Biometrika (2017), **xx**, x, *pp*. 1–31

Printed in Great Britain

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

BY YUKUN LIU

School of Statistics, East China Normal University, Shanghai 200241, China

ykliu@sfs.ecnu.edu.cn

PENGFEI LI

Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada

pengfei.li@uwaterloo.ca

AND JING QIN

National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, U.S.A.

jingqin@niaid.nih.gov

This is a supplementary document to the corresponding paper submitted to *Biometrika*. §1 reviews the results in the main paper. §2 presents some preliminary preparation. §3 contains the proofs of Theorems 1-2, Corollaries 1-2, the consistency of $\hat\sigma^2$ and $\hat\sigma_s^2$, the semiparametric efficiency of $\hat N$, and the consistency of $\hat f_w(x)$. §4 discusses the numerical implementation of the empirical-likelihood-based methods. §5 provides some additional simulation results. §6 proposes a bootstrap procedure to improve the performance of the empirical-likelihood-ratio-based confidence interval.

1. MAIN RESULTS IN THE MAIN PAPER

1·1. *General case*

Recall that we model the probability of capture on occasion j ($j = 1, \ldots, k$) by the logistic regression model $g_j(x) = g(x, \beta_j)$, where

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

We show that the profile empirical log-likelihood of (N, β, α) is, up to a constant not dependent on the unknown parameters,

$$\ell(N, \beta, \alpha) = \log\left\{\frac{\Gamma(N+1)}{\Gamma(N-n+1)}\right\} + (N-n)\log\alpha - \sum_{i=1}^{n}\log[1 + \lambda\{\phi(x_i, \beta) - \alpha\}]$$
$$+ \sum_{i=1}^{n}\sum_{j=1}^{k}\left[d_{ij}\log g(x_i, \beta_j) + (1-d_{ij})\log\{1 - g(x_i, \beta_j)\}\right], \tag{2}$$

where λ is the solution of

$$\sum_{i=1}^{n} \frac{\phi(x_i,\beta)-\alpha}{1+\lambda\{\phi(x_i,\beta)-\alpha\}} = 0 \tag{3}$$

and $\phi(x,\beta) = \prod_{j=1}^{k}\{1-g(x,\beta_j)\}$.

The maximum empirical likelihood estimators $(\hat{N},\hat{\beta},\hat{\alpha})$ of (N,β,α) are defined to be

$$(\hat{N},\hat{\beta},\hat{\alpha}) = \arg\max_{N,\beta,\alpha} \ell(N,\beta,\alpha). \tag{4}$$

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

$$R(N,\beta,\alpha) = 2\{\ell(\hat{N},\hat{\beta},\hat{\alpha}) - \ell(N,\beta,\alpha)\}, \tag{5}$$
$$R'(N) = 2\{\ell(\hat{N},\hat{\beta},\hat{\alpha}) - \ell(N,\hat{\beta}_N,\hat{\alpha}_N)\}, \tag{6}$$

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

Let N_0, $\beta_0 = (\beta_{10}^{\mathrm{T}},\ldots,\beta_{k0}^{\mathrm{T}})^{\mathrm{T}}$, and α_0 be the true values of N, β, and α, respectively. Denote

$$G_1(x) = \left\{g(x,\beta_{10}),\ldots,g(x,\beta_{k0})\right\}^{\mathrm{T}}, \quad G_2(x) = \mathrm{diag}\{G_1(x)\},\ \phi_* = E\left[\{1-\phi(X,\beta_0)\}^{-1}\right].$$

We use $\otimes$ to denote the Kronecker product operator. Define

$$V = \begin{pmatrix} V_{11} & 0 & V_{13} & 0 \\ 0 & V_{22} & V_{23} & V_{24} \\ V_{31} & V_{32} & V_{33} & V_{34} \\ 0 & V_{42} & V_{43} & V_{44} \end{pmatrix}, \tag{7}$$

$$W \equiv (W_{ij})_{1\le i,j\le 3} = \begin{pmatrix} -V_{11} & 0 & -V_{13} \\ 0 & -V_{22}+V_{24}V_{44}^{-1}V_{42} & -V_{23}+V_{24}V_{44}^{-1}V_{43} \\ -V_{31} & -V_{32}+V_{34}V_{44}^{-1}V_{42} & -V_{33}+V_{34}V_{44}^{-1}V_{43} \end{pmatrix}, \tag{8}$$

where

$$V_{11} = 1-\alpha_0^{-1},\ V_{13} = \alpha_0^{-1},$$
$$V_{22} = E\left[\left\{\frac{\phi(X,\beta_0)}{1-\phi(X,\beta_0)}G_1(X)G_1^{\mathrm{T}}(X) + G_2^2(X) - G_2(X)\right\} \otimes \{q(X)q(X)^{\mathrm{T}}\}\right],$$
$$V_{23} = V_{32}^{\mathrm{T}} = E\left\{\frac{\phi(X,\beta_0)}{1-\phi(X,\beta_0)}G_1(X)\otimes q(X)\right\},\ V_{24} = V_{42}^{\mathrm{T}} = (1-\alpha_0)^2 V_{23},$$
$$V_{33} = \phi_* - \alpha_0^{-1},\ V_{34} = V_{43} = (1-\alpha_0)^2\phi_*,\ V_{44} = (1-\alpha_0)^4\phi_* - (1-\alpha_0)^3.$$

With the above preparation, we have the following theorems.

THEOREM 1. *Assume that the support of X is compact, the capture probability function is $g_j(x) = g(x,\beta_j)$ as defined in* (1) *and the vector-valued function $q(x)$ is b-variate with linearly independent components. Let (N_0,β_0,α_0) be the true value of (N,β,α) with $\alpha_0 \in (0,1)$. If W defined in* (8) *is nonsingular, then as N_0 goes to infinity, we have*

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

Denote by $\ell_c(\beta) = \log L_c(\beta)$ the conditional log-likelihood given the observed data, where $L_c(\beta)$ defined in (3) in the main paper is the conditional likelihood. The maximum conditional likelihood estimator of N is defined as

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

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

THEOREM 2. *Under the assumptions in Theorem* 1, *as N_0 goes to infinity, we have*

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

Based on the form of σ^2 in Theorem 2, an estimator of σ^2 can be constructed as follows:

$$\hat{\sigma}^2 = \hat{\phi}_* - 1 - \hat{V}_{32}\hat{V}_{22}^{-1}\hat{V}_{23}, \tag{9}$$

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

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

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

1·2. *Special case*

When the β_j's are all equal, $\phi(x, \beta)$ reduces to $\phi_s(x, \beta_s) = \{1 - g(x, \beta_s)\}^k$, where β_s denotes the common value of the β_j's. In this situation, the profile empirical log-likelihood $\ell_s(N, \beta_s, \alpha)$ can be directly obtained from the profile empirical log-likelihood in (2):

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

where λ is the solution to

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

With the profile empirical log-likelihood $\ell_s(N, \beta_s, \alpha)$, we define the maximum empirical likelihood estimators $(\hat{N}_s, \hat{\beta}_s, \hat{\alpha}_s)$ of (N, β_s, α), the empirical likelihood ratio $R_s(N, \beta_s, \alpha)$ for (N, β_s, α) and the empirical likelihood ratio $R_s'(N)$ for N similarly to the definitions of $(\hat{N}, \hat{\beta}, \hat{\alpha})$, $R(N, \beta, \alpha)$, and $R'(N)$ in (4), (5), and (6). To present the asymptotics, we define a new W matrix, namely W_s, which is W with ϕ_*, V_{23}, V_{24}, and V_{22} in (8) replaced by $\phi_{s*} = E[\{1 - \phi_s(X, \beta_{s0})\}^{-1}]$ and

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

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

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

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

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

Given the observations, the conditional log-likelihood is

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

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

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

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

COROLLARY 2. *Under the assumptions in Corollary 1, as N_0 goes to infinity, we have*

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

Similarly to $\hat{\sigma}^2$ in (9), a consistent estimator of σ_s^2 can be constructed as

$$\hat{\sigma}_s^2 = \hat{\phi}_{s*} - 1 - \hat{V}_{32s}\hat{V}_{22s}^{-1}\hat{V}_{32s}^{\mathrm{T}}, \tag{10}$$

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

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

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

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

2. PREPARATION

2·1. *Reexpression*

It can be verified that

$$\ell(N, \beta, \alpha) = h(N, \beta, \alpha, \lambda_{N,\beta,\alpha}),$$

where

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

and $\lambda_{N,\beta,\alpha}$ is the solution to $\partial h/\partial\lambda = 0$.

Let $\hat\lambda$ be the solution to (3) with $(\hat\beta,\hat\alpha)$ in place of (β,α). We first discuss some asymptotic properties of $\hat\lambda$. It can be verified that $(\hat N,\hat\beta,\hat\alpha,\hat\lambda)$ satisfy

$$\frac{\partial h(N,\beta,\alpha,\lambda)}{\partial N}=0,\quad \frac{\partial h(N,\beta,\alpha,\lambda)}{\partial \beta}=0,\quad \frac{\partial h(N,\beta,\alpha,\lambda)}{\partial \alpha}=0,\quad \frac{\partial h(N,\beta,\alpha,\lambda)}{\partial \lambda}=0.$$

Note that

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

together imply that $(\hat N,\hat\beta,\hat\alpha,\hat\lambda)$ satisfy

$$\lambda = -\frac{1-n/N}{(n/N)\alpha}. \tag{11}$$

By the fact that $n\sim B(N_0,1-\alpha_0)$ and the law of large numbers, the right-hand side of (11) at the true values of (N,β,α) converges to a constant (denoted by λ_0) in probability. That is,

$$-\frac{1-n/N}{(n/N)\alpha} \to \lambda_0 = -1/(1-\alpha_0)$$

in probability. When $(\hat N,\hat\beta,\hat\alpha)$ is consistent, we can further verify that

$$\hat\lambda = -\frac{1-n/\hat N}{(n/\hat N)\hat\alpha} \to \lambda_0$$

in probability.

Next, we define more notation. Let

$$\gamma^{\mathrm T} = (\gamma_1,\gamma_2^{\mathrm T},\gamma_3,\gamma_4) = N_0^{1/2}\{(N/N_0)-1,(\beta-\beta_0)^{\mathrm T},\alpha-\alpha_0,\lambda-\lambda_0\},$$

and define

$$\hat\gamma^{\mathrm T} = (\hat\gamma_1,\hat\gamma_2^{\mathrm T},\hat\gamma_3,\hat\gamma_4) = N_0^{1/2}\{(\hat N/N_0)-1,(\hat\beta-\beta_0)^{\mathrm T},\hat\alpha-\alpha_0,\hat\lambda-\lambda_0\}.$$

Define

$$H(\gamma) = h(N,\beta,\alpha,\lambda) = h(N_0+N_0^{1/2}\gamma_1,\ \beta_0+N_0^{-1/2}\gamma_2,\ \alpha_0+N_0^{-1/2}\gamma_3,\ \lambda_0+N_0^{-1/2}\gamma_4).$$

It can be verified that $\hat\gamma$ is the solution to $\partial H(\gamma)/\partial\gamma = 0$.

To investigate the asymptotic properties of $(\hat N,\hat\beta,\hat\alpha,\hat\lambda)$, we need their approximations, which can be obtained via the second-order Taylor expansion of $H(\gamma)$ around $\gamma=0$. In this subsection, we derive the forms of $\partial H(0)/\partial\gamma$ and $\partial^2 H(0)/(\partial\gamma\partial\gamma^{\mathrm T})$ and study their properties.

2·2. *First and second derivatives of $H(\gamma)$ at $\gamma = 0$*

Recall that $G_1(x) = \{g(x, \beta_{10}), \ldots, g(x, \beta_{k0})\}^{\mathrm{T}}$. After some calculus, we have

$$\begin{aligned}
\partial H(0)/\partial\gamma_1 &= N_0^{1/2}\{S_1(N_0, n) + \log\alpha_0\},\\
\partial H(0)/\partial\gamma_2 &= N_0^{-1/2}\sum_{i=1}^{n}\left\{d_i - \frac{G_1(x_i)}{1-\phi(x_i,\beta_0)}\right\}\otimes q(x_i),\\
\partial H(0)/\partial\gamma_3 &= N_0^{-1/2}\left[\frac{N_0}{\alpha_0} - \sum_{i=1}^{n}\left\{\frac{1}{1-\phi(x_i,\beta_0)} + \frac{1}{\alpha_0}\right\}\right],\\
\partial H(0)/\partial\gamma_4 &= -(1-\alpha_0)N_0^{-1/2}\sum_{i=1}^{n}\frac{\phi(x_i,\beta_0)-\alpha_0}{1-\phi(x_i,\beta_0)}.
\end{aligned}$$

Here

$$S_c(N, n) = \frac{d^c \log\{\Gamma(N)\}}{dN^c} - \frac{d^c \log\{\Gamma(N-n+1)\}}{dN^c}$$

for nonnegative integer c. Using the properties of the polygamma functions, we have

$$S_c(N, n) = (-1)^{c-1}(c-1)!\sum_{k=N-n+1}^{N} k^{-c}; \tag{12}$$

see for example Murty & Saradha (2009).

