\mathcal{I}_{\mathrm{Wald}}$	90.95	90.60	100.0	99.35
$\mathcal{I}_{\mathrm{Chao}}$	87.25	85.50	87.35	86.25
Level = 95%				
$\mathcal{I}_{\mathrm{LR}}$	94.90	95.35	97.20	97.93
$\mathcal{I}_{\mathrm{Wald}}$	96.25	96.20	100.0	100.0
$\mathcal{I}_{\mathrm{Chao}}$	92.85	91.30	92.85	91.85
Level = 99%				
$\mathcal{I}_{\mathrm{LR}}$	98.90	99.15	99.75	99.65
$\mathcal{I}_{\mathrm{Wald}}$	99.90	99.80	100.0	100.0
$\mathcal{I}_{\mathrm{Chao}}$	98.45	97.20	97.95	97.70

Table 2 Simulated coverage probabilities (%) and interval scores of $\mathcal{I}_{\mathrm{LR}}$, $\mathcal{I}_{\mathrm{Wald}}$ and $\mathcal{I}_{\mathrm{Chao}}$ in Scenarios C and D

	Scenario C		Scenario D	
	$K = 2$	4	$p(\theta_0) = 0.1$	0.25
Two-sided				
Level = 90%				
$\mathcal{I}_{\mathrm{LR}}$	90.15 (159.28)	90.35 (111.02)	90.90 (165.58)	90.25 (96.72)
$\mathcal{I}_{\mathrm{Wald}}$	92.90 (168.52)	91.40 (113.31)	93.60 (191.85)	91.65 (99.17)
$\mathcal{I}_{\mathrm{Chao}}$	91.10 (165.98)	90.70 (113.73)	92.15 (183.26)	90.45 (100.36)
Level = 95%				
$\mathcal{I}_{\mathrm{LR}}$	95.25 (183.02)	94.80 (127.33)	95.75 (190.81)	95.00 (111.54)
$\mathcal{I}_{\mathrm{Wald}}$	96.45 (199.65)	95.60 (131.73)	97.00 (236.47)	95.65 (115.31)
$\mathcal{I}_{\mathrm{Chao}}$	95.65 (195.68)	95.15 (133.21)	95.85 (219.99)	95.25 (117.47)
Level = 99%				
$\mathcal{I}_{\mathrm{LR}}$	98.95 (234.68)	98.90 (165.93)	99.25 (242.38)	98.65 (144.57)
$\mathcal{I}_{\mathrm{Wald}}$	99.50 (275.85)	99.25 (172.24)	99.25 (317.13)	98.90 (150.76)
$\mathcal{I}_{\mathrm{Chao}}$	99.35 (266.54)	99.05 (176.15)	98.85 (309.58)	98.60 (157.26)
Lower limit				
Level = 90%				
$\mathcal{I}_{\mathrm{LR}}$	89.95	90.20	91.90	91.25
$\mathcal{I}_{\mathrm{Wald}}$	94.15	92.70	96.40	93.40
$\mathcal{I}_{\mathrm{Chao}}$	93.30	92.70	95.55	94.10
Level = 95%				
$\mathcal{I}_{\mathrm{LR}}$	94.95	94.90	96.45	95.70
$\mathcal{I}_{\mathrm{Wald}}$	97.65	96.20	98.90	96.90
$\mathcal{I}_{\mathrm{Chao}}$	97.10	96.45	98.70	97.25
Level = 99%				
$\mathcal{I}_{\mathrm{LR}}$	99.20	99.20	99.10	98.85
$\mathcal{I}_{\mathrm{Wald}}$	99.70	99.25	99.95	99.15
$\mathcal{I}_{\mathrm{Chao}}$	99.70	99.50	99.85	99.60
Upper limit				
Level = 90%				
$\mathcal{I}_{\mathrm{LR}}$	90.05	90.85	88.30	89.45
$\mathcal{I}_{\mathrm{Wald}}$	90.10	89.95	89.30	89.20
$\mathcal{I}_{\mathrm{Chao}}$	88.75	88.80	87.85	87.95
Level = 95%				
$\mathcal{I}_{\mathrm{LR}}$	95.20	95.45	94.45	94.55
$\mathcal{I}_{\mathrm{Wald}}$	95.25	95.20	94.70	94.75
$\mathcal{I}_{\mathrm{Chao}}$	94.00	94.25	93.45	93.20
Level = 99%				
$\mathcal{I}_{\mathrm{LR}}$	99.15	98.85	99.00	98.95
$\mathcal{I}_{\mathrm{Wald}}$	99.25	99.15	98.60	98.70
$\mathcal{I}_{\mathrm{Chao}}$	98.85	98.35	98.25	98.10

