Article

A Penalized Empirical Likelihood Approach for Estimating Population Sizes under the Negative Binomial Regression Model

Yulu Ji and Yang Liu *

School of Mathematical Sciences, Soochow University, Suzhou 215006, China; 20244507016@stu.suda.edu.cn
* Correspondence: liuyang2023@suda.edu.cn

Abstract: In capture–recapture experiments, the presence of overdispersion and heterogeneity necessitates the use of the negative binomial regression model for inferring population sizes. However, within this model, existing methods based on likelihood and ratio regression for estimating the dispersion parameter often face boundary and nonidentifiability issues. These problems can result in nonsensically large point estimates and unbounded upper limits of confidence intervals for the population size. We present a penalized empirical likelihood technique for solving these two problems by imposing a half-normal prior on the population size. Based on the proposed approach, a maximum penalized empirical likelihood estimator with asymptotic normality and a penalized empirical likelihood ratio statistic with asymptotic chi-square distribution are derived. To improve numerical performance, we present an effective expectation-maximization (EM) algorithm. In the M-step, optimization for the model parameters could be achieved by fitting a standard negative binomial regression model via the R basic function `glm.nb()`. This approach ensures the convergence and reliability of the numerical algorithm. Using simulations, we analyze several synthetic datasets to illustrate three advantages of our methods in finite-sample cases: complete mitigation of the boundary problem, more efficient maximum penalized empirical likelihood estimates, and more precise penalized empirical likelihood ratio interval estimates compared to the estimates obtained without penalty. These advantages are further demonstrated in a case study estimating the abundance of black bears (*Ursus americanus*) at the U.S. Army's Fort Drum Military Installation in northern New York.

Keywords: population size; negative binomial regression model; penalized empirical likelihood; EM algorithm

MSC: 62F10

1. Introduction

The use of capture–recapture data to infer population sizes is widely sought after across multiple fields, such as ecology and epidemiology. The capture–recapture sampling dates back to fisheries [1,2] and is used to estimate the abundance of species in the ecosystem [3]. Additionally, the number of populations of drug or alcohol addicts can be inferred by considering each visit to institutions as one capture [4,5]. Similarly, epidemiologists may leverage multiple incomplete lists of patients to estimate disease prevalence [6].

In traditional practice, the Poisson regression model is commonly used for modeling the capture–recapture data. However, the literature has demonstrated that the Poisson regression model may perform inadequately when the number of captures is subject to overdispersion, which means that the corresponding expectation is larger than variance. In capture–recapture experiments, the overdispersion often arises from individual heterogeneity. For example, animal species may exhibit flocking and spatiotemporal aggregation behaviors [7]. As argued in ref. [8], the population size may be underestimated if the model used ignores heterogeneity and overdispersion. Without any adjustments,

using the Poisson regression model can yield unreliable statistical inferences, leading to incorrect interpretations. In this situation, ref. [9] argued that the negative binomial regression model might be more suitable.

Various estimation methods have been investigated for capture–recapture data within the framework of the negative binomial regression model. Based on conditional likelihood, ref. [8] introduced a maximum likelihood estimator for the regression parameter. Additionally, they derived a Horvitz–Thompson type estimator and a corresponding Wald-type interval estimator for estimating the population size. Building on the recapture probability ratios, refs. [10–12] investigated issues related to model diagnostics and population size estimation. Their simulations and empirical investigations revealed that the negative binomial regression model outperforms the Poisson regression model in the presence of overdispersion. However, these methods face challenges, and at least two specific issues arise. First, the likelihood function may suffer from identification and boundary problems for the dispersion parameter, leading to nonsensically large values of the Horvitz–Thompson type estimator associated with large standard errors; see, for example, refs. [10,11,13]. Second, the Newton–Raphson procedure may not be reliable and even fail to converge for maximizing the conditional likelihood function, as illustrated in refs. [14,15].