Next we simplify $\partial H(0)/\partial N$ using (12). Since x^{-1} is a monotone decreasing function, (12) implies that

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

Since n follows $B(N_0, 1-\alpha_0)$, by the central limit theorem we have $n/N_0 = 1-\alpha_0 + O_p(N_0^{-1/2})$ and further

$$S_1(N_0, n) = \log\left(\frac{N_0}{N_0 - n}\right) + O_p(N_0^{-1}) = -\log\alpha_0 + \frac{(n/N_0) - 1 + \alpha_0}{\alpha_0} + O_p(N_0^{-1}).$$

Hence,

$$\partial H(0)/\partial\gamma_1 = N_0^{1/2}\{S_1(N_0, n) + \log\alpha_0\} = N_0^{1/2}\left\{\frac{(n/N_0) - (1-\alpha_0)}{\alpha_0}\right\} + O_p(N_0^{-1/2}).$$

Let

$$u_{n1} = N_0^{1/2}\left\{\frac{n/N_0 - (1-\alpha_0)}{\alpha_0}\right\},\ u_{n2} = \frac{\partial H(0)}{\partial\gamma_2},\ u_{n3} = \frac{\partial H(0)}{\partial\gamma_3},\ u_{n4} = \frac{\partial H(0)}{\partial\gamma_4}, \tag{13}$$

and $u_n = (u_{n1}, u_{n2}^{\mathrm{T}}, u_{n3}, u_{n4})^{\mathrm{T}}$. Then

$$\frac{\partial H(0)}{\partial\gamma} = u_n + O_p(N_0^{-1/2}).$$

Next we calculate the second derivatives of $H(\gamma)$ at $\gamma = 0$. Recall that $G_2(x) = \mathrm{diag}\{G_1(x)\}$. After some calculation, it can be verified that

$$\frac{\partial^2 H(0)}{\partial\gamma\partial\gamma^{\mathrm{T}}} = \begin{pmatrix} \frac{\partial^2 H(0)}{\partial\gamma_1^2} & 0 & \frac{\partial^2 H(0)}{\partial\gamma_1\gamma_3} & 0 \\ 0 & \frac{\partial^2 H(0)}{\partial\gamma_2\partial\gamma_2^{\mathrm{T}}} & \frac{\partial^2 H(0)}{\partial\gamma_2\partial\gamma_3} & \frac{\partial^2 H(0)}{\partial\gamma_2\partial\gamma_4} \\ \frac{\partial^2 H(0)}{\partial\gamma_3\partial\gamma_1} & \frac{\partial^2 H(0)}{\partial\gamma_3\partial\gamma_2^{\mathrm{T}}} & \frac{\partial^2 H(0)}{\partial\gamma_3^2} & \frac{\partial^2 H(0)}{\partial\gamma_3\partial\gamma_4} \\ 0 & \frac{\partial^2 H(0)}{\partial\gamma_4\partial\gamma_2^{\mathrm{T}}} & \frac{\partial^2 H(0)}{\partial\gamma_4\partial\gamma_3} & \frac{\partial^2 H(0)}{\partial\gamma_4^2} \end{pmatrix}, \tag{14}$$

with

$$\frac{\partial^2 H(0)}{\partial\gamma_1^2} = N_0 S_2(N_0, n), \ \frac{\partial^2 H(0)}{\partial\gamma_1\gamma_3} = \frac{\partial^2 H(0)}{\partial\gamma_3\gamma_1} = \frac{1}{\alpha_0},$$

$$\frac{\partial^2 H(0)}{\partial\gamma_2\partial\gamma_2^{\mathrm{T}}} = \frac{1}{N_0}\sum_{i=1}^n \frac{G_1(x_i)G_1(x_i)^{\mathrm{T}}\phi(x_i,\beta_0) - \{1-\phi(x_i,\beta_0)\}\{G_2(x_i) - G_2^2(x_i)\}}{\{1-\phi(x_i,\beta_0)\}^2} \otimes \{q(x_i)q(x_i)^{\mathrm{T}}\},$$

$$\frac{\partial^2 H(0)}{\partial\gamma_2\partial\gamma_3} = \left\{\frac{\partial^2 H(0)}{\partial\gamma_3\partial\gamma_2^{\mathrm{T}}}\right\}^{\mathrm{T}} = \frac{1}{N_0}\sum_{i=1}^n \frac{\phi(x_i,\beta_0)}{\{1-\phi(x_i,\beta_0)\}^2} G_1(x_i)\otimes q(x_i),$$

$$\frac{\partial^2 H(0)}{\partial\gamma_2\partial\gamma_4} = \left\{\frac{\partial^2 H(0)}{\partial\gamma_4\partial\gamma_2^{\mathrm{T}}}\right\}^{\mathrm{T}} = (1-\alpha_0)^2 \frac{\partial^2 H(0)}{\partial\gamma_2\partial\gamma_3},$$

$$\frac{\partial^2 H(0)}{\partial\gamma_3^2} = \frac{1}{N_0}\sum_{i=1}^n \frac{1}{\{1-\phi(x_i,\beta_0)\}^2} - \frac{1-(n/N_0)}{\alpha_0^2},$$

$$\frac{\partial^2 H(0)}{\partial\gamma_3\partial\gamma_4} = \frac{\partial^2 H(0)}{\partial\gamma_4\partial\gamma_3} = \frac{1}{N_0}\sum_{i=1}^n \frac{(1-\alpha_0)^2}{\{1-\phi(x_i,\beta_0)\}^2},$$

$$\frac{\partial^2 H(0)}{\partial\gamma_4^2} = \frac{1}{N_0}\sum_{i=1}^n \frac{(1-\alpha_0)^2\{\phi(x_i,\beta_0)-\alpha_0\}^2}{\{1-\phi(x_i,\beta_0)\}^2}.$$

2·3. *Some useful technical lemmas*

Recall that $\partial H(0)/\partial\gamma = u_n + O_p(N_0^{-1/2})$. In the proof of Theorem 1, we need the limit of $\partial^2 H(0)/(\partial\gamma\partial\gamma^{\mathrm{T}})$ and the expectation and variance of u_n defined in (13). The following lemmas ease much of the calculation burden in our proofs.

LEMMA 1. *Suppose $r(x)$ is a given nonzero function of x and $X \sim F(x)$. Then*

(a) if $E\left[r(X)\{1-\phi(X,\beta_0)\}\right] < \infty$, we have

$$E\left\{\frac{1}{N_0}\sum_{i=1}^n r(x_i)\right\} = E\left[r(X)\{1-\phi(X,\beta_0)\}\right]; \tag{15}$$

(b) if $E\left[r^2(X)\{1-\phi(X,\beta_0)\}\right] < \infty$, we have

$$\frac{1}{N_0}\sum_{i=1}^n r(x_i) - E\left[r(X)\{1-\phi(X,\beta_0)\}\right] = O_p(N_0^{-1/2}); \tag{16}$$

(c) if $E\left\{g(X,\beta_{j0})r(X)\right\}<\infty$, we have

$$E\left\{\frac{1}{N_0}\sum_{i=1}^{n} d_{ij}r(x_i)\right\}=E\left\{g(X,\beta_{j0})r(X)\right\}. \tag{17}$$

For (a), we define N_0 indicator variables $I_1,\ldots,I_{N_0}$ for the N_0 individuals in the population such that $I_i=1$ if the ith individual has been captured at least once and 0 otherwise, $i=1,\ldots,N_0$. Then

$$\frac{1}{N_0}\sum_{i=1}^{n} r(x_i)=\frac{1}{N_0}\sum_{i=1}^{N_0} r(X_i)I_i,$$

which is the summation of independent and identically distributed random variables. Hence, (15) follows from the fact that

$$E\{r(X_i)I_i\}=E[E\{r(X_i)I_i\mid X_i\}]=E\{r(X_i)E(I_i\mid X_i)\}=E[r(X)\{1-\phi(X;\beta_0)\}],$$

where we use $E(I_i\mid X_i)=\mathrm{pr}(I_i=1\mid X_i)=\phi(X_i,\beta_0)$ in the last equation.

For (b), we first write

$$\frac{1}{N_0}\sum_{i=1}^{n} r(x_i)-E[r(X)\{1-\phi(X;\beta_0)\}]=\frac{1}{N_0}\sum_{i=1}^{N_0}[r(X_i)I_i-E\{r(X_i)I_i\}].$$

Because $E\left[r^2(X)\left\{1-\phi(X,\beta_0)\right\}\right]<\infty$ and $r(x)$ is nonzero, by the central limit theorem we have

$$N_0^{1/2}\left(\frac{1}{N_0}\sum_{i=1}^{n} r(x_i)-E[r(X)\{1-\phi(X;\beta_0)\}]\right)\to N\Big[0,\mathrm{var}\{r(X_1)I_1\}\Big]$$

in distribution, which implies (16).

For (c), we define $d_i^*=(d_{i1}^*,\ldots,d_{ik}^*)^{\mathrm{T}}$ to be the capture history for the individual with the characteristic X_i, $i=1,\ldots,N_0$. Then

$$\frac{1}{N_0}\sum_{i=1}^{n} d_{ij}r(x_i)=\frac{1}{N_0}\sum_{i=1}^{N_0} d_{ij}^*r(X_i)I_i.$$

Note that $d_{ij}^*I_i=d_{ij}^*$. Then

$$E\{d_{ij}^*r(X_i)I_i\}=E[E\{d_{ij}^*r(X_i)|X_i\}]=E\{r(X_i)E(d_{ij}^*\mid X_i)\}=E\{r(X)g(X,\beta_{j0})\},$$

where we use $E(d_{ij}^*\mid X_i)=\mathrm{pr}(d_{ij}^*=1\mid X_i)=g(X_i,\beta_{j0})$. This completes the proof.

From Lemma 1 and (14), we have the following result regarding the limit of $\partial^2H(0)/(\partial\gamma\partial\gamma^{\mathrm{T}})$.

LEMMA 2. *Under the conditions of Theorem* 1*, we have $\partial^2H(0)/(\partial\gamma\partial\gamma^{\mathrm{T}})=V+O_p(N_0^{-1/2})$, where V is defined in (7).*

We concentrate on the result

$$\frac{\partial^2H(0)}{\partial\gamma_1^2}=N_0S_2(N_0,n)=V_{11}+O_p(N_0^{-1/2}).$$

The other results are either trivial or follow from the application of (15) and (16) in Lemma 1.

From (12) and the fact that x^{-2} is a monotone decreasing function of x, we have

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

Recall that $n/N_0 = 1-\alpha_0 + O_p(N_0^{-1/2})$. Then

$$S_2(N_0,n) = -\frac{n}{N_0(N_0-n)} + O_p(N_0^{-2}) = -\frac{1-\alpha_0}{N_0\alpha_0} - O_p(N_0^{-3/2}).$$

Therefore,

$$\frac{\partial^2 H(0)}{\partial \gamma_1^2} = N_0 S_2(N_0,n) = -\frac{1-\alpha_0}{\alpha_0} + O_p(N_0^{-1/2}) = V_{11} + O_p(N_0^{-1/2}).$$

This completes the proof.

From Lemma 1 and (13), we have the following lemma, which summarizes the properties of u_n.

LEMMA 3. *Under the conditions of Theorem* 1*, we have* $E(u_n)=0$, $\mathrm{var}(u_n)=\Sigma$, *and as* $N_0\to\infty$, $u_n \to N(0,\Sigma)$ *in distribution, where*

$$\Sigma = \begin{pmatrix} -V_{11} & 0 & -V_{13} & 0 \\ 0 & -V_{22} & 0 & 0 \\ -V_{31} & 0 & 2V_{34}(1-\alpha_0)^{-2} - V_{33} & V_{44}(1-\alpha_0)^{-2} \\ 0 & 0 & V_{44}(1-\alpha_0)^{-2} & V_{44} \end{pmatrix}.$$

The results that $E(u_n)=0$ and $\mathrm{var}(u_n)=\Sigma$ follow from (15) and (17) in Lemma 1 and some tedious algebra work. With these results, the limiting distribution of u_n follows from the fact that u_n can be expressed as a summation of independent and identically distributed random vectors, as demonstrated in the proof of Lemma 1.

3. PROOFS OF THE MAIN RESULTS IN MAIN PAPER

3·1. *Proof of Theorem* 1

Using a similar argument to that in the proofs of Lemma 1 and Theorem 1 of Qin & Lawless (1994), we have

$$\hat\gamma^{\mathrm{T}} = N_0^{1/2}\{(\hat N/N_0)-1, (\hat\beta-\beta_0)^{\mathrm{T}}, \hat\alpha-\alpha_0, \hat\lambda-\lambda_0\} = O_p(1).$$

Next we investigate the asymptotic approximations of $\hat\gamma$ and the likelihood ratio statistics. The following lemma from Hjort & Pollard (2011) will simplify our derivation.