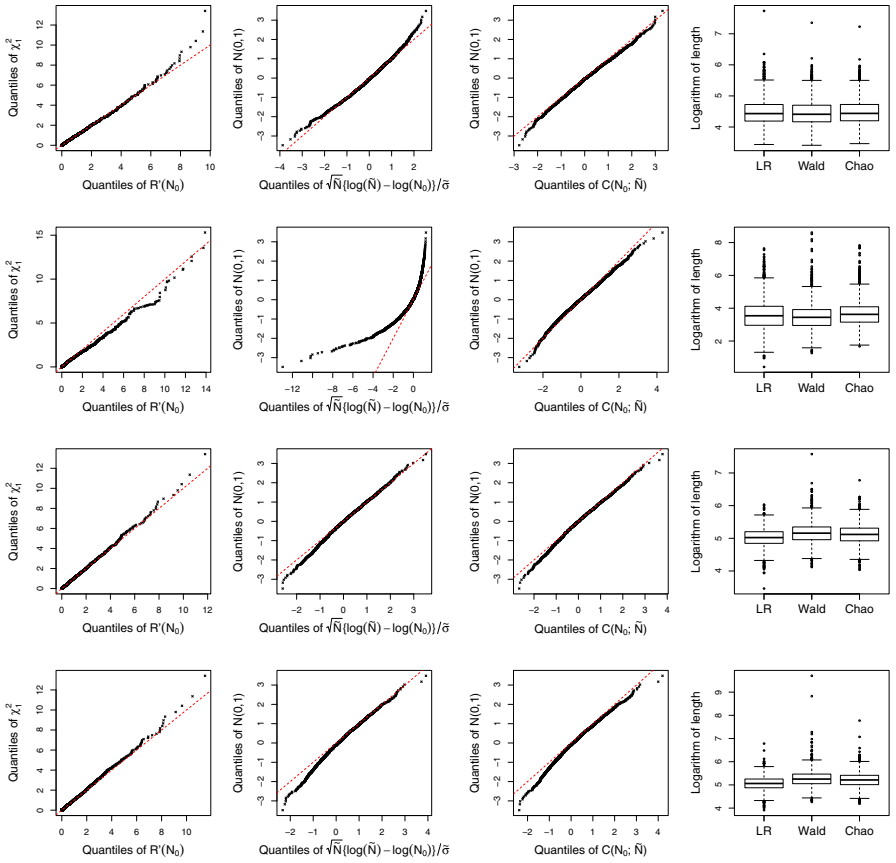


Fig. 1 QQ-plots and box-plots for scenario A with $p = 0.5$ (row 1), B with $\phi = 0.8$ (row 2), C with $K = 2$ (row 3), and D with $p = 0.1$ (row 4). Column 1: QQ-plots of $R'(N_0)$ versus χ^2_1; Column 2: QQ-plots of $\sqrt{\tilde{N}}\{\log(\tilde{N}) - \log(N_0)\}/\tilde{\sigma}$ versus $N(0, 1)$; Column 3: QQ-plots of $C(N_0; \tilde{N})$ versus $N(0, 1)$; Column 4: box-plots of the average lengths of $\mathcal{I}_{\mathrm{LR}}$, $\mathcal{I}_{\mathrm{Wald}}$ and $\mathcal{I}_{\mathrm{Chao}}$

we can see from Figs. 1 and 2. However $\mathcal{I}_{\mathrm{Wald}}$ and $\mathcal{I}_{\mathrm{Chao}}$ may have undercoverage or overcoverage. In terms of interval length, $\mathcal{I}_{\mathrm{LR}}$ has the shortest average lengths in all the two scenarios. In scenario A, $\mathcal{I}_{\mathrm{LR}}$, $\mathcal{I}_{\mathrm{Wald}}$ and $\mathcal{I}_{\mathrm{Chao}}$ have close coverage probabilities. This can be explained by the corresponding QQ-plots, in which both the distributions of $R'(N_0)$, $\sqrt{\tilde{N}}\{\log(\tilde{N}) - \log(N_0)\}/\tilde{\sigma}$ and $C(N_0; \tilde{N})$ are close to their limiting distributions. Meanwhile $\mathcal{I}_{\mathrm{LR}}$ and $\mathcal{I}_{\mathrm{Wald}}$ seem to have shorter length than $\mathcal{I}_{\mathrm{Chao}}$.