Empirical likelihood offers an alternative approach to mitigating the two issues mentioned earlier. Using empirical likelihood techniques, ref. [16] investigated the full semiparametric likelihood inference for population sizes in capture–recapture studies. Additionally, this method was extended by refs. [17–20] to include the continuous time case, one inflation and missing covariates. Experience shows that the semiparametric empirical-likelihood-based method usually outperforms the conditional-likelihood-based method in numerical experiments.

However, our simulation studies indicate that the point and interval estimators derived using maximum empirical likelihood estimation are prone to boundary issues, particularly when a small proportion of individuals are captured and severe overdispersion is present. For instance, as shown in our simulation studies with $\boldsymbol{\beta}_0 = (-0.5, 0.3)^{\top}$, $k_0 = 0.5$, and $N_0 = 250$, where the capture probability is as low as 40%, nearly 67% of simulated cases yield unreasonable empirical likelihood ratio interval estimates whose upper limits exceed the number of captured individuals by over 100 times. This boundary problem is similarly observed in Section 4 for a case study. A potential cause of this issue may be that in such scenarios, the empirical log-likelihood ratio function flattens out as the population size increases.

The following sections are arranged as follows. Section 2 revisits the semiparametric empirical likelihood method within the negative binomial regression model and introduces a penalized empirical likelihood estimation approach to addressing the boundary problem. Furthermore, the EM algorithm is designed to enhance the reliability of the estimation process. Section 3 presents a number of simulations to illustrate how the proposed methods perform in the finite-sample settings. These methods are put into practice for analyzing the black bear data from New York in Section 4. Finally, a discussion is presented in Section 5.

2. Methodology

2.1. Model and Data

Suppose there are N individuals in a population. Each individual is characterized by the number of captures, denoted by Y, a nonnegative interger-valued variable. A naive approach to modeling Y is to assume a Poisson distribution, $Y \sim \text{Poisson}(\theta)$, where $\theta > 0$ represents the rate or expected number of captures. The Poisson model inherently assumes that all individuals in the population are homogeneous, meaning they share the same rate parameter θ. In addition, as noted in the introduction, a key limitation of the Poisson model is its inability to handle overdispersed data, where the variance exceeds the mean.

To address the heterogeneity and overdispersion, one can assume that individuals have varying rates. In practice, a common approach is to model the distribution of these

rates using a gamma distribution as a prior. Specifically, $\theta \sim \text{Gamma}(r, p/(1-p))$, where $r > 0$ is the shape parameter and $p/(1-p)$ is the scale parameter with $p \in (0,1)$. With this prior, the marginal probability mass function of Y can be derived in a closed form:

$$p_Y(y) = \int_0^\infty p(y \mid \theta)p(\theta)d\theta = \frac{\Gamma(y+r)}{\Gamma(y+1)\Gamma(r)}p^y(1-p)^r, \quad y = 0, 1, 2, \ldots.$$

which corresponds to the Poisson–Gamma or negative binomial distribution in the probability textbook. This distribution models the number of successes before the rth failure occurs in a sequence of independent Bernoulli trials, with p representing the probability of success in each trial.

As highlighted in ref. [9], a common reparameterization of $p_Y(y)$ is often used to interpret the counting processes in ecological and biodiversity studies. This reparameterization expresses $p_Y(y)$ in terms of its mean μ and a dispersion or aggregation parameter k, which controls the variation in counts. By setting $p = \mu/(k+\mu)$ and $k = r$, $p_Y(y)$ can be reformulated as:

$$p_Y(y) = \frac{\Gamma(y+k)}{\Gamma(y+1)\Gamma(k)}\left(\frac{\mu}{k+\mu}\right)^y\left(\frac{k}{k+\mu}\right)^k.$$

When the individual covariates, denoted by X, are available, it becomes necessary to account for the heterogeneity induced by these covariates. To do so, a parametric model is used, specifically $\mu(x; \beta) = \exp(\beta^\top x)$, which relates the mean parameter to the individual covariates. Thus, given $X = x$, the conditional probability mass function of Y is expressed by:

$$P(Y = y \mid X = x) = \frac{\Gamma(y+k)}{\Gamma(y+1)\Gamma(k)}\left\{\frac{\mu(x;\beta)}{k+\mu(x;\beta)}\right\}^y\left\{\frac{k}{k+\mu(x;\beta)}\right\}^k =: f(y, x; \beta, k), \quad (1)$$

where β represents the unknown regression coefficients and $k > 0$ represents the dispersion parameter. This formulation is referred to as the negative binomial regression model. As a special case, Equation (1) reduces to $p_Y(y)$ when all coefficients, except the intercept, are zero. The negative binomial regression model also includes the geometric regression model when $k = 1$ and reduces to the Poisson regression model as $k \to \infty$.