LEMMA 4. *Assume that* $\theta^{\mathrm{T}} = (\theta_1^{\mathrm{T}}, \theta_2^{\mathrm{T}})$ *where* θ_1 *and* θ_2 *are* r*- and* s*-dimensional vectors, respectively. Let* $\theta_0^{\mathrm{T}} = (\theta_{10}^{\mathrm{T}}, \theta_{20}^{\mathrm{T}})$ *be its true value, and* $\gamma = (\gamma_1^{\mathrm{T}}, \gamma_2^{\mathrm{T}})^{\mathrm{T}} = n^{1/2}(\theta-\theta_0)$ *where* n *is the sample size. Suppose that for* $\theta = \theta_0 + O_p(n^{-1/2})$*, we have*

$$H(\theta) = C_n + 2a_n^{\mathrm{T}}\gamma - \gamma^{\mathrm{T}}A\gamma + \varepsilon_n(\theta)$$

where $a_n = O_p(1)$*,* V *is a positive definite matrix,* C_n *depends only on* θ_0*,* A *is nonsingular, and* $\varepsilon_n(\theta) = O_p(n^{-1/2})$ *for any fixed* θ*. According to* $\theta = (\theta_1^{\mathrm{T}}, \theta_2^{\mathrm{T}})^{\mathrm{T}}$*, we partition* A *into*

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},$$

and partition a_n^{T} *into* $(a_{n1}^{\mathrm{T}}, a_{n2}^{\mathrm{T}})$*. As* $n\to\infty$*, if* $a_n \to N(0,A)$ *in distribution, then*

(a) *the maximizer $\hat{\theta}$ of $H(\theta)$ satisfies*

$$n^{1/2}(\hat{\theta}-\theta_0) = A^{-1}a_n + O_p(n^{-1/2}) \to N(0, A^{-1})$$

in distribution;

(b) $\max_\theta H(\theta) - H(\theta_0) = a_n^{\mathrm{T}} A^{-1} a_n + o_p(1) \to \chi^2_{r+s}$ *in distribution, and*

(c) $\max_\theta H(\theta) - \max_{\theta_2} H(\theta_{10}, \theta_2) = a_n^{\mathrm{T}} A^{-1} a_n - a_{n2}^{\mathrm{T}} A_{22}^{-1} a_{n2} + o_p(1) \to \chi^2_r$ *in distribution.*

Applying the second-order Taylor expansion to $H(\gamma)$ at $\gamma = 0$, we have

$$H(\gamma) = H(0) + \left\{\frac{\partial H(0)}{\partial \gamma}\right\}^{\mathrm{T}} \gamma + \frac{1}{2}\gamma^{\mathrm{T}} \frac{\partial^2 H(0)}{\partial \gamma \partial \gamma^{\mathrm{T}}}\gamma + O_p(N_0^{-1/2}).$$

Recall that $\partial H(0)/\partial\gamma = u_n + O_p(N_0^{-1/2})$. Further, using Lemma 2, we get

$$H(\gamma) = H(0) + u_n^{\mathrm{T}}\gamma + \frac{1}{2}\gamma^{\mathrm{T}} V \gamma + O_p(N_0^{-1/2}). \tag{18}$$

Next we profile out γ_4 and obtain the profile log-likelihood function $\ell(N, \beta, \alpha)$. Recall that for the given β and α, λ is the solution of

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

Equivalently, γ_4 is the solution of

$$\frac{\partial H(\gamma)}{\partial \gamma_4} = 0.$$

Applying the first-order Taylor expansion, we get

$$0 = \frac{\partial H(0)}{\partial \gamma_4} + \frac{\partial^2 H(0)}{\partial\gamma_1 \partial\gamma_4}\gamma_1 + \frac{\partial^2 H(0)}{\partial\gamma_2^{\mathrm{T}} \partial\gamma_4}\gamma_2 + \frac{\partial^2 H(0)}{\partial\gamma_3 \partial\gamma_4}\gamma_3 + \frac{\partial^2 H(0)}{\partial\gamma_4^2}\gamma_4 + O_p(N_0^{-1/2}). \tag{19}$$

With (13) and Lemma 2, (19) is simplified to

$$0 = u_{n4} + V_{42}\gamma_2 + V_{43}\gamma_3 + V_{44}\gamma_4 + O_p(N_0^{-1/2}), \tag{20}$$

which implies that

$$\gamma_4 = -V_{44}^{-1} u_{n4} - V_{44}^{-1}(0, V_{42}, V_{43})\gamma_{-4} + O_p(N_0^{-1/2}), \tag{21}$$

where $\gamma_{-4}^{\mathrm{T}} = (\gamma_1, \gamma_2^{\mathrm{T}}, \gamma_3)$. Substituting (21) into (18), we get an approximation of the profile likelihood,

$$\ell(N, \beta, \alpha) = H(0) - 0.5V_{44}^{-1}u_{n4}^2 + t^{\mathrm{T}}\gamma_{-4} - 0.5\gamma_{-4}^{\mathrm{T}} W \gamma_{-4} + O_p(N_0^{-1/2}) \tag{22}$$

where W is defined in (8) and $t^{\mathrm{T}} = (t_1, t_2^{\mathrm{T}}, t_3)$ with

$$t_1 = u_{n1},\ t_2 = u_{n2} - V_{24}V_{44}^{-1}u_{n4},\ t_3 = u_{n3} - V_{34}V_{44}^{-1}u_{n4}.$$

From Lemma 3, the form of t, and some tedious algebra work, it can be verified that $\mathrm{var}(t) = W$. Hence, $t \to N(0, W)$ in distribution.

Note that in (22), $H(0) - 0.5V_{44}^{-1}u_{n4}^2$ does not depend on γ. Applying Part (a) of Lemma 4, we get

$$\hat{\gamma}_{-4} = N_0^{1/2}\{(\hat{N}/N_0) - 1, (\hat{\beta} - \beta_0)^{\mathrm{T}}, \hat{\alpha} - \alpha_0\}^{\mathrm{T}} = W^{-1}t + O_p(N_0^{-1/2}). \tag{23}$$

With the asymptotic order $N_0^{1/2}\{(\hat N/N_0)-1\} = O_p(1)$, we have

$$N_0^{1/2}\{(\hat N/N_0)-1\} = N_0^{1/2}\log(\hat N/N_0) + O_p(N_0^{-1/2}).$$

Hence,

$$N_0^{1/2}\{\log(\hat N/N_0), (\hat\beta-\beta_0)^{\mathrm{T}}, \hat\alpha-\alpha_0\}^{\mathrm{T}} = W^{-1}t + O_p(N_0^{-1/2}),$$

which converges in distribution to $N(0, W^{-1})$ as claimed in Part (a) of Theorem 1.

Part (b) is a direct application of Parts (b) and (c) of Lemma 4. This completes the proof.

3·2. *Proof of Theorem* 2

We first derive an approximation to $\tilde N$, which depends on that of $\tilde\beta$. Note that $\tilde\beta$ satisfies $\partial\ell_c(\tilde\beta)/\partial\beta = 0$. It can be verified that

$$N_0^{-1/2}\frac{\partial\ell_c(\beta_0)}{\partial\beta} = u_{n2}, \quad \frac{1}{N_0}\frac{\partial^2\ell_c(\beta_0)}{\partial\beta\partial\beta^{\mathrm{T}}} = \frac{\partial^2 H(0)}{\partial\gamma_2\partial\gamma_2^{\mathrm{T}}} = V_{22} + O_p(N_0^{-1/2}).$$

Applying the first-order Taylor expansion to $\partial\ell_c(\tilde\beta)/\partial\beta$ gives

$$N_0^{1/2}(\tilde\beta-\beta_0) = -V_{22}^{-1}u_{n2} + O_p(N_0^{-1/2}). \tag{24}$$

Further, note that the partial derivative of $\sum_{i=1}^n\{1-\phi(x_i,\beta)\}^{-1}$ at $\beta=\beta_0$ is

$$-\sum_{i=1}^n \frac{\phi(x_i,\beta_0)}{\{1-\phi(x_i,\beta_0)\}^2} G_1(x_i,\beta_0)\otimes q(x_i) = -N_0\{V_{32} + O_p(N_0^{-1/2})\}.$$

Using (24), we have

$$\begin{aligned}
N_0^{-1/2}(\tilde N - N_0) &= N_0^{-1/2}\left\{\sum_{i=1}^n \frac{1}{1-\phi(x_i,\tilde\beta)} - N_0\right\}\\
&= N_0^{1/2}\left\{\frac{1}{N_0}\sum_{i=1}^n\frac{1}{1-\phi(x_i,\beta)} - 1\right\} + V_{32}V_{22}^{-1}u_{n2} + O_p(N_0^{-1/2})\\
&= -(u_{n1}+u_{n3}) + V_{32}V_{22}^{-1}u_{n2} + O_p(N_0^{-1/2}).
\end{aligned}$$

Recall that the approximation of $\hat N$ is given in (23). Denote W^{-1} by $(W^{ij})_{1\le i,j\le 3}$. Then the first component of $\hat\gamma_{-4}$ in (23), namely $N_0^{-1/2}(\hat N - N_0)$, can be rewritten as

$$\begin{aligned}
N_0^{-1/2}(\hat N - N_0) &= W^{11}t_1 + W^{13}t_2 + W^{12}t_3 + O_p(N_0^{-1/2})\\
&= W^{11}u_{n1} + W^{13}u_{n3} + W^{12}u_{n2} - (W^{13}V_{34}V_{44}^{-1} + W^{12}V_{24}V_{44}^{-1})u_{n4} + O_p(N_0^{-1/2}).
\end{aligned}$$

With the form of u_n in (13), it can be verified that

$$u_{n4} = (1-\alpha_0)u_{n1} + (1-\alpha_0)^2u_{n3} + O_p(N_0^{-1/2}).$$

Hence,

$$\begin{aligned}
N_0^{-1/2}(\hat N - N_0) = \{&W^{11} - (1-\alpha_0)(W^{13}V_{34}V_{44}^{-1} + W^{12}V_{24}V_{44}^{-1})\}u_{n1} + W^{12}u_{n2}\\
&+\{W^{13} - (1-\alpha_0)^2(W^{13}V_{34}V_{44}^{-1} + W^{12}V_{24}V_{44}^{-1})\}u_{n3} + O_p(N_0^{-1/2}).
\end{aligned}$$

Therefore, if we can prove that

$$W^{11} - (1-\alpha_0)(W^{13}V_{34}V_{44}^{-1} + W^{12}V_{24}V_{44}^{-1}) = -1, \tag{25}$$
$$W^{12} = V_{32}V_{22}^{-1}, \tag{26}$$
$$W^{13} - (1-\alpha_0)^2(W^{13}V_{34}V_{44}^{-1} + W^{12}V_{24}V_{44}^{-1}) = -1, \tag{27}$$

then

$$N_0^{-1/2}(\tilde{N} - N_0) = N_0^{-1/2}(\hat{N} - N_0) + O_p(N_0^{-1/2}),$$

which means $\tilde{N} = \hat{N} + O_p(1)$ and

$$\begin{aligned} N_0^{-1/2}(\hat{N} - N_0) &= N_0^{-1/2}(\tilde{N} - N_0) + O_p(N_0^{-1/2}) \\ &= -(u_{n1} + u_{n3}) + V_{32}V_{22}^{-1}u_{n2} + O_p(N_0^{-1/2}). \end{aligned} \tag{28}$$

With Lemma 3, we will further have that

$$\sigma^2 = \mathrm{var}(u_{n1} + u_{n3} - V_{32}V_{22}^{-1}u_{n2}) = \phi_* - 1 - V_{32}V_{22}^{-1}V_{23},$$

and hence both $N_0^{-1/2}(\tilde{N} - N_0)$ and $N_0^{-1/2}(\hat{N} - N_0)$ converge in distribution to $N(0, \sigma^2)$, which can easily be used to verify the other results in Part (b). This completes the proofs of Parts (a) and (b).

Lastly, we verify that (25)–(27) are correct. Let $\xi = \phi_* - (1-\alpha_0)^{-1}$. Using the relationships

$$\begin{aligned} V_{24} &= (1-\alpha_0)^2 V_{23}, \\ V_{33} &= \xi + \frac{1}{1-\alpha_0} - \frac{1}{\alpha_0}, \\ V_{34} &= (1-\alpha_0)^2\left(\xi + \frac{1}{1-\alpha_0}\right), \\ V_{44} &= (1-\alpha_0)^4\xi, \end{aligned}$$

we can simplify the left-hand sides of (25) and (27) to

$$W^{11} - (1-\alpha_0)(W^{13}V_{34}V_{44}^{-1} + W^{12}V_{24}V_{44}^{-1}) = W^{11} - \frac{W^{12}V_{23}}{(1-\alpha_0)\xi} - W^{13}\frac{\xi + \frac{1}{1-\alpha_0}}{(1-\alpha_0)\xi}, \tag{29}$$
$$W^{13} - (1-\alpha_0)^2(W^{13}V_{34}V_{44}^{-1} + W^{12}V_{24}V_{44}^{-1}) = -\frac{W^{12}V_{23}}{\xi} - \frac{W^{13}}{(1-\alpha_0)\xi}. \tag{30}$$

Further, W in (8) is simplified to

$$W = (W_{ij})_{1\le i,j\le 3} = \begin{pmatrix} \frac{1-\alpha_0}{\alpha_0} & 0 & -\frac{1}{\alpha_0} \\ 0 & -V_{22} + \frac{1}{\xi}V_{23}V_{32} & \frac{1}{(1-\alpha_0)\xi}V_{23} \\ -\frac{1}{\alpha_0} & \frac{1}{(1-\alpha_0)\xi}V_{32} & \frac{1}{(1-\alpha_0)^2\xi} + \frac{1}{\alpha_0(1-\alpha_0)} \end{pmatrix}.$$

Since $W^{-1} = (W^{ij})_{1\le i,j\le 3}$, from the first row of $W^{-1} \times W = I$, we have

$$\frac{1-\alpha_0}{\alpha_0}W^{11} - \frac{1}{\alpha_0}W^{13} = 1, \tag{31}$$
$$W^{12}\left(-V_{22} + \frac{1}{\xi}V_{23}V_{32}\right) + \frac{1}{(1-\alpha_0)\xi}W^{13}V_{32} = 0, \tag{32}$$
$$-\frac{1}{\alpha_0}W^{11} + \frac{1}{(1-\alpha_0)\xi}W^{12}V_{23} + \left\{\frac{1}{(1-\alpha_0)^2\xi} + \frac{1}{\alpha_0(1-\alpha_0)}\right\}W^{13} = 0. \tag{33}$$

It follows from (31) and (33) that

$$\frac{W^{12}V_{23}}{\xi} + \frac{W^{13}}{(1-\alpha_0)\xi} = 1, \tag{34}$$

$$-W^{11} + \frac{1}{(1-\alpha_0)\xi}W^{12}V_{23} + \left\{\frac{1}{(1-\alpha_0)^2\xi} + \frac{1}{(1-\alpha_0)}\right\}W^{13} = 1. \tag{35}$$

Combining (34)–(35) with (29)–(30), we then verify that (25) and (27) are correct.