In Scenario B, $\mathcal{I}_{\mathrm{Chao}}$ outperforms $\mathcal{I}_{\mathrm{LR}}$ and $\mathcal{I}_{\mathrm{Wald}}$ although all intervals have severe undercoverage. This probably can be explained by their QQ-plots, in which the approximation of the limiting distribution to the finite-sample distribution of $C(N_0; \tilde{N})$ is better than those of $R'(N_0)$ and $\sqrt{\tilde{N}}\{\log(\tilde{N}) - \log(N_0)\}/\tilde{\sigma}$. It is worth mentioning that $\mathcal{I}_{\mathrm{LR}}$ always has very close two-sided coverage probabilities to those of $\mathcal{I}_{\mathrm{Chao}}$. Their severe undercoverage makes it unfair to compare their interval lengths. A possible

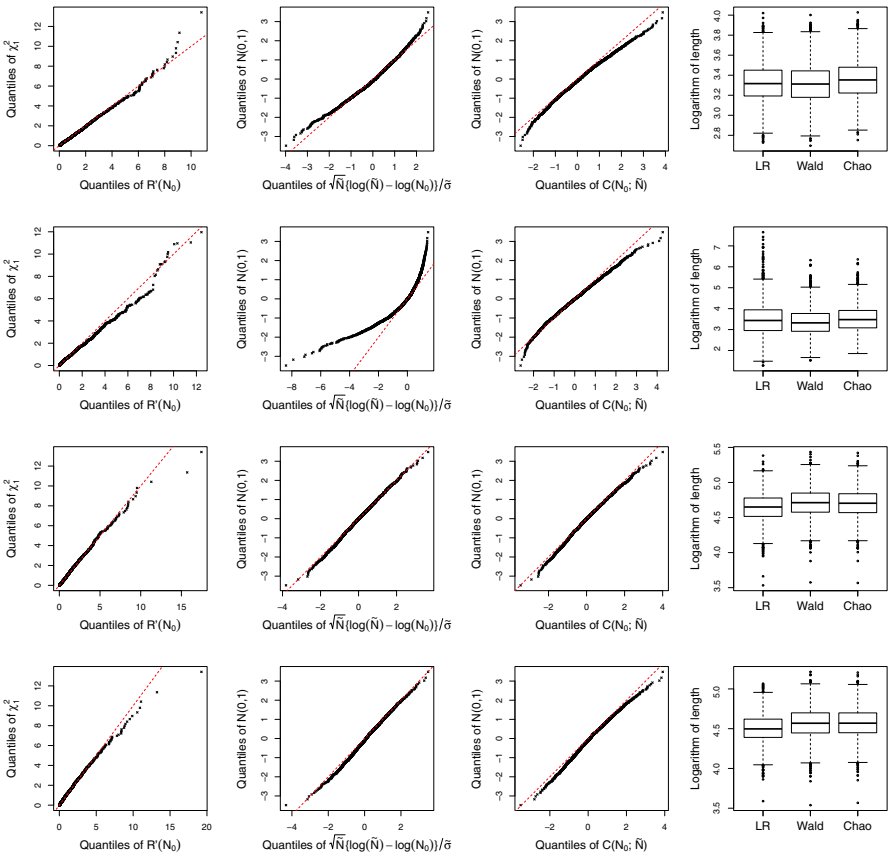


Fig. 2 QQ-plots and box-plots for scenario A with $p = 0.8$ (row 1), B with $\phi = 1.5$ (row 2), C with $K = 4$ (row 3), and D with $p = 0.25$ (row 4). Column 1: QQ-plots of $R'(N_0)$ versus χ^2_1; Column 2: QQ-plots of $\sqrt{\tilde{N}}\{\log(\tilde{N}) - \log(N_0)\}/\tilde{\sigma}$ versus $N(0, 1)$; Column 3: QQ-plots of $C(N_0; \tilde{N})$ versus $N(0, 1)$; Column 4: box-plots of the average lengths of $\mathcal{I}_{\text{LR}}$, $\mathcal{I}_{\text{Wald}}$ and $\mathcal{I}_{\text{Chao}}$

for these undesirable performance is that there are too many unknown parameters (the dimension of θ is 11) in the model but there are too few observations since the abundance is only 100.

In addition, the lower-tail quantiles of $N(0, 1)$ are much larger than those of $\sqrt{\tilde{N}}\{\log(\tilde{N}) - \log(N_0)\}/\tilde{\sigma}$ in scenario B, which explains the sever under-coverage of the lower limit of $\mathcal{I}_{\text{Wald}}$. The inverse occurs in scenarios C and D with lower capture probabilities. The upper-tail quantiles of $N(0, 1)$ are much larger than those of $\sqrt{\tilde{N}}\{\log(\tilde{N}) - \log(N_0)\}/\tilde{\sigma}$ in scenario B, but are very close to the latter in other scenarios. This probably explains why the upper limit of $\mathcal{I}_{\text{Wald}}$ has sever overcoverage in scenario B, but produces very accurate coverage probabilities in other scenarios.