Given $X = x$, the conditional expectation of Y is equal to $\mu(x; \beta)$, while the conditional variance is expressed as $\mu(x; \beta) + \{\mu(x; \beta)\}^2/k$, indicating a quadratic relationship. The parameter k controls the degree of overdispersion: as k decreases, the variance increases, leading to greater overdispersion. Overdispersion is commonly observed in capture–recapture studies, where the variance significantly exceeds the mean. Consequently, the negative binomial regression model is often more appropriate for modeling capture–recapture data under conditions of severe overdispersion, as compared to the Poisson model.

Because the event $\{Y = 0\}$ is unobservable in capture–recapture studies, the zero-truncated version of model (1) is considered:

$$\begin{aligned} P(Y = y \mid Y > 0, X = x) &= \frac{P(Y = y \mid X = x)}{P(Y > 0 \mid X = x)} \\ &= \frac{f(y, x; \beta, k)}{1 - \phi(x; \beta, k)}, \quad y = 1, 2, \ldots, \end{aligned} \quad (2)$$

where $\phi(x; \beta, k) = [k/\{k + \mu(x; \beta)\}]^k$ represents the conditional probability that an individual with a covariate x is not captured at all.

Consider a study that captured n distinct individuals, with $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_n)$ denoting their individual covariates and capture frequencies, respectively. Under

the model (2), ref. [8] proposed a maximum conditional likelihood estimator $(\widetilde{\boldsymbol{\beta}}, \widetilde{k})$ by maximizing:

$$\prod_{i=1}^{n} \frac{f(y_i, \boldsymbol{x}_i; \boldsymbol{\beta}, k)}{1 - \phi(\boldsymbol{x}_i; \boldsymbol{\beta}, k)}.$$

According to the principle of inverse probability weighting, the Horvitz–Thompson type estimator of N is defined as $\widetilde{N} = \sum_{i=1}^{n} \{1 - \phi(\boldsymbol{x}_i; \widetilde{\boldsymbol{\beta}}, \widetilde{k})\}^{-1}$. However, $\widetilde{N}$ might be inflated due to small detection probabilities.

2.2. Semiparametric Empirical Likelihood

The semiparametric empirical likelihood, initially derived from ref. [16], is an appealing technique for implementing the full likelihood method when capture–recapture data contain individual covariates. Taking the negative binomial regression model as an example, we provide a brief introduction to this technique.

Considering that n distinct individuals out of a total of N individuals were captured, n follows a binomial distribution and the corresponding probability is as follows:

$$P(n) = \binom{N}{n} (1 - \alpha)^n \alpha^{N-n},$$

where $\alpha = P(Y = 0)$ represents the probability that a generic individual was not captured at all. For the given n individuals, the conditional probability of their covariates and capture counts is as follows:

$$\prod_{i=1}^{n} \frac{P(Y = y_i, \boldsymbol{X} = \boldsymbol{x}_i)}{P(Y > 0)} = \prod_{i=1}^{n} \frac{f(y_i, \boldsymbol{x}_i; \boldsymbol{\beta}, k) P(\boldsymbol{X} = \boldsymbol{x}_i)}{1 - \alpha}.$$

Multiplying these two expressions yields the full likelihood function:

$$\binom{N}{n} \alpha^{N-n} \times \prod_{i=1}^{n} \{f(y_i, \boldsymbol{x}_i; \boldsymbol{\beta}, k) P(\boldsymbol{X} = \boldsymbol{x}_i)\}. \tag{3}$$