We now verify (26). From (34), we get

$$W^{13} = (1-\alpha_0)\xi - (1-\alpha_0)W^{12}V_{23}. \tag{36}$$

Substituting (36) into (32) gives $-W^{12}V_{22} + V_{32} = 0$, which implies that (26) is correct. This completes the proof.

3·3. *Consistency of* $\hat{\sigma}^2$

The proof of Theorem 2 indicates that $\tilde{\beta}$ is a root-N_0 estimator of β_0. Therefore,

$$\hat{\phi}_* = \frac{N_0}{\tilde{N}} \cdot \frac{1}{N_0}\sum_{i=1}^{n}\{1-\phi(x_i,\beta_0)\}^{-2} + O_p(N_0^{-1/2}).$$

Theorems 1 and 2 imply $\tilde{N}/N_0 = 1 + O_p(N_0^{-1/2})$. Lemma 1 implies

$$\frac{1}{N_0}\sum_{i=1}^{n}\{1-\phi(x_i,\beta_0)\}^{-2} = \phi^* + O_p(N_0^{-1/2}).$$

Combining the above results, we have

$$\hat{\phi}_* = \left\{1 + O_p(N_0^{-1/2})\right\}\left\{\phi^* + O_p(N_0^{-1/2})\right\} + O_p(N_0^{-1/2}) = \phi^* + O_p(N_0^{-1/2}).$$

With a similar analysis, we found that

$$\begin{aligned}
\hat{V}_{23} &= \frac{N_0}{\tilde{N}} \cdot \frac{1}{N_0}\sum_{i=1}^{n}\frac{\phi(x_i,\beta_0)}{\{1-\phi(x_i,\beta_0)\}^2}G_1(x_i,\beta_0)\otimes q(x_i) + O_p(N_0^{-1/2}) \\
&= E\left[\frac{\phi(x_i,\beta_0)}{1-\phi(x_i,\beta_0)}G_1(x_i,\beta_0)\otimes q(x_i)\right] + O_p(N_0^{-1/2}) \\
&= V_{23} + O_p(N_0^{-1/2}).
\end{aligned}$$

In addition,

$$\begin{aligned}
\hat{V}_{22} = -\frac{N_0}{\tilde{N}} \cdot \frac{1}{N_0}\sum_{i=1}^{n}\left[\left\{d_i - \frac{G_1(x_i,\beta_0)}{1-\phi(x_i,\beta_0)}\right\}\left\{d_i - \frac{G_1(x_i,\beta_0)}{1-\phi(x_i,\beta_0)}\right\}^{\mathrm{T}}\right]\otimes\{q(x_i)q(x_i)^{\mathrm{T}}\} \\
+O_p(N_0^{-1/2}).
\end{aligned}$$

Applying Lemma 1 and the result that $\tilde{N}/N_0 = 1 + O_p(N_0^{-1/2})$, we have

$$\begin{aligned}
\hat{V}_{22} = -E\left\{\frac{1}{N_0}\sum_{i=1}^{n}\left[\left\{d_i - \frac{G_1(x_i,\beta_0)}{1-\phi(x_i,\beta_0)}\right\}\left\{d_i - \frac{G_1(x_i,\beta_0)}{1-\phi(x_i,\beta_0)}\right\}^{\mathrm{T}}\right]\otimes\{q(x_i)q(x_i)^{\mathrm{T}}\}\right\} \\
+O_p(N_0^{-1/2}).
\end{aligned}$$

Using the fact that

$$E(u_{n2}) = E\left[\frac{1}{N_0^{1/2}}\sum_{i=1}^{n}\left\{d_i - \frac{G_1(x_i,\beta_0)}{1-\phi(x_i,\beta_0)}\right\}\otimes q(x_i)\right] = 0,$$

we further have

$$\hat{V}_{22} = \mathrm{var}\left[\frac{1}{N_0^{1/2}}\sum_{i=1}^{n}\left\{d_i - \frac{G_1(x_i,\beta_0)}{1-\phi(x_i,\beta_0)}\right\}\otimes q(x_i)\right] + O_p(N_0^{-1/2}) = V_{22} + O_p(N_0^{-1/2}).$$

Consequently, $\hat{\sigma}^2 = \hat{\phi}_* - 1 - \hat{V}_{32}\hat{V}_{22}^{-1}\hat{V}_{23}$ is a root-N_0 consistent estimator of σ^2.

3·4. *Proof of Corollaries* 1 *and* 2 *and consistency of* $\hat{\sigma}_s^2$

The proofs of Corollaries 1 and 2 are similar to those of Theorems 1 and 2, and the proof of the consistency of $\hat{\sigma}_s^2$ in (10) is similar to that of $\hat{\sigma}^2$. Hence, the details are omitted here.

3·5. *Semiparametric efficiency of* $\hat{N}$

Let $dF(x) = f(x,\theta)dx$ denote a parametric submodel such that $f(x,\theta_0)$ is the true density function of X. Further, let $\hat{N}_p(f,\theta)$ denote the parametric maximum likelihood estimator of N under the parametric submodel $f(x,\theta)$ for the marginal distribution of X. According to Fewster & Jupp (2009), as $N_0 \to \infty$,

$$N_0^{-1/2}\{\hat{N}_p(f,\theta) - N_0\} \to N\Big(0, \sigma_p^2(f,\theta)\Big)$$

for some $\sigma_p^2(f,\theta) > 0$. In this section, we establish the semiparametric efficiency of $\hat{N}$ by showing that the asymptotic variance σ^2 of $\hat{N}$ satisfies

$$\sigma^2 = \sup \sigma_p^2(f,\theta), \tag{37}$$

where the supremum is taken over all parametric submodels for $dF(x)$.

We need some preparation. Let $\eta = 1/(1-\alpha)$, $\eta_0 = 1/(1-\alpha_0)$, and $\hat{\eta} = 1/(1-\hat{\alpha})$. Since we treat N as a continuous parameter, $\hat{N}$ and $\hat{\alpha}$ should satisfy

$$S_1(\hat{N}, n) + \log\hat{\alpha} = 0.$$

Recall that

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

Then

$$\hat{N} = n\hat{\eta} + O_p(1), \tag{38}$$

which implies that

$$\hat{N} - N_0 = n(\hat{\eta} - \eta_0) + n\eta_0 - N_0 + O_p(1). \tag{39}$$

Combining (28) and (38), we get

$$\hat{\eta} - \eta_0 = n^{-1}\sum_{i=1}^{n}\left[V_{32}V_{22}^{-1}\left\{d_i - \frac{G_1(x_i)}{1-\phi(x_i,\beta_0)}\right\}\otimes q(x_i) + \frac{1}{1-\phi(x_i,\beta_0)} - \eta_0\right] + o_p(N_0^{-1/2}). \tag{40}$$

By the central limit theorem and Slutsky's theorem,

$$n^{1/2}(\hat{\eta} - \eta_0) \mid n \to N(0, \sigma_\eta^2)$$

as $n \to \infty$, for some $\sigma_\eta^2 > 0$.

Let $\hat{\eta}_p(f, \theta)$ denote the parametric maximum likelihood estimator of η under the parametric submodel $f(x, \theta)$ for the marginal distribution of X. Similarly to (38), we have

$$\hat{N}_p(f, \theta) = n\hat{\eta}_p(f, \theta) + O_p(1).$$

According to Fewster & Jupp (2009),

$$n^{1/2}\{\hat{\eta}_p(f, \theta) - \eta_0\} \mid n \to N\Big(0, \sigma_{p,\eta}^2(f, \theta)\Big)$$

as $n \to \infty$, for some $\sigma_{p,\eta}^2(f, \theta) > 0$.

We return to the proof of (37). The roadmap is as follows. In the first step, we show that conditional on n, $\hat{\eta}$ is a semiparametric efficient estimator of η, which implies that

$$\sigma_\eta^2 = \sup \sigma_{p,\eta}^2(f, \theta), \tag{41}$$

where the supremum is taken over all parametric submodels for $dF(x)$. In the second step, we show that

$$\sigma^2 = \eta_0^{-1}\sigma_\eta^2 + \eta_0 - 1, \quad \sigma_p^2(f, \theta) = \eta_0^{-1}\sigma_{p,\eta}^2(f, \theta) + \eta_0 - 1,$$

which together with (41) imply (37).

We start with the first step. Let D and X respectively denote the capture history and characteristic of an ideal individual, with D_+ the number of captures in the k occasions, and $\Delta = I(D_+ > 0)$ with $I(\cdot)$ an indicator function. With (40), conditional on n, the influence function of $\hat{\eta}$ is

$$\varphi_\eta(X, D) = V_{32}V_{22}^{-1}\left\{D - \frac{G_1(X)}{1 - \phi(X, \beta_0)}\right\} \otimes q(X) + \frac{1}{1 - \phi(X, \beta_0)} - \eta_0.$$

Referring to the established theory for the semiparametric efficiency bound, for example Chapter 3 of Bickel et al. (1993) and Newey (1990), we need to show only the following two parts to establish the semiparametric efficiency of $\hat{\eta}$ conditional on n:

(a) $\hat{\eta}$ is a regular estimator of η_0;
(b) there exists a parametric submodel with $h_\xi(x, d)$ the joint density of X and D such that the true model is $h_0(x, d)$ and

$$\varphi_\eta(x, d) = \left.\frac{\partial \log h_\xi(x, d)}{\partial \xi}\right|_{\xi=0}.$$

We first consider (a). Following the procedure for the derivation of the likelihood in §2 of the main paper, the joint distribution of X and D conditioning on that it is captured is

$$h(x, d; \theta, \beta) = \{1 - \alpha(\theta, \beta)\}^{-1} f(x, \theta) \prod_{j=1}^{k} g(x, \beta_j)^{d_j}\{1 - g(x, \beta_j)\}^{1-d_j},$$

where $\alpha(\theta, \beta) = \int \phi(x, \beta)f(x, \theta)dx$.

Let

$$B_1(x, d) = \frac{\partial \log h(x, d; \theta_0, \beta_0)}{\partial \theta}, \quad B_2(x, d) = \frac{\partial \log h(x, d; \theta_0, \beta_0)}{\partial \beta}.$$

By Theorem 2.2 in Newey (1990), arguing that $\hat{\eta}$ is a regular estimator of η_0 is equivalent to showing that

$$E_0\left\{\varphi_\eta(X,D)B_1(X,D)\right\} = \partial\eta/\partial\theta = \eta_0^2\partial\alpha/\partial\theta \tag{42}$$

and

$$E_0\left\{\varphi_\eta(X,D)B_2(X,D)\right\} = \partial\eta/\partial\beta = \eta_0^2\partial\alpha/\partial\beta, \tag{43}$$

where E_0 indicates that the expectation is taken under $h(x,d;\theta_0,\beta_0)$. Let $f'(x,\theta) = \partial f(x,\theta)/\partial\theta$. After some calculus, it can be verified that

$$\begin{aligned} B_1(x,d) &= \frac{f'(x,\theta_0)}{f(x,\theta_0)} + \eta_0 E_0\left\{\phi(X,\beta_0)\frac{f'(X,\theta_0)}{f(X,\theta_0)}\right\}, \\ B_2(x,d) &= \{D - G_1(x)\}\otimes q(x) - \eta_0 E_0\{\phi(X,\beta_0)G_1(X)\otimes q(X)\}. \end{aligned}$$

We now consider (42). Note that

$$E_0\{\varphi_\eta(X,D)\} = 0.$$

Hence,

$$\begin{aligned} E_0\left\{\varphi_\eta(X,D)B_1(X,D)\right\} &= E_0\left\{\varphi_\eta(X,D)\frac{f'(X,\theta_0)}{f(X,\theta_0)}\right\} \\ &= E_0\left\{V_{32}V_{22}^{-1}\left\{D - \frac{G_1(X)}{1-\phi(X,\beta_0)}\right\}\otimes q(X)\frac{f'(X,\theta_0)}{f(X,\theta_0)}\right\} \qquad (44) \\ &\quad + E_0\left[\left\{\frac{1}{1-\phi(X,\beta_0)} - \eta_0\right\}\frac{f'(X,\theta_0)}{f(X,\theta_0)}\right\}. \qquad (45) \end{aligned}$$

The term in (44) is equal to zero because

$$E_0(D \mid X) = \frac{G_1(X)}{1-\phi(X,\beta_0)}.$$

The term in (45) is equal to

$$\eta_0 E\left\{\frac{f'(X,\theta_0)}{f(X,\theta_0)}\right\} - \eta_0^2 E\left[\{1-\phi(X,\beta_0)\}\frac{f'(X,\theta_0)}{f(X,\theta_0)}\right] = \eta_0^2 E\left\{\phi(X,\beta_0)\frac{f'(X,\theta_0)}{f(X,\theta_0)}\right\},$$

where E is the expectation with respect to the distribution of X given that the individual has been captured at least once. Therefore,

$$E_0\left\{\varphi_\eta(X,D)B_1(X,D)\right\} = E\left\{\eta_0^2\phi(X,\beta_0)f'(X,\theta_0)\right\} = \eta_0^2\frac{\partial}{\partial\theta}E\left\{\phi(X,\beta_0)f(X,\theta_0)\right\} = \eta_0^2\frac{\partial\alpha}{\partial\theta}.$$

This proves (42).