Overall, the proposed likelihood ratio confidence interval usually (except in scenario B) has very accurate coverage probability with reasonable interval length and interval score, and exhibits more stable performance than $\mathcal{I}_{\text{Wald}}$ and $\mathcal{I}_{\text{Chao}}$ under varieties of

binomial detectability models, especially in scenarios C and D. There is still much room for improvement in its performance in scenario B.

4 Real applications

We illustrate the proposed interval estimation method by analyzing three data sets. The first two can be suitably modelled by capture–recapture models, while the last is obtained in a distance sampling experiment.

4.1 Possum data and Mouse data

We first analyze two capture–recapture data: the possum data, cf. Heinze et al. (2004); Huggins and Hwang (2007) and the mouse data, cf. Stoklosa et al. (2011). The possum data, collected over five consecutive nights in November 2003, records the capture histories of mountain Pygmy Possum (Burramys parvus) at mount Hotham in the snowfields of Victoria, Australia. The mouse data records the captures of the Harvest mouse (Micromys minutus) over 14 occasions at Wulin Recreation Area in Shei-Pa natinal Park, Taiwan, in the summer of 2008.

We assume both data follow M_0 models, which were employed in scenario A in our simulation study. Table 3 presents the maximum full likelihood estimates (MLE) and maximum conditional likelihood estimates (MCLE) of the abundances for the two populations, and the interval estimates $\mathcal{I}_{\mathrm{LR}}$, $\mathcal{I}_{\mathrm{Wald}}$, and $\mathcal{I}_{\mathrm{Chao}}$ at confidence levels 95% and 99%. The MLE and MCLE point estimates are quite close to each other for the possums data and the mouse data. This observation is consistent with the theoretical result in Lemma 1.

Table 3 Point and interval estimates for the possums data and the mouse data

Model	Data	n	MLE	MCLE	$\mathcal{I}_{\mathrm{LR}}$	$\mathcal{I}_{\mathrm{Wald}}$	$\mathcal{I}_{\mathrm{Chao}}$
M_0	Confidence level: 95%						
	Possum	43	47.3	48.1	[43.2, 54.7]	[43.2, 53.6]	[44.9, 56.3]
	Mouse	142	157.3	158.0	[149.0, 168.4]	[149.3, 167.1]	[151.2, 169.6]
	Confidence level: 99%						
	Possum				[43.0, 57.9]	[41.7, 55.5]	[44.4, 61.0]
	Mouse				[146.9, 172.6]	[146.7, 170.1]	[149.8, 174.8]
M_h	Confidence level: 95%						
	Possum	43	48.6	49.1	[44.2, 54.7]	[44.0, 54.7]	[45.6, 57.0]
	Mouse	142	175.3	175.8	[162.9, 190.1]	[162.6, 190.1]	[164.6, 192.6]
	Confidence level: 99%						
	Possum				[43.2, 57.1]	[42.6, 56.6]	[45.0, 61.3]
	Mouse				[159.5, 195.3]	[158.7, 194.9]	[162.0, 199.4]

The three interval estimates are also very close to each other. However, for the possum data, the lower limit, 41.7, of $\mathcal{I}_{\text{Wald}}$ at level 99% is smaller than the sample size $n = 43$, which is unreasonable. In comparison, the lower limits of $\mathcal{I}_{\text{LR}}$ and $\mathcal{I}_{\text{Chao}}$ in all the four cases are no less than the corresponding n. According to our simulation experience in scenario A, $\mathcal{I}_{\text{Chao}}$ tends to produce undercoverage under M_0. Hence we believe that $\mathcal{I}_{\text{LR}}$ is the most reliable two-side confidence interval for these two data.

Since both data include body mass, which can be regarded as a covariate, we can apply the Alho-Huggins model to them, together with the M_h model. Under these models, by the maximum conditional likelihood method, we find that the fitted capture probability of an individual given its covariate x is $g(x) = \frac{\exp(2.12-0.07x)}{1+\exp(2.12-0.07x)}$ for the Possum data and $g(x) = \frac{\exp(-4.08+0.27x)}{1+\exp(-4.08+0.27x)}$ for the Mouse data. We further assume that the covariates follows a log-normal distribution $LN(\mu, \sigma^2)$ and denote its density function by $f(x;\theta)$ with $\theta = (\mu, \sigma^2)$. These lead to a model for the marginal capture probability, $p(\theta) = \int_0^\infty f(x;\theta)\phi(x)dx$, and a model for the conditional distribution of the covariate, $k(x;\theta) = f(x;\theta)\phi(x)/p(\theta)$, where $\phi(x) = 1-\{1-g(x)\}^m$ and m is the capture times. The numerical results paralleling those under the M_0 model are also presented in Table 3. Our general observations are similar to those under the M_0 model, except that all numbers increase.