In Equation (3), the marginal probability $P(\boldsymbol{X} = \boldsymbol{x}_i)$ is unknown and shall be addressed by the empirical likelihood method; see refs. [21,22] for more details. Technically, we assume that $P(\boldsymbol{X} = \boldsymbol{x}_i) = p_i$ for $i = 1, 2, \ldots, n$, where $p_i \in (0, 1)$ is subject to the constraint $\sum_{i=1}^{n} p_i = 1$. With this substitution, we call the full likelihood the semiparametric empirical likelihood and refer to its logarithm as the empirical log-likelihood function, namely:

$$\widetilde{\ell}(N, \boldsymbol{\beta}, k, \alpha, \{p_i\}) = \log \binom{N}{n} + (N - n) \log(\alpha) + \sum_{i=1}^{n} \log\{f(y_i, \boldsymbol{x}_i; \boldsymbol{\beta}, k)\}$$

$$+ \sum_{i=1}^{n} \log(p_i).$$

By the definition of α and the iterated expectation theorem, it follows that $\alpha = E\{P(Y = 0 \mid \boldsymbol{X})\}$, or equivalently:

$$\sum_{i=1}^{n} \{\phi(\boldsymbol{x}_i; \boldsymbol{\beta}, k) - \alpha\} p_i = 0.$$

With the constraints for p_i's, the profile empirical log-likelihood function can be derived using the Lagrange multiplier method on Equation (3):

$$\ell(N, \boldsymbol{\beta}, k, \alpha) = \log \binom{N}{n} + (N - n) \log(\alpha) + \sum_{i=1}^{n} \log \binom{y_i + k - 1}{y_i}$$
$$+ \sum_{i=1}^{n} \left[k \log \left\{ \frac{k}{k + \mu(\boldsymbol{x}_i; \boldsymbol{\beta})} \right\} + y_i \log \left\{ \frac{\mu(\boldsymbol{x}_i; \boldsymbol{\beta})}{k + \mu(\boldsymbol{x}_i; \boldsymbol{\beta})} \right\} \right]$$
$$- \sum_{i=1}^{n} \log[1 + \xi\{\phi(\boldsymbol{x}_i; \boldsymbol{\beta}, k) - \alpha\}] - n \log(n),$$

where ξ is the Lagrange multiplier, satisfying:

$$\sum_{i=1}^{n} \frac{\phi(\boldsymbol{x}_i; \boldsymbol{\beta}, k) - \alpha}{1 + \xi\{\phi(\boldsymbol{x}_i; \boldsymbol{\beta}, k) - \alpha\}} = 0.$$

Notice that there are a finite number of unknown parameters in the profile empirical log-likelihood function. By maximizing this function, we obtain the maximum empirical likelihood estimator, expressed as $(\widehat{N}, \widehat{\boldsymbol{\beta}}, \widehat{k}, \widehat{\alpha}) = \arg\max\{\ell(N, \boldsymbol{\beta}, k, \alpha)\}$.

2.3. Penalized Empirical Likelihood Inference

When the number of captures exhibits severe overdispersion, both the estimators $\widetilde{N}$ and $\widehat{N}$, proposed in Sections 2.1 and 2.2, respectively, may exhibit spuriously large values, potentially leading to misleading conclusions. This issue has been addressed in ref. [13] (p. 84) for the Horvitz–Thompson type estimator. Our simulation studies further confirm that the empirical likelihood estimators may also suffer from the boundary problem. This issue may arise due to the limited information available about the population size, causing the profile empirical log-likelihood to fail in distinguishing between different values of large N.