We proceed to show (43). Since $E_0\{\varphi_\eta(X,D)\} = 0$, we have

$$\begin{aligned} &E_0\left\{\varphi_\eta(X,D)B_2(X,D)\right\} \\ &= E_0\left\{\varphi_\eta(X,D)\{D - G_1(x)\}\otimes q(X)\right\} \\ &= E_0\left[\{D - G_1(X)\}\left\{D - \frac{G_1(X)}{1-\phi(X,\beta_0)}\right\}^{\mathrm{T}}\otimes\left\{q(X)q^{\mathrm{T}}(X)\right\}V_{22}^{-1}V_{23}\right] \qquad (46) \\ &\quad + E_0\left[\left\{\frac{1}{1-\phi(X,\beta_0)} - \eta_0\right\}\{D - G_1(X)\}\otimes q(X)\right]. \qquad (47) \end{aligned}$$

For (46), conditional on X, we have

$$
\begin{aligned}
&E_0\left[\{D-G_1(X)\}\left\{D-\frac{G_1(X)}{1-\phi(X,\beta_0)}\right\}^{\mathrm{T}}\otimes\left\{q(X)q^{\mathrm{T}}(X)\right\}V_{22}^{-1}V_{23}\right]\\
&=E_0\left[\left\{\frac{G_2(X)-G_2^2(X)}{1-\phi(X,\beta_0)}-\frac{\phi(X,\beta_0)G_1(X)G_1^{\mathrm{T}}(X)}{\{1-\phi(X,\beta_0)\}^2}\right\}\otimes\left\{q(X)q^{\mathrm{T}}(X)\right\}V_{22}^{-1}V_{23}\right]\\
&=-\eta_0E\left[\left\{-G_2(X)+G_2^2(X)+\frac{\phi(X,\beta_0)G_1(X)G_1^{\mathrm{T}}(X)}{1-\phi(X,\beta_0)}\right\}\otimes\left\{q(X)q^{\mathrm{T}}(X)\right\}V_{22}^{-1}V_{23}\right]\\
&=-\eta_0V_{22}V_{22}^{-1}V_{23}=-\eta_0V_{23}, \qquad (48)
\end{aligned}
$$

where in the penultimate step we have used the definition of V_{22}.

Similarly, for (47), we get

$$
\begin{aligned}
&E_0\left[\left\{\frac{1}{1-\phi(X,\beta_0)}-\eta_0\right\}\{D-G_1(X)\}\otimes q(X)\right]\\
&=E_0\left[\left\{\frac{1}{1-\phi(X,\beta_0)}-\eta_0\right\}\frac{\phi(X,\beta_0)}{1-\phi(X,\beta_0)}G_1(X)\otimes q(X)\right]\\
&=\eta_0E\left[\left\{\frac{1}{1-\phi(X,\beta_0)}-\eta_0\right\}\phi(X,\beta_0)G_1(X)\otimes q(X)\right]\\
&=\eta_0V_{23}-\eta_0^2E\left\{\phi(X,\beta_0)G_1(X)\otimes q(X)\right\}, \qquad (49)
\end{aligned}
$$

where in the last step we have used the definition of V_{23}. Combining (46)–(49), we obtain

$$E_0\left\{\varphi_\eta(X,D)B_2(X,D)\right\}=-\eta_0^2E\left\{\phi(X,\beta_0)G_1(X)\otimes q(X)\right\}$$

which is exactly $\eta_0^2\partial\alpha/\partial\beta$. This completes the proof of (a).

For (b), we consider the following function

$$h_\xi(x,d)=\left\{1+\xi\varphi_\eta(x,d)\right\}(1-\alpha_0)^{-1}f_0(x)\prod_{j=1}^{k}g(x,\beta_{j0})^{d_j}\{1-g(x,\beta_{j0})\}^{1-d_j}, \qquad (50)$$

where $f_0(x)$ is the true density of X. If X has a compact support C, then $\max_{x\in C}\phi(x,\beta_0)<1$ and $\varphi_\eta(x,d)$ is bounded. Then it is easy to check that for sufficiently small ξ this $h_\xi(x,d)$ is a parametric submodel and

$$\varphi_\eta(x,d)=\left.\frac{\partial\log h_\xi(x,d)}{\partial\xi}\right|_{\xi=0}.$$

This completes the proof of (b), and hence the semiparametric efficiency of $\hat\eta$ is established.

We now move to the second step of proving (37) by identifying the relationship between σ^2 and σ_η^2. The relationship between $\sigma_p^2(f,\theta)$ and $\sigma_{p,\eta}^2(f,\theta)$ can be similarly proved.

Recall that $\hat N=n\hat\eta+O_p(1)$. This implies that

$$n^{-1/2}(\hat N-n\eta_0)=n^{1/2}(\hat\eta-\eta_0)+o_p(1).$$

Therefore,

$$n^{-1/2}(\hat N-n\eta_0)\mid n\to N\left(0,\sigma_\eta^2\right)$$

as $n\to\infty$. Note that $N_0/n=\eta_0+o_p(1)$. By Slutsky's theorem, we further have

$$N_0^{-1/2}(\hat N-n\eta_0)\mid n\to N\left(0,\eta_0^{-1}\sigma_\eta^2\right)$$

as $n \to \infty$.

Because the above limiting distribution does not depend on n, we conclude that as $N_0 \to \infty$, $N_0^{-1/2}(\hat{N} - n\eta_0)$ is asymptotically independent of n or $N_0^{-1/2}(n\eta_0 - N_0)$, and

$$N_0^{-1/2}(\hat{N} - n\eta_0) \sim N(0, \sigma_\eta^2/\eta_0).$$

Recall that $n \sim B(N_0, 1 - \alpha_0 = \eta_0^{-1})$, which implies that $N_0^{-1/2}(n\eta_0 - N_0) \sim N(0, \eta_0 - 1)$. Hence,

$$N_0^{-1/2}(\hat{N} - N_0) = N_0^{-1/2}(\hat{N} - n\eta_0) + N^{-1/2}(n\eta_0 - N_0) \sim N(0, \eta_0^{-1}\sigma_\eta^2 + \eta_0 - 1). \quad (51)$$

That is,

$$\sigma^2 = \eta_0^{-1}\sigma_\eta^2 + \eta_0 - 1. \quad (52)$$

Similarly, we have

$$\sigma_p^2(f, \theta) = \eta_0^{-1}\sigma_{p,\eta}^2(f, \theta) + \eta_0 - 1. \quad (53)$$

Combining (52)–(53) with (41) leads to (37). This completes the proof of (37).

In practice, we may round $\hat{N}$ to the closest integer $\hat{N}_t$. Then

$$|\hat{N} - \hat{N}_t| \leq 1.$$

Hence,

$$\hat{N} = \hat{N}_t + O_p(1),$$

which implies that $N_0^{-1/2}(\hat{N} - N_0)$ and $N_0^{-1/2}(\hat{N}_t - N_0)$ have the same limiting distribution. That is, $\hat{N}_t$ is also semiparametric efficient in the sense that the asymptotic variance of $\hat{N}_t$ is the supremum of the asymptotic variances of the maximum parametric likelihood estimator of N under all parametric submodels.

3·6. *Consistency of the weighted kernel density estimator* $\hat{f}_w(x)$

Given the maximum empirical likelihood estimators $\hat{\beta}_s$ and $\hat{\alpha}_s$, let $\hat{p}_{si} = n^{-1}[1 + \hat{\lambda}_s\{\phi_s(x_i, \hat{\beta}_s) - \hat{\alpha}_s\}]^{-1}$ and let $\hat{\lambda}_s$ be the solution to

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

We propose a weighted kernel estimator

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

for the covariate density function $f(x)$, where $K(x)$ is a kernel function, usually chosen to be the standard normal density function, and h a bandwidth. In contrast, the usual kernel density estimator is defined as

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

Next we restate the properties of $\hat{f}_w(x)$ and $\hat{f}_u(x)$ in the following proposition.

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

$$\hat{f}_w(x) = f(x) + o_p(1), \;\; \hat{f}_u(x) = (1-\alpha_0)^{-1}\{1-\phi_s(x,\beta_0)\}f(x) + o_p(1).$$

We now give a proof for the above proposition. The proof of Theorem 1 and Corollary 1 implies that

$$\hat{\lambda}_s = -(1-\alpha_0)^{-1} + O_p(N_0^{-1/2}), \quad \hat{\beta}_s = \beta_{s0} + O_p(N_0^{-1/2}), \quad \hat{\alpha}_s = \alpha_0 + O_p(N_0^{-1/2}).$$

Because the support of X is compact, there must exist $\epsilon_0 \in (0,\alpha_0)$ such that $\epsilon_0 \le \phi_s(x,\beta_{s0}) \le 1-\epsilon_0$ uniformly over all x. Using the first-order Taylor expansion and the condition that $K(x)$ is a bounded function, we have that

$$\begin{aligned}
\hat{f}_w(x) &= (1-\alpha_0)\frac{1}{n}\sum_{i=1}^{n}\frac{K\{(x_i-x)h^{-1}\}h^{-1}}{1-\phi_s(x_i,\beta_{s0})} + O_p\{1/(N_0h^2)^{1/2}\} \\
&= (1-\alpha_0)\frac{1}{N_0}\frac{N_0}{n}\sum_{i=1}^{n}\frac{K\{(x_i-x)h^{-1}\}h^{-1}}{1-\phi_s(x_i,\beta_{s0})} + o_p(1),
\end{aligned}$$

where in the last step, we have used the condition $N_0h^2 \to \infty$ as $N_0 \to \infty$.

Recall that $n/N_0 = 1-\alpha_0 + o_p(1)$. Then

$$\hat{f}_w(x) = \frac{1}{N_0}\{1+o_p(1)\}\sum_{i=1}^{N_0} I(d_{i+}^* > 0)\frac{K\{(X_i-x)h^{-1}\}h^{-1}}{1-\phi_s(X_i,\beta_{s0})} + o_p(1),$$

where d_{i+}^* is the number of times that the individual with covariate X_i has been captured in the k occasions. By the law of large numbers, we further have

$$\begin{aligned}
\hat{f}_w(x) &= E\left[I(d_{i+}^* > 0)\frac{K\{(X_i-x)h^{-1}\}h^{-1}}{1-\phi_s(X_i,\beta_{s0})}\right]\{1+o_p(1)\} + o_p(1) \\
&= E\left[K\{(X_i-x)h^{-1}\}h^{-1}\right]\{1+o_p(1)\} + o_p(1).
\end{aligned}$$

If $K(x)$ is a bounded, symmetric, and continuous density function, then it satisfies the conditions in Theorem 1A of Parzen (1962). Applying that theorem, we have

$$E\left[K\{(X_i-x)h^{-1}\}h^{-1}\right] = f(x) + o_p(1),$$

where $h = o(1)$ as $N_0 \to \infty$. Hence, we have shown the consistency of the proposed weighted kernel density estimator $\hat{f}_w(x)$.

For the usual kernel density estimator, we similarly have

$$\hat{f}_u(x) = N_0^{-1}\left\{(1-\alpha_0)^{-1} + o_p(1)\right\}\sum_{i=1}^{N_0} I(d_{i+}^* > 0)K\{(X_i-x)h^{-1}\}h^{-1} + o_p(1).$$

By the law of large numbers, we get that

$$\begin{aligned}
\hat{f}_u(x) &= \left\{(1-\alpha_0)^{-1} + o_p(1)\right\} E\left[I(d_{i+}^* > 0)K\{(X_i - x)h^{-1}\}h^{-1}\}\right] + o_p(1) \\
&= \left\{(1-\alpha_0)^{-1} + o_p(1)\right\} E\left[\{1-\phi_s(X_i,\beta_0)\}h^{-1}K\{(X_i - x)h^{-1}\}\right] + o_p(1) \\
&= \left\{(1-\alpha_0)^{-1} + o_p(1)\right\} \int \{1-\phi_s(y,\beta_0)\}h^{-1}K\{(y-x)h^{-1}\}f(y)dy + o_p(1) \\
&= (1-\alpha_0)^{-1}\{1-\phi_s(x,\beta_0)\}f(x) + o_p(1).
\end{aligned}$$

This completes the proof of Proposition 1.