4.2 Hawaiian Amakihi data

We next apply the proposed estimation method to a distance sampling data and estimate the abundance of Hawaiian amakihi. The data were collected on the island of Hawaii from point transects in seven survey periods between July 1992 and April 1995, cf. Fancy et al. (1997) and Marques et al. (2007), and are available from the R package `Distance`.

The radii W (unit: meter) of the point transects in each period, the numbers n of observations, and the months in which the surveys were conducted are given in Table 4. Let x ($x \le W$) denote the distance between a generic amakihi and the observer. For illustration purpose, we take x as a covariate, and assume that the overall capture probability and the conditional density function of x are respectively

$$p(\theta) = \frac{1}{\theta W^2}\left(1-e^{-\theta W^2}\right) \quad \text{and} \quad k(x;\theta) = \frac{2\theta x e^{-\theta x^2}}{1-e^{-\theta W^2}}, \quad 0 \le x \le W.$$

For each survey period, we calculate the MLE and MCLE, and the likelihood ratio interval and Wald-type interval estimates at the confidence level 95%. We present in Table 4 the MLEs and MCLEs, and the interval estimates $\mathcal{I}_{\text{LR}}$, $\mathcal{I}_{\text{Wald}}$, and $\mathcal{I}_{\text{Chao}}$ at confidence levels 95%. Again the MLEs are very close to the corresponding MCLEs. The interval estimate is shifting to the right from $\mathcal{I}_{\text{LR}}$ to $\mathcal{I}_{\text{Wald}}$ to $\mathcal{I}_{\text{Chao}}$. If the assumed models are correct then $\mathcal{I}_{\text{LR}}$ is the most reliable because it has the best coverage accuracy, the shortest interval length and the smallest interval score among the three intervals in scenario D, which is a distance sampling model.

Table 4 Information and estimates for the Hawaiian amakihi data

No.	n	W	MLE	MCLE	$\mathcal{I}_{\mathrm{LR}}$	$\mathcal{I}_{\mathrm{Wald}}$	$\mathcal{I}_{\mathrm{Chao}}$
July 1992	183	250	2449.8	2450.9	[1997.8, 2987.0]	[2004.5, 2996.7]	[2009.2, 2999.4]
December 1992	158	110	666.0	667.3	[533.3, 826.8]	[536.0, 830.7]	[540.8, 835.5]
April 1993	274	130	1364.2	1365.4	[1155.3, 1605.5]	[1158.4, 1609.5]	[1163.0, 1614.0]
July 1993	210	100	1107.5	1108.6	[916.2, 1332.8]	[919.3, 1337.0]	[923.8, 1341.3]
January 1994	262	210	1956.3	1957.4	[1653.2, 2306.7]	[1657.2, 2312.0]	[1661.6, 2315.7]
April 1994	148	110	726.4	727.6	[578.7, 906.0]	[581.6, 910.2]	[586.2, 914.6]
April 1995	250	120	1032.8	1034.1	[865.8, 1227.5]	[868.6, 1231.2]	[873.5, 1236.1]

Acknowledgements We are grateful to the editor and two anonymous referees for their insightful and constructive comments which led to an improved presentation of this article. The research was supported by National Natural Science Foundation of China (Grant Nos. 11501354, 11771144, 11371142, and 11501208), Program of Shanghai Subject Chief Scientist (14XD1401600) and the 111 Project (B14019).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Appendix

A. Regularity conditions on $k(x;\theta)$

We assume that $k(x;\theta)$ satisfies the following regularity conditions, which are from §4.2.2 of Serfling (1980).

(R1) Let Θ be the parameter space of θ and θ_0 be its true value. Suppose Θ is an open set and θ_0 belongs to Θ.

(R2) For each $\theta \in \Theta$, the derivatives

$$\frac{\partial \log k(x;\theta)}{\partial\theta}, \quad \frac{\partial^2 \log k(x;\theta)}{\partial\theta^2}, \quad \frac{\partial^3 \log k(x;\theta)}{\partial\theta^3}$$

exist for all x.