To mitigate this problem, we intuitively incorporate prior information on the population size to reduce the probability of large values. We achieve this by augmenting the likelihood functions with an appropriate penalty term. Correspondingly, the penalized empirical log-likelihood function and its profile version are defined as:

$$\widetilde{\ell}_p(N, \boldsymbol{\beta}, k, \alpha, \{p_i\}) = \widetilde{\ell}(N, \boldsymbol{\beta}, k, \alpha, \{p_i\}) + f_p(N),$$
$$\ell_p(N, \boldsymbol{\beta}, k, \alpha) = \ell(N, \boldsymbol{\beta}, k, \alpha) + f_p(N). \tag{4}$$

where the penalty term $f_p(N)$ takes the form of $-C(N - \nu)^2 I(N > \nu)$, where ν is a lower bound of N, $C \geq 0$ is a tuning parameter, and $I(\cdot)$ is the indicator function. For specific values of ν and C, a maximum penalized empirical likelihood estimator is proposed, namely, $(\widehat{N}_p, \widehat{\boldsymbol{\beta}}_p, \widehat{k}_p, \widehat{\alpha}_p) = \arg\max\{\ell_p(N, \boldsymbol{\beta}, k, \alpha)\}$.

From a Bayesian perspective, adding the penalty term $f_p(N)$ into the log-likelihood is equivalent to imposing a prior for N that has a mixture of the half-normal distribution $\mathrm{N}(\nu, (2C)^{-1})$ for $N > \nu$ and a uniform distribution $\mathrm{U}(n, \nu)$ for $n \leq N \leq \nu$. In other words, this penalty has no effect on the likelihood when $n \leq N \leq \nu$ and gradually decreases the likelihood when $N > \nu$. The larger the population size, the more pronounced the decrease. Consequently, the penalized method encourages large values of $\widehat{N}$ to shrink towards ν. In practice, we recommend using the Chao estimator as the lower bound ν; see ref. [23] for details about this estimator. Alternatively, the generalized Chao estimator proposed in ref. [24] can also be considered.

To derive the large-sample properties of the estimator $(\widehat{N}_p, \widehat{\boldsymbol{\beta}}_p, \widehat{k}_p, \widehat{\alpha}_p)$, we define some notation when the parameter vector $(N, \boldsymbol{\beta}, k, \alpha)$ takes its true value, namely, $(N_0, \boldsymbol{\beta}_0, k_0, \alpha_0)$. Define $\psi(\boldsymbol{x}; \boldsymbol{\beta}_0, k_0) = -\log\{1 + \mu(\boldsymbol{x}; \boldsymbol{\beta}_0)/k_0\} + \{1 + k_0/\mu(\boldsymbol{x}; \boldsymbol{\beta}_0)\}^{-1}$, $\varphi = \mathbb{E}\{1 - \phi(\boldsymbol{X}; \boldsymbol{\beta}_0, k_0)\}^{-1}$, and:

$$S_2(y + k_0 - 1, y) = \begin{cases} -\sum_{k=k_0}^{y+k_0-1} \frac{1}{k^2}, & y = 1, 2, \dots \\ 0, & y = 0 \end{cases}.$$

Let:

$$\mathbf{W} = \begin{bmatrix} -V_{11} & \mathbf{0}_{s\times 1}^{\top} & 0 & -V_{14} \\ \mathbf{0}_{s\times 1} & -\mathbf{V}_{22} + \mathbf{V}_{25}V_{55}^{-1}\mathbf{V}_{52} & -\mathbf{V}_{23} + \mathbf{V}_{25}V_{55}^{-1}V_{53} & -\mathbf{V}_{24} + \mathbf{V}_{25}V_{55}^{-1}V_{54} \\ 0 & -\mathbf{V}_{32} + \mathbf{V}_{35}V_{55}^{-1}\mathbf{V}_{52} & -\mathbf{V}_{33} + \mathbf{V}_{35}V_{55}^{-1}V_{53} & -\mathbf{V}_{34} + \mathbf{V}_{35}V_{55}^{-1}V_{54} \\ -V_{41} & -\mathbf{V}_{42} + \mathbf{V}_{45}V_{55}^{-1}\mathbf{V}_{52} & -\mathbf{V}_{43} + \mathbf{V}_{45}V_{55}^{-1}V_{53} & -\mathbf{V}_{44} + \mathbf{V}_{45}V_{55}^{-1}V_{54} \end{bmatrix}, \tag{5}$$

where $V_{11} = 1 - \alpha