4. NUMERICAL IMPLEMENTATION OF EMPIRICAL LIKELIHOOD METHODS

In the numerical calculation of empirical likelihood methods, a crucial step is to calculate the Lagrange multiplier λ. Recall that given (β, α), the empirical log-likelihood achieves its maximum in general when

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

where the Lagrange multiplier λ is the solution to

$$\sum_{i=1}^{n} \frac{\phi(x_i,\beta)-\alpha}{1+\lambda\{\phi(x_i,\beta)-\alpha\}} = 0. \tag{54}$$

The fact that the p_i's are probability weights implies that $0 < p_i < 1$ for all $1 \le i \le n$ or equivalently

$$1+\lambda\{\phi(x_i,\beta)-\alpha\} > 1/n, \quad 1 \le i \le n. \tag{55}$$

Owen (1988) showed that the solution of (54) exists under constraint (55) if and only if $\min_i\{\phi(x_i,\beta)-\alpha\} < 0 < \max_i\{\phi(x_i,\beta)-\alpha\}$. In this situation, the solution is unique, and constraint (55) implies that λ should lie in

$$J(\beta,\alpha) = \left(-\frac{1-n^{-1}}{\max_i\{\phi(x_i,\beta)-\alpha\}},\ -\frac{1-n^{-1}}{\min_i\{\phi(x_i,\beta)-\alpha\}}\right).$$

We can use the R function `uniroot` to search for the solution of (54) in the interval $J(\beta,\alpha)$.

Under certain regularity conditions,

$$\lim_{N_0\to\infty} \mathrm{pr}\left[\min_i\{\phi(x_i,\beta)-\alpha\} < 0 < \max_i\{\phi(x_i,\beta)-\alpha\}\right] = 1.$$

See Owen (1988). For certain values of (β,α) and a finite sample size, we may not have $\min_i\{\phi(x_i,\beta)-\alpha\} < 0 < \max_i\{\phi(x_i,\beta)-\alpha\}$. In this situation, the solution of (54) does not exist, and hence the profile empirical log-likelihood $\ell(N,\beta,\alpha)$ in (2) is not well defined. To overcome this difficulty, we follow a method proposed by Owen (1990) in our numerical implementation.

Recall that

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

It can easily be verified that $h(N,\beta,\alpha,\lambda)$ is strictly convex in λ and the solution of (54), if it exists, minimizes $h(N,\beta,\alpha,\lambda)$ with respect to λ for the given (N,β,α). Hence, we can minimize $h(N,\beta,\alpha,\lambda)$ to find the solution of (54). However, $h(N,\beta,\alpha,\lambda)$ is not always well defined.

Following the idea in Owen (1990), we first extend the definition of $h(N,\beta,\alpha,\lambda)$ to $h_*(N,\beta,\alpha,\lambda)$, where

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

Here

$$\log_*(z) = \begin{cases} \log(z), & z > c_n, \\ \log(c_n) - 1.5 + 2z/c_n - 0.5(z/c_n)^2, & z \le c_n, \end{cases}$$

where $c_n > 0$ is usually chosen to be very small, e.g. $c_n = 1/n$ or 10^{-5}. The function $\log_*(z)$ is twice continuously differentiable and strictly concave throughout the whole real line. Hence, for given (N,β,α), $h_*(N,\beta,\alpha,\lambda)$ is strictly convex and is well defined for all (N,β,α,λ). For small c_n, $h_*(N,\beta,\alpha,\lambda)$ is a very close approximation to $h(N,\beta,\alpha,\lambda)$ when the latter is well defined.

We next minimize $h_*(N,\beta,\alpha,\lambda)$ with respect to λ to calculate the Lagrange multiplier for the given (N,β,α) and define the profile empirical log-likelihood of (N,β,α) as

$$\ell(N,\beta,\alpha) = \arg\min_{\lambda} h_*(N,\beta,\alpha,\lambda).$$

The optimization problem can easily be solved using the `R` function `optimize`. Our simulation experience indicates that this procedure is computationally efficient and stable.

By implementing the idea in Owen (1990), we overcome the non-definition problem of the profile empirical log-likelihood $\ell(N,\beta,\alpha)$. The resulting $\ell(N,\beta,\alpha)$ is always well defined and is a smooth function of (N,β,α). When calculating the maximum empirical likelihood estimator of (N,β,α), we use a divide-and-conquer strategy to maximize $\ell(N,\beta,\alpha)$.

Note that $\ell(N,\beta,\alpha)$ can be rewritten

$$\ell(N,\beta,\alpha) = h_1(N,\alpha) + h_{23}(\beta,\alpha)$$

where

$$h_1(N,\alpha) = \log\left\{\frac{\Gamma(N+1)}{\Gamma(N-n+1)}\right\} + (N-n)\log\alpha,$$

$$h_{23}(\beta,\alpha) = \min_{\lambda} h_{2*}(\beta,\alpha,\lambda) + h_3(\beta),$$

with

$$h_{2*}(\beta, \alpha, \lambda) = -\sum_{i=1}^{n} \log *[1 + \lambda\{\phi(x_i, \beta) - \alpha\}],$$

$$h_3(\beta) = \sum_{i=1}^{n}\sum_{j=1}^{k} [d_{ij} \log g(x_i, \beta_j) + (1 - d_{ij}) \log\{1 - g(x_i, \beta_j)\}] .$$

We propose to maximize $\ell(N, \beta, \alpha)$ via the following algorithm:

Step 1. Given β and α, obtain $\min_\lambda h_{2*}(\beta, \alpha, \lambda)$ and hence $h_{23}(\beta, \alpha)$. This step can be carried out using the `R` function `optimize`.

Step 2. Given α, maximize $h_1(N, \alpha)$ with respect to N to obtain $\max_N h_1(N, \alpha)$ and maximize $h_{23}(\beta, \alpha)$ with respect to β to obtain $\max_\beta h_{23}(\beta, \alpha)$. Let

$$h_{123}(\alpha) = \max_N h_1(N, \alpha) + \max_\beta h_{23}(\beta, \alpha).$$

This step can be carried out by applying the `R` functions `optimize` and `nlminb` respectively to $h_1(N, \alpha)$ and $h_{23}(\beta, \alpha)$ for the given α.

Step 3. Maximizing $h_{123}(\alpha)$ with respect to α gives the maximum empirical likelihood estimator $\hat{\alpha}$. This step can be carried out by applying the `R` function `optimize` to $h_{123}(\alpha)$. Then maximizing $h_1(N, \hat{\alpha})$ with respect to N gives $\hat{N}$ and maximizing $h_{23}(\beta, \hat{\alpha})$ with respect to β gives $\hat{\beta}$.

The above algorithm has been implemented for both the general case and the special case. In `abun.R`, the `gabun` function implements the empirical likelihood and conditional likelihood methods for the general case, and the `sabun` function implements these methods for the special case. See the accompanying `example.R` for the use of these functions. Both `R` files are available at `http://sas.uwaterloo.ca/~p4li/publications/abun.zip`.

Next we use simulation to compare the computational times for calculating the maximum empirical likelihood estimator, $\hat{N}$ or $\hat{N}_s$, and the maximum conditional likelihood estimator, $\tilde{N}$ or $\tilde{N}_s$, of N. In the simulation, we generate random samples from Scenario S1 and record the times to calculate $\hat{N}$ and $\tilde{N}$ under the M_{th} model, and $\hat{N}_s$ and $\tilde{N}_s$ under the M_{h} model. Based on 100 repetitions, we record the averages of the times in seconds on an IMAC with a 3.4-GHz Intel Core i7 processor. The results are summarized in Table 1. Under both M_{h} and M_{th}, the time to calculate the maximum empirical likelihood estimator increases as N_0 or k increases. The averages of the times to calculate $\hat{N}$ under the M_{th} model are less than 8 seconds when $N_0 = 5000$ and $k = 4$; and the averages of the times to calculate $\hat{N}_s$ under the M_{h} model are less than 3 seconds when $N_0 = 5000$ and $k = 16$. We acknowledge that it takes more time to calculate the maximum empirical likelihood estimator than the maximum conditional likelihood estimator. However, this is the price to pay for a more efficient method.

5. ADDITIONAL SIMULATION RESULTS

5·1. *Some plots*

In this section, we first display quantile-quantile plots of the empirical likelihood ratio $R'(N_0)$ of N versus the χ^2_1 distribution, the pivotal $(\tilde{N} - N_0)/(\tilde{N}^{1/2}\hat{\sigma})$ versus the $N(0, 1)$ distribution, the pivotal $\tilde{N}^{1/2} \log(\tilde{N}/N_0)/\hat{\sigma}$ versus the $N(0, 1)$ distribution, and the pivotal $C(N_0; \tilde{N})$ versus the $N(0, 1)$ distribution for Scenario G1 with $N_0 = 200$. The quantile-quantile plots for $k = 2$

Table 1. *Average times in seconds to compute the maximum empirical likelihood and maximum conditional likelihood estimators of N under $M_{\rm th}$ and $M_{\rm h}$ models.*

		Model $M_{\rm th}$			Model $M_{\rm h}$	
N_0	k	$\hat{N}$	$\tilde{N}$	k	$\hat{N}_s$	$\tilde{N}_s$
100	2	0.26	0.01	2	0.13	< 0.01
100	3	0.29	0.03	8	0.14	< 0.01
100	4	0.34	0.06	16	0.17	< 0.01
1000	2	1.10	0.14	2	0.37	< 0.01
1000	3	1.55	0.39	8	0.57	0.01
1000	4	1.63	0.79	16	0.61	0.01
5000	2	6.13	0.96	2	1.53	0.01
5000	3	7.53	2.76	8	2.63	0.04
5000	4	7.90	5.68	16	2.88	0.04

and $k = 3$ are in Figures 1 and 2, respectively. The plots for the remaining cases are similar and omitted. These two figures indicate that the distribution of the empirical likelihood ratio $R'(N_0)$ is quite close to χ_1^2, and the distributions of $(\tilde{N} - N_0)/(\tilde{N}^{1/2}\hat{\sigma})$ and $\tilde{N}^{1/2}\log(\tilde{N}/N_0)/\hat{\sigma}$ are not close to normal. They also show that the distribution of $C(N_0; \tilde{N})$ is quite close to normal. These observations may explain why the empirical-likelihood-ratio-based confidence intervals $\mathcal{I}_1$ always have more accurate coverage probabilities than the Wald-type confidence intervals $\mathcal{I}_2$ and $\mathcal{I}_3$ but only a slight advantage over $\mathcal{I}_4$.

We next display boxplots of the logarithms of the lengths of $\mathcal{I}_1, \ldots, \mathcal{I}_4$ under Scenario G1 in Figure 3. Together with the results for the coverage probabilities, we observe that $\mathcal{I}_1$ has slightly longer lengths than I_2 and I_3 but much better coverage accuracy. Further, $\mathcal{I}_1$ in general has shorter lengths than $\mathcal{I}_4$, but better or comparable coverage accuracy. The plots and conclusions for the remaining cases are similar and omitted.

The plots of $\hat{N}$ versus $\tilde{N}$ and $\log \hat{N}$ versus $\log \tilde{N}$ in Figure 4 show that the two abundance estimators $\tilde{N}$ and $\hat{N}$ are indeed quite close, although $\tilde{N}$ is slightly larger than $\hat{N}$ in general.

5·2. *Simulation results for small N_0*

In this section, we conduct more simulations for $N_0 = 100, 150$ under Scenarios G1, G2, S1, and S2 to determine how the asymptotic results work for small N_0. The simulated coverage probabilities of $\mathcal{I}_1, \ldots, \mathcal{I}_4$ under Scenarios G1 and G2 and those of $\mathcal{I}_{1s}, \ldots, \mathcal{I}_{4s}$ under Scenarios S1 and S2 at the nominal level 95% are summarized in Table 2.

We can see that the asymptotic theory works reasonably well for all four types of confidence intervals and all sample sizes considered in the simulation under Scenarios G1 and G2 with $k = 3$, especially for the empirical-likelihood-ratio-based confidence interval $\mathcal{I}_1$ and the Wald-type confidence interval $\mathcal{I}_4$. When $k = 2$, $\mathcal{I}_1$ has better coverage probabilities than the other three confidence intervals. However, the general trend for all the confidence intervals is that the asymptotic theory performs worse as N_0 decreases. Some finite-sample correction may be required in the application of $\mathcal{I}_1$ to small N_0 and $k = 2$ under $M_{\rm th}$ models.

For Scenarios S1 and S2, the asymptotic theory works reasonably well for the empirical-likelihood-ratio-based confidence interval $\mathcal{I}_{1s}$ with $k = 2$. The coverage for $\mathcal{I}_{1s}$ is much better than that for the other confidence intervals. In particular, in Scenario S1 with $N_0 = 100$ and $k = 2$, the coverage gain of $\mathcal{I}_{1s}$ over the other three intervals is at least 7%. When $k = 2$, $\mathcal{I}_{4s}$ can have worse coverage probabilities than $\mathcal{I}_{2s}$ and $\mathcal{I}_{3s}$. When $k = 8$, the asymptotic theory does

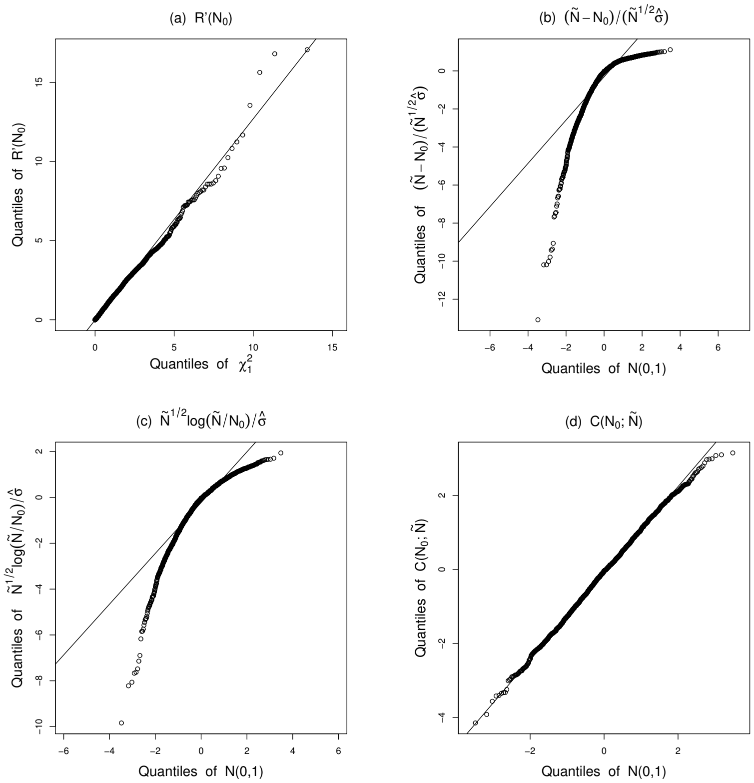

Fig. 1. Simulation results for Scenario G1 with $N_0 = 200$ and $k = 2$. Panel (a) is a quantile-quantile plot of the empirical likelihood ratio $R'(N_0)$ with the theoretical χ_1^2 quantiles. Panel (b) is a quantile-quantile plot of $(\tilde{N} - N_0)/(\tilde{N}^{1/2}\hat{\sigma})$ with the theoretical standard normal quantiles. Panel (c) is a quantile-quantile plot of $\tilde{N}^{1/2}\log(\tilde{N}/N_0)/\hat{\sigma}$ with the theoretical standard normal quantiles. Panel (d) is a quantile-quantile plot of $C(N_0; \tilde{N})$ with the theoretical standard normal quantiles. In all panels, the solid line is the identity line.