(R3) There exist functions $g(x)$, $h(x)$ and $H(x)$ such that for θ in a neighborhood of θ_0,

$$\left|\frac{\partial \log k(x;\theta)}{\partial\theta}\right| \le g(x), \quad \left|\frac{\partial^2 \log k(x;\theta)}{\partial\theta^2}\right| \le h(x), \quad \left|\frac{\partial^3 \log k(x;\theta)}{\partial\theta^3}\right| \le H(x)$$

hold for all x and

$$\int g(x)dx < \infty, \quad \int h(x)dx < \infty, \quad \int H(x)k(x;\theta)dx < \infty.$$

(R4) For each $\theta \in \Theta$, $0 < \int \{\partial \log k(x;\theta)/\partial\theta\}^2 k(x;\theta)dx < \infty$.

B. Technical preparations

We make technical preparations for the proof of Theorem 1. For any positive real number x greater than n, define the digamma function $\psi_0(x) = d\log\{\Gamma(x)\}/dx$ and $S_1(x,n) = \psi_0(x+1) - \psi_0(x-n+1)$. For $a = 1, 2, \ldots$, we define the polygamma functions

$$\psi_a(x) = \frac{d^{a+1}\log\{\Gamma(x)\}}{dx^{a+1}} = \frac{d^a\psi_0(x)}{dx^a} = (-1)^{a+1}a!\sum_{k=0}^{\infty}\frac{1}{(x+k)^{a+1}}, \tag{6}$$

$$S_a(x,u) = \psi_{a-1}(x+1) - \psi_{a-1}(x-u+1) = (-1)^{a-1}(a-1)!\sum_{k=x-u+1}^{x} k^{-a}. \tag{7}$$

It is clear that $\psi_1(x) = d\psi_0(x)/dx$ and therefore $S_2(x,n) = dS_1(x,n)/dx$.

Since x^{-1} and x^{-2} are both monotone decreasing functions for $x > 0$, it follows from Eqs. (6) and (7) that

$$\log\{(N+1)/(N+1-n)\} < S_1(N,n) < \log\{N/(N-n)\},$$
$$-n/\{N(N-n)\} < S_2(N,n) < -n/\{(N+1)(N+1-n)\}.$$

Note that the number n of detected observations follows a binomial distribution $B(N_0, p(\theta_0))$. By the central limit theorem,

$$\sqrt{N_0}\left\{\frac{n}{N_0} - p(\theta_0)\right\} \xrightarrow{d} N\left(0,\ p(\theta_0)\{1-p(\theta_0)\}\right),$$

as $N_0 \to \infty$. Therefore, it follows that

$$\begin{aligned} S_1(N_0,n) &= \log\{N_0/(N_0-n)\} + O_p\left(N_0^{-1}\right) \\ &= -\log\{1-p(\theta_0)\} + \frac{(n/N_0) - p(\theta_0)}{1-p(\theta_0)} + O_p\left(N_0^{-1}\right), \\ S_2(N_0,n) &= -\frac{n}{N_0(N_0-n)} + O_p\left(N_0^{-2}\right) \\ &= -\frac{p(\theta_0)}{N_0\{1-p(\theta_0)\}} + O_p\left(N_0^{-3/2}\right). \end{aligned}$$

The following lemma from Hjort and Pollard (2011) can ease much of the technical burden in our proof of Theorem 1.

Lemma 2 *Assume that $\theta^\top = (\theta_1^\top, \theta_2^\top)$ where θ_1 and θ_2 are r- and s-dimensional vectors, respectively. Let $\theta_0^\top = (\theta_{10}^\top, \theta_{20}^\top)$ be its true value, and $\gamma = (\gamma_1^\top, \gamma_2^\top)^\top = \sqrt{n}(\theta - \theta_0)$ where n is the sample size. Suppose for $\theta = \theta_0 + O_p(n^{-1/2})$, it holds that*

$$H(\theta) = C_n + a_n^\top\gamma - \frac{1}{2}\gamma^\top A\gamma + \varepsilon_n(\theta)$$

where $a_n = O_p(1)$, A is a positive definite matrix, C_n does not depend on θ, and $\varepsilon_n(\theta) = o_p(1)$ for any fixed θ. According to $\theta = (\theta_1^\top, \theta_2^\top)^\top$, we partition A into $A = (A_{ij})_{1\le i,j\le 2}$, and partition $a_n^\top$ into $(a_{n1}^\top, a_{n2}^\top)$. As $n \to \infty$, if $a_n \xrightarrow{d} N(0,A)$, then

(a) *the maximizer* $\hat{\theta}$ *of* $H(\theta)$ *satisfies* $\sqrt{n}(\hat{\theta}-\theta_0) = A^{-1}a_n + o_p(1) \xrightarrow{d} N(0, A^{-1})$,
(b) $2\{\max_\theta H(\theta) - H(\theta_0)\} = a_n^\top A^{-1} a_n + o_p(1) \xrightarrow{d} \chi^2_{r+s}$, *and*
(c) $2\{\max_\theta H(\theta) - \max_{\theta_2} H(\theta_{10}, \theta_2)\} = a_n^\top A^{-1} a_n - a_{n2}^\top A_{22}^{-1} a_{n2} + o_p(1) \xrightarrow{d} \chi^2_r$.