Table 2. *Coverage probabilities in percentages for $\mathcal{I}_1, \ldots, \mathcal{I}_4$ under Scenarios G1 and G2 and $\mathcal{I}_{1s}, \ldots, \mathcal{I}_{4s}$ under Scenarios S1 and S2 with $N_0 = 100, 150$. Here the nominal level is 95%.*

		Scenario G1				Scenario G2			
N_0	k	$\mathcal{I}_1$	$\mathcal{I}_2$	$\mathcal{I}_3$	$\mathcal{I}_4$	$\mathcal{I}_1$	$\mathcal{I}_2$	$\mathcal{I}_3$	$\mathcal{I}_4$
100	2	90.3	83.4	86.2	88.4	91.2	87.7	90.3	89.9
100	3	93.9	91.1	92.7	95.1	94.5	91.8	93.7	95.0
150	2	91.8	85.8	88.5	90.2	92.7	88.2	90.3	91.9
150	3	93.3	92.0	92.8	93.9	94.3	92.8	94.3	95.1
		Scenario S1				Scenario S2			
N_0	k	$\mathcal{I}_{1s}$	$\mathcal{I}_{2s}$	$\mathcal{I}_{3s}$	$\mathcal{I}_{4s}$	$\mathcal{I}_{1s}$	$\mathcal{I}_{2s}$	$\mathcal{I}_{3s}$	$\mathcal{I}_{4s}$
100	2	93.6	84.0	86.7	82.8	91.6	86.8	89.0	87.8
100	8	90.1	83.0	84.5	91.2	87.5	84.2	85.6	89.3
150	2	93.8	84.6	87.7	85.0	92.4	88.1	89.8	90.0
150	8	90.0	85.0	86.6	90.8	88.1	85.5	86.9	89.1

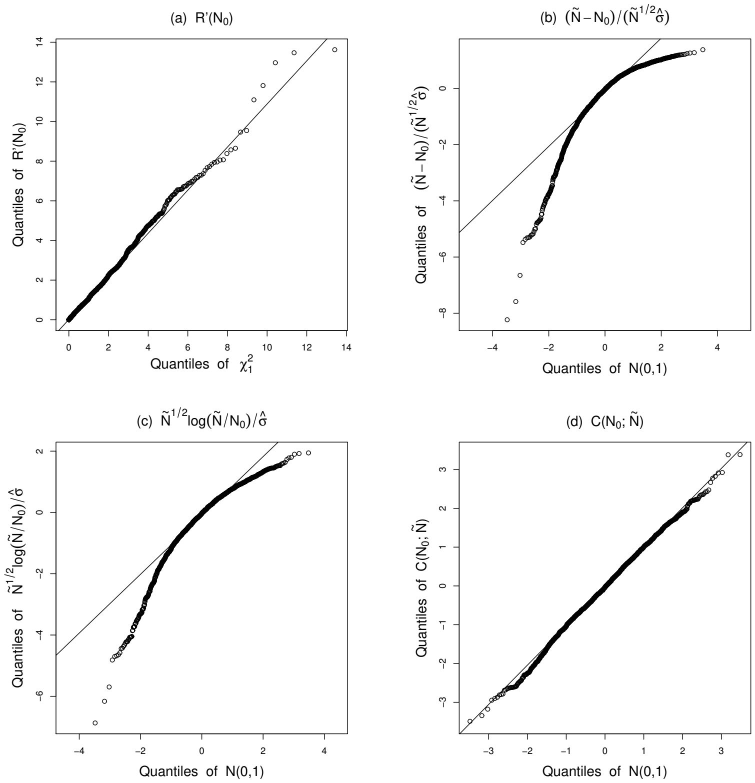

Fig. 2. Simulation results for Scenario G1 with $N_0 = 200$ and $k = 3$. Panel (a) is a quantile-quantile plot of the empirical likelihood ratio $R'(N_0)$ with the theoretical χ_1^2 quantiles. Panel (b) is a quantile-quantile plot of $(\tilde{N} - N_0)/(\tilde{N}^{1/2}\hat{\sigma})$ with the theoretical standard normal quantiles. Panel (c) is a quantile-quantile plot of $\tilde{N}^{1/2}\log(\tilde{N}/N_0)/\hat{\sigma}$ with the theoretical standard normal quantiles. Panel (d) is a quantile-quantile plot of $C(N_0;\tilde{N})$ with the theoretical standard normal quantiles. In all panels, the solid line is the identity line.

not work well for any of the confidence intervals. Again, some finite-sample correction may be required in the application of $\mathcal{I}_{1s}$ when k is large.

5·3. *Simulation results for the special case with large N_0*

In §4 of the main paper, we noticed that under Scenarios S1 and S2, the empirical-likelihood-ratio-based confidence interval $\mathcal{I}_{1s}$ has reduced coverage probabilities as k increases. We now conduct more simulations with $N_0 = 1000, 5000, 10000$ under Scenarios S1 and S2 with 2000 repetitions. The simulated coverage probabilities of $\mathcal{I}_{1s}$, …, $\mathcal{I}_{4s}$ are summarized in Table 3. Clearly, the undesirable trend for $\mathcal{I}_{1s}$ persists when N_0 is increased to 10000 but is less severe when N_0 is increased to 1000.

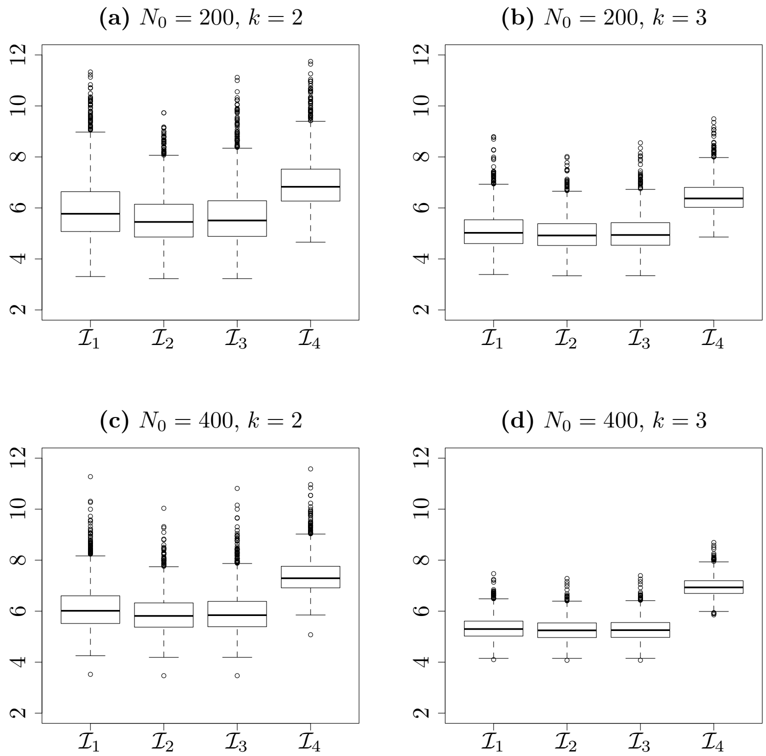

Fig. 3. Boxplots of the logarithm of lengths of $\mathcal{I}_1, \ldots, \mathcal{I}_4$ under Scenario G1.

Table 3. *Coverage probabilities in percentages of $\mathcal{I}_{1s}$, …, $\mathcal{I}_{4s}$ at the nominal level 95% under Scenarios S1 and S2 with $N_0 = 1000, 5000, 10000$.*

		Scenario S1				Scenario S2			
N_0	k	$\mathcal{I}_{1s}$	$\mathcal{I}_{2s}$	$\mathcal{I}_{3s}$	$\mathcal{I}_{4s}$	$\mathcal{I}_{1s}$	$\mathcal{I}_{2s}$	$\mathcal{I}_{3s}$	$\mathcal{I}_{4s}$
1000	2	93.7	89.1	91.0	91.7	93.2	91.0	92.0	93.0
1000	8	93.7	90.1	91.2	94.1	92.5	91.7	92.2	93.3
5000	2	94.1	90.9	92.2	93.2	93.6	93.1	93.5	93.8
5000	8	93.7	92.5	92.8	93.9	93.1	93.1	93.1	93.7
10000	2	94.3	92.9	93.7	94.3	94.5	94.4	94.3	94.7
10000	8	93.9	93.1	93.4	94.1	93.1	92.9	92.9	93.5

5·4. *Simulation results for one-tailed interval estimation*

In the general case, let

$$\begin{aligned}
\omega_1 &= \text{sign}(\hat{N} - N_0)\{R'(N_0)\}^{1/2}, \\
\omega_2 &= (\tilde{N} - N_0)/(\tilde{N}^{1/2}\hat{\sigma}), \\
\omega_3 &= \tilde{N}^{1/2}\log(\tilde{N}/N_0)/\hat{\sigma}, \\
\omega_4 &= C(N_0; \tilde{N}).
\end{aligned}$$

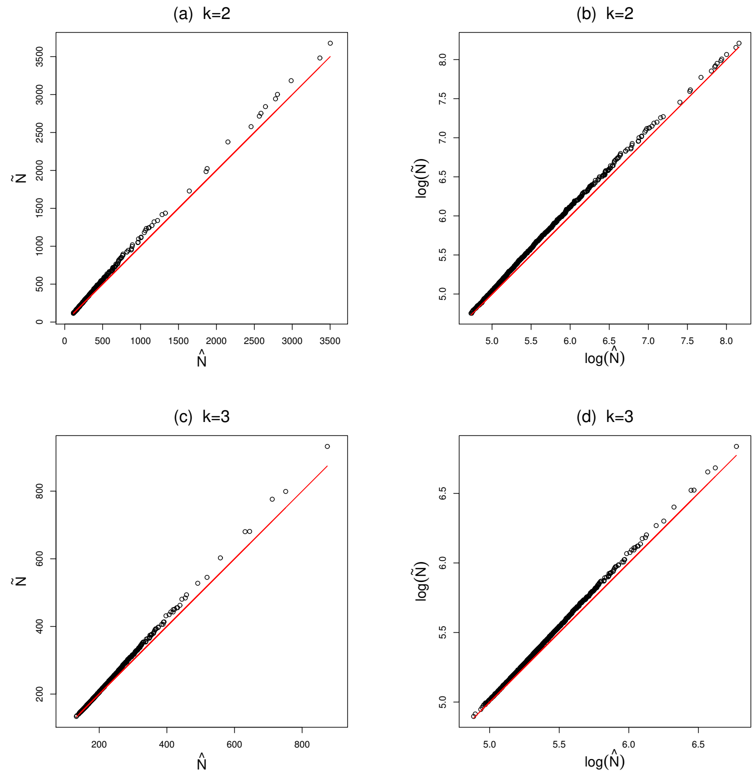

Fig. 4. Comparison of $\hat{N}$ and $\tilde{N}$ for Scenario G1 with $N_0 = 200$. Panels (a) and (c) are plots of $\hat{N}$ versus $\tilde{N}$ for $k = 2$ and $k = 3$. Panels (b) and (d) are plots of $\log \hat{N}$ versus $\log \tilde{N}$ for $k = 2$ and $k = 3$.

That is, ω_1 denotes the signed square root of the empirical likelihood ratio statistic $R'(N_0)$, and $\omega_2, \ldots, \omega_4$ denote three asymptotic pivotal statistics based on the maximum conditional likelihood estimator $\tilde{N}$. Based on the asymptotic results developed in the main paper, $\omega_1, \ldots, \omega_4$ all have the limiting distribution $N(0, 1)$ as $N_0 \to \infty$. In §3 of the main paper, we discussed the two-sided coverage probabilities of the confidence intervals based on $\omega_1, \ldots, \omega_4$. In this section, we study the one-sided coverage probabilities of the confidence intervals based on $\omega_1, \ldots, \omega_4$.