C. Proof of Theorem 1

Using a similar argument to that in the proofs of Lemma 1 and Theorem 1 of Qin and Lawless (1994), we have $\hat{N} = N_0 + O_p(N_0^{1/2})$ and $\hat{\theta} - \theta_0 = O_p(N_0^{-1/2})$. Since the results in Theorem 1 are about the properties of $(\hat{N}, \hat{\theta})$, our proof begins by studying the behavior of $\ell(N, \theta)$ for (N, θ) such that $((N - N_0)/N_0, \theta - \theta_0) = O_p(N_0^{-1/2})$.

Let $\alpha = (\alpha_1, \alpha_2^\top)^\top$ with $\alpha_1 = N_0^{-1/2}(N - N_0)$ and $\alpha_2 = N_0^{1/2}(\theta - \theta_0)$. Define $H(\alpha) = \ell(N_0 + N_0^{1/2}\alpha_1, \theta_0 + N_0^{-1/2}\alpha_2)$. The likelihood ratio function of (N, θ) can be expressed as

$$R(N, \theta) = 2\{H(\alpha) - H(0)\}.$$

By the second-order Taylor expansion, we have

$$H(\alpha) = H(0) + \alpha^\top u + \frac{1}{2}\alpha^\top V \alpha + o_p(1),$$

where $u \equiv (u_1, u_2^\top)^\top = \partial H(0)/\partial\alpha$ and V is the leading term of $\partial^2 H(0)/(\partial\alpha\partial\alpha^\top)$.

To proceed, we need the expressions of u and V. It can be seen that

$$\begin{aligned}
\frac{\partial \ell(N, \theta)}{\partial N} &= S_1(N, n) + \log\{1 - p(\theta)\}, \\
\frac{\partial \ell(N, \theta)}{\partial \theta} &= \frac{n - Np(\theta)}{p(\theta)\{1 - p(\theta)\}} \frac{dp(\theta)}{d\theta} + \sum_{i=1}^{n} \frac{\partial \log\{k(x_i; \theta)\}}{\partial \theta}.
\end{aligned}$$

According to the properties of the digamma functions, we further have

$$\begin{aligned}
u_1 = \frac{\partial H(0)}{\partial \alpha_1} &= N_0^{1/2} \frac{\partial \ell(N_0, \theta_0)}{\partial N} \\
&= N_0^{1/2} \left[S_1(N_0, n) + \log\{1 - p(\theta_0)\} \right] \\
&= N_0^{1/2} \frac{(n/N_0) - p(\theta_0)}{1 - p(\theta_0)} + O_p\left(N_0^{-1/2}\right).
\end{aligned}$$

and

$$\begin{aligned}
u_2 = \frac{\partial H(0)}{\partial \alpha_2} &= N_0^{-1/2} \frac{\partial \ell(N_0, \theta_0)}{\partial \theta} \\
&= N_0^{-1/2} \left[\frac{n - N_0 p(\theta_0)}{1 - p(\theta_0)} \frac{d \log\{p(\theta_0)\}}{d\theta} + \sum_{i=1}^{n} \frac{d \log\{k(x_i, \theta_0)\}}{d\theta} \right]
\end{aligned}$$

$$= N_0^{1/2} \frac{(n/N_0) - p(\theta_0)}{1 - p(\theta_0)} p_1(\theta_0) + \{p(\theta_0)\}^{1/2} n^{-1/2} \sum_{i=1}^{n} \frac{d \log\{k(x_i, \theta_0)\}}{d\theta} + O_p\left(N_0^{-1/2}\right).$$

By the central limit theorem, it can be shown that $u \xrightarrow{d} N(0, \Sigma)$.

Write $V = (V_{ij})_{1 \le i,j \le 2}$. It can be seen that V_{11} is the leading term of

$$\frac{\partial H(0)}{\partial \alpha_1^2} = N_0 \frac{\partial \ell(N_0, \theta_0)}{\partial N^2} = N_0 S_2(N_0, n) = -\frac{p(\theta_0)}{1 - p(\theta_0)} + O_p\left(N_0^{-1/2}\right),$$

where we have used an approximate of $S_2(N_0, n)$. This implies that

$$V_{11} = -\frac{p(\theta_0)}{1 - p(\theta_0)}.$$

With tedious algebra, we similarly have

$$V_{21} = -\frac{p(\theta_0)}{1 - p(\theta_0)} p_1(\theta_0), \quad V_{22} = -\frac{p(\theta_0)}{1 - p(\theta_0)} p_1(\theta_0)\{p_1(\theta_0)\}^{\top} - p(\theta_0) I(\theta_0).$$

Since $u \xrightarrow{d} N(0, \Sigma)$ and $V = -\Sigma$, Theorem 1 is proved by applying Lemma 2. □

References

Alho JM (1990) Logistic regression in capture–recapture models. Biometrics 46:623–635

Barnard J, Emam K, Zubrow D (2003) Using capture–recapture models for the reinspection decision. Softw Qual Prof 5:11–20

Borchers DL, Zucchini W, Fewster RM (1998) Mark-recapture models for line transect surveys. Biometrics 54:1207–1220

Borchers DL, Buckland ST, Zucchini W (2002) Estimating animal abundance: closed population. Springer, London

Borchers DL, Stevenson BC, Kidney D, Thomas L, Marques TA (2015) A unifying model for capture-recapture and distance sampling surveys of wildlife populations. J Am Stat Assoc 110:195–204

Buckland ST, Anderson DR, Burnham KP, Laake JL, Borchers DL, Thomas L (2001) Introduction to distance sampling. Oxford University Press, Oxford

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

Chao A, Chu W, Hsu CH (2000) Capture–recapture when time and behavioral response affect capture probabilities. Biometrics 56:427–433

Chao A, Tsay PK, Lin SH, Shau WY, Chao DY (2001) The applications of capture–recapture models to epidemiological data. Stat Med 20:3123–3157

Chen SX, Lloyd CJ (2002) Estimation of population size from biased samples using non-parametric binary regression. Stat. Sin. 12:505–518

Cormack RM (1992) Interval estimation for mark-recapture studies of closed population. Biometrics 48:567–576

Evans MA, Bonett DG (1994) Bias reduction for multiple recapture estimators of closed population size. Biometrics 50:388–395

Evans MA, Kim H, O'Bren TE (1996) An application of profile-likelihood based confidence interval to capture–recapture estimators. J Agric Biol Environ Stat 1(1):131–140

Fancy SG, Snetsinger TJ, Jacobi JD (1997) Translocation of the Palila, an endangered Hawaiian honeycreeper. Pac Conserv Biol 3:39–46
Fewster RM, Jupp PE (2009) Inference on population size in binomial detectability models. Biometrika 96:805–820
Gneiting T, Raftery AE (2007) Strictly proper scoring rules, prediction, and estimation. J Am Stat Assoc 102:359–378
Heinze D, Broome L, Mansergh I (2004) A review of the ecology and conservation of the mountain pygmy-possum Burramys parvus. In: Goldingay RL, Jackson SM (eds) The biology of Australian Possums and Gliders. Baulkham Hills, Surrey Beatty & Sons, pp 254–267
Hjort NL, Pollard D (2011) Asymptotics for minimisers of convex processes. arXiv:1107.3806v1
Hogan H (2000) Accuracy and coverage evaluation 2000: decomposition of dual system estimate components. U.S. Census Bureau, Washington
Huggins RM (1989) On the statistical analysis of capture experiments. Biometrika 76:133–140
Huggins R, Hwang WH (2007) Non-parametric estimation of population size from capture–recapture data when the capture probability depends on a covariate. J R Stat Soc Ser C 56:429–443
Liu Y, Li P, Qin J (2017) Maximum empirical likelihood estimation for abundance in a closed population from capture–recapture data. Biometrika 104(3):527–543
Marques FFC, Buckland ST (2004) Covariate models for the detection function. In: Buckland ST, Anderson DR, Burnham KP, Laake JL, Borchers DL, Thomas L (eds) Advanced distance sampling: estimating abundance of biological populations. Oxford University Press, Oxford
Marques TA, Thomas L, Fancy SG, Buckland ST (2007) Improving estimates of bird density using multiple-covariate distance sampling. Auk 124(4):1229–1243
Otis DL, Burnham KP, White GC, Anderson DR (1978) Statistical inference from capture data on closed animal populations. Wildl Monogr 62:1–135
Pollock KH (2000) Capture–recapture models. J Am Stat Assoc 95:293–296
Qin J, Lawless J (1994) Empirical likelihood and general estimating equations. Ann Stat 22(1):300–325
Sanathanan L (1972) Estimating the size of a multinomial population. Ann Math Stat 43:142–152
Serfling RJ (1980) Approximation theorem of mathematical statistics. Wiley, New York
Stoklosa J, Hwang WH, Wu SH, Huggins R (2011) Heterogeneous capture–recapture models with covariates: a partial likelihood approach for closed populations. Biometrics 67:1659–1665