For each of the four statistics $\omega_1, \ldots, \omega_4$, we calculate the simulated probabilities that the statistic is smaller than the 1%, 2.5%, 5%, 95%, 97.5%, and 99% quantiles of $N(0, 1)$ based on 2000 repetitions. The results for Scenarios G1 and G2 with $N_0 = 200$, 400 are summarized in Table 4. Similarly to $\omega_1, \ldots, \omega_4$, we can define $\omega_{1s}, \ldots, \omega_{4s}$ for the special case. The simulation results for Scenarios S1 and S2 with $N_0 = 200$, 400 are summarized in Table 5.

We observe that the distributions of ω_2 and ω_3, including ω_{2s} and ω_{3s}, are much larger than and not close to the standard normal distribution. This observation is consistent with Figures 1 and 2, where the normal quantiles are larger than those of ω_2 and ω_3. Compared with ω_2 and ω_3, the distributions of ω_1 and ω_4, including ω_{1s} and ω_{4s}, are closer to the standard normal distribution. It can be seen that the quantiles of ω_1 and ω_{1s} are uniformly less than the standard normal. This could explain the stable performance of the two-sided confidence interval based on the empirical likelihood ratio or equivalently based on ω_1 and ω_{1s}: there may be location shifts in their distributions. DiCiccio and Romano (1989)'s adjustment method may be applied to reduce the bias in the one-sided coverage probabilities. At the same time, we observe that ω_4 and ω_{4s} seem to shrink towards 0, the median of the standard normal, when $k = 2$. Compared with the standard normal, the probabilities are larger at the lower-half normal quantiles and smaller at the upper-half normal quantiles. As k increases, the shrinkage is alleviated and the distribution of ω_4 becomes closer to the standard normal. This explains why $\mathcal{I}_4$ and $\mathcal{I}_{4s}$ have good performance for large k but severe undercoverage for small k such as $k = 2$.

Table 4. *Simulated probabilities that $\omega_1, \ldots, \omega_4$ are smaller than 1%, 2.5%, 5%, 95%, 97.5%, and 99% quantiles of $N(0, 1)$ under Scenarios G1 and G2 with $N_0 = 200$, 400.*

		Scenario G1				Scenario G2			
		$N_0 = 200$		$N_0 = 400$		$N_0 = 200$		$N_0 = 400$	
Statistic	Level	$k = 2$	$k = 3$	$k = 2$	$k = 3$	$k = 2$	$k = 3$	$k = 2$	$k = 3$
ω_1	1%	2.6	2.3	3.1	1.8	2.4	1.5	1.8	1.3
ω_1	2.5%	5.6	4.8	6.1	4.2	5.7	3.3	4.7	2.7
ω_1	5%	10.2	7.8	10.5	7.9	9.9	6.0	8.2	5.8
ω_1	95%	96.0	96.3	96.2	96.0	96.2	96.1	95.7	96.3
ω_1	97.5%	98.3	98.0	98.2	97.4	98.4	98.1	97.7	98.4
ω_1	99%	99.4	99.3	99.3	99.1	99.2	99.2	99.0	99.1
ω_2	1%	11.1	6.5	9.9	5.6	10.6	4.5	7.3	3.6
ω_2	2.5%	13.8	8.5	12.4	8.0	13.4	6.6	10.2	5.9
ω_2	5%	17.1	11.2	15.6	10.7	16.7	10.2	13.4	8.9
ω_2	95%	100.0	100.0	100.0	99.9	100.0	100.0	100.0	99.8
ω_2	97.5%	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
ω_2	99%	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
ω_3	1%	7.8	5.4	7.8	4.0	7.2	3.4	5.8	2.9
ω_3	2.5%	11.2	7.3	10.8	6.6	10.5	5.6	8.4	4.9
ω_3	5%	14.3	9.7	13.3	10.1	13.8	8.2	11.2	7.6
ω_3	95%	99.6	99.7	99.5	98.4	99.5	99.4	99.1	98.6
ω_3	97.5%	100.0	100.0	99.8	99.8	99.9	100.0	100.0	99.7
ω_3	99%	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
ω_4	1%	2.5	1.8	3.4	1.3	2.1	1.2	2.2	1.1
ω_4	2.5%	5.5	4.0	6.5	3.4	4.8	2.4	4.6	2.3
ω_4	5%	9.8	6.4	10.5	6.5	8.9	4.8	7.5	4.8
ω_4	95%	93.3	95.3	95.0	95.2	93.4	94.9	94.7	95.9
ω_4	97.5%	96.3	98.2	97.8	97.6	96.7	97.9	97.3	98.0
ω_4	99%	98.8	99.3	99.4	99.2	98.4	99.3	99.0	99.2

6. BOOTSTRAP PROCEDURE

The proposed empirical-likelihood-based framework enables us to use a bootstrap method to calibrate the finite-sample distribution of a statistic. As an illustration, we concentrate on the signed square root of the empirical likelihood ratio statistic, ω_{1s}, for the special case with all β_j equal to β_s.

Next we discuss how to obtain the bootstrap distribution of ω_{1s}. Recall that $\hat{p}_{si} = n^{-1}[1 + \hat{\lambda}_s\{\phi_s(x_i, \hat{\beta}_s) - \hat{\alpha}_s\}]^{-1}$, where $\hat{\beta}_s$ and $\hat{\alpha}_s$ are the maximum empirical likelihood estimators of

Table 5. *Simulated probabilities that $\omega_{1s}, \ldots, \omega_{4s}$ are smaller than 1%, 2.5%, 5%, 95%, 97.5%, and 99% quantiles of $N(0,1)$ under Scenarios S1 and S2 with $N_0 = 200$, 400.*

		Scenario S1				Scenario S2			
		$N_0 = 200$		$N_0 = 400$		$N_0 = 200$		$N_0 = 400$	
Statistic	Level	$k=2$	$k=8$	$k=2$	$k=8$	$k=2$	$k=8$	$k=2$	$k=8$
ω_{1s}	1%	2.6	4.6	3.2	3.2	2.3	4.9	2.1	3.9
ω_{1s}	2.5%	5.1	7.9	5.7	6.1	5.4	8.9	4.7	7.3
ω_{1s}	5%	9.3	13.8	9.4	11.7	9.2	13.9	8.6	12.6
ω_{1s}	95%	96.8	97.3	97.1	97.6	95.7	97.9	95.9	95.9
ω_{1s}	97.5%	98.7	99.3	98.3	98.8	98.0	99.1	97.9	98.1
ω_{1s}	99%	99.5	99.7	99.4	99.6	99.3	99.6	99.2	99.2
ω_{2s}	1%	13.1	11.3	12.1	8.6	9.8	9.9	7.9	8.0
ω_{2s}	2.5%	16.1	15.2	15.7	13.2	13.3	13.5	10.5	11.7
ω_{2s}	5%	18.5	19.0	18.8	17.3	15.9	17.4	14.0	15.4
ω_{2s}	95%	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.9
ω_{2s}	97.5%	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
ω_{2s}	99%	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
ω_{3s}	1%	9.8	9.3	9.4	7.0	7.5	8.4	5.8	7.1
ω_{3s}	2.5%	12.9	13.4	12.3	11.4	10.4	12.2	9.0	10.8
ω_{3s}	5%	15.9	17.2	15.9	15.8	14.0	16.2	12.2	14.5
ω_{3s}	95%	99.3	100.0	99.5	99.9	99.8	99.6	99.5	99.1
ω_{3s}	97.5%	99.8	100.0	99.7	100.0	100.0	100.0	100.0	99.9
ω_{3s}	99%	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
ω_{4s}	1%	5.1	3.9	5.8	3.2	2.5	2.7	2.2	2.7
ω_{4s}	2.5%	7.5	6.4	8.8	5.5	5.4	6.0	4.8	5.8
ω_{4s}	5%	12.1	10.9	12.1	10.3	9.0	10.4	8.4	10.4
ω_{4s}	95%	92.6	95.3	94.1	97.0	93.8	95.6	95.0	94.9
ω_{4s}	97.5%	94.8	98.1	96.3	98.8	96.3	98.2	97.2	97.3
ω_{4s}	99%	96.5	99.2	97.7	99.7	98.6	99.2	99.1	99.0

β_s and α, and $\hat{\lambda}_s$ is the solution to

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

Then we can estimate the cumulative distribution function $F(x)$ by

$$\hat{F}_s(x) = \sum_{i=1}^{n} \hat{p}_{si} I(x_i \le x).$$

Similarly to the consistency of $\hat{f}_w(x)$, we can show that $\hat{F}_s(x)$ is consistent with $F(x)$.

Based on $\hat{N}_s$, $\hat{\beta}_s$, and $\hat{F}_s(x)$, we propose the following bootstrap procedure to obtain the bootstrap distribution of ω_{1s}.

Step 1. Sample $X_{i,b}$ $(i = 1, \ldots, \hat{N}_s)$ from $\hat{F}_s(x)$.

Step 2. For each $X_{i,b}$, generate the number of captures $d^*_{i+,b}$ in the k occasions from $B\big(k, g(X_{i,b}, \hat{\beta}_s)\big)$. Let n_b be the number of individuals that have been captured at least once.

We use $x_{i,b}$ $(i = 1, \ldots, n_b)$ and $d_{i+,b}$ $(i = 1, \ldots, n_b)$ to denote the covariate and the number of captures for these n_b individuals.

Step 3. Based on the bootstrap sample $(x_{i,b}, d_{i+,b})$ $(i = 1, \ldots, n_b)$, calculate the maximum empirical likelihood estimator $\hat{N}_{s,b}$, the empirical likelihood ratio statistic $R'_{s,b}(\hat{N}_s)$ of the abundance N, and the signed square root of the empirical likelihood ratio statistic

$$\omega_{1s,b} = \text{sign}(\hat{N}_{s,b} - \hat{N}_s)\{R'_{s,b}(\hat{N}_s)\}^{1/2}.$$

Step 4. Repeat steps 1–3 B times and obtain $\{\omega_{1s,1}, \ldots, \omega_{1s,B}\}$. The empirical distribution of $\{\omega_{1s,1}, \ldots, \omega_{1s,B}\}$ is an accurate approximation of the bootstrap distribution of ω_{1s}.

We run a simulation under Scenarios S1 and S2 with $N_0 = 200$, 400 and $k = 8$ to check the approximation of the bootstrap distribution to the finite distribution of ω_{1s}. The simulated probabilities that ω_{1s} is smaller than the 1%, 2.5%, 5%, 95%, 97.5%, and 99% quantiles of $N(0, 1)$ and those of the bootstrap distribution are summarized in Table 6. In the calculation of the bootstrap distribution, we use $B = 999$. Clearly, the bootstrap procedure can significantly improve the coverage probabilities on both tails. It can also improve the two-tailed coverage probability. For example, under Scenario S2, the two-tailed coverage probability has been improved from 90.3% to 92.6% at the nominal level 95%. It is well known that the adjustment to improve the coverage precision is much harder for the one-sided confidence interval than for the two-sided. DiCiccio and Romano (1989) studied the correction for the signed root of the empirical likelihood ratio statistic. Further research is warranted.

Table 6. *Simulated probabilities that ω_{1s} are smaller than 1%, 2.5%, 5%, 95%, 97.5%, and 99% quantiles of $N(0, 1)$ and those of bootstrap distribution under Scenarios S1 and S2 with $N_0 = 200$, 400 and $k = 8$.*

	Scenario S1				Scenario S2			
	$N_0 = 200$		$N_0 = 400$		$N_0 = 200$		$N_0 = 400$	
Level	$N(0,1)$	Bootstrap	$N(0,1)$	Bootstrap	$N(0,1)$	Bootstrap	$N(0,1)$	Bootstrap
1%	4.6	3.6	3.2	2.6	4.9	3.5	3.9	2.6
2.5%	7.9	6.3	6.1	4.4	8.9	5.9	7.3	5.4
5%	13.8	10.2	11.7	8.4	13.9	10.2	12.6	9.2
95%	97.3	95.0	97.6	95.6	97.9	96.4	95.9	94.6
97.5%	99.3	98.0	98.8	97.9	99.1	98.5	98.1	97.1
99%	99.7	99.5	99.6	99.2	99.6	99.5	99.2	98.7

In this section, we use ω_{1s} as an illustration. The above bootstrap procedure can also be applied to obtain the bootstrap percentile confidence interval and other types of confidence intervals for N. We leave a thorough comparison of the different types of bootstrap confidence intervals for N and their asymptotic properties to future research.

REFERENCES

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

BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. & WELLNER, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Baltimore: The Johns Hopkins University Press.

DICICCIO, T. J. & ROMANO, J. P. (1989). On adjustments based on the signed root of the empirical likelihood ratio statistic. *Biometrika* **76**, 447–456.

FEWSTER, R. M. & JUPP, P. E. (2009). Inference on population size in binomial detectability models. *Biometrika* **96**, 805–820.

HJORT, N. L. & POLLARD, D. (2011). Asymptotics for minimisers of convex processes. *arXiv:1107.3806*.

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

MURTY, M. R. & SARADHA, N. (2009). Special values of the polygamma functions. *International Journal of Number Theory* **5**, 257–270.

NEWEY, W. K. (1990). Semiparametric efficiency bounds. *Journal of Applied Econometrics* **5**, 99–135.

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

OWEN, A. B. (1990). Empirical likelihood ratio confidence regions. *The Annals of Statistics* **18**, 90–120.

PARZEN, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics* **33**, 1065–1076.

QIN, J. & LAWLESS, J. (1994). Empirical likelihood and general equations. *The Annals of Statistics* **22**, 300–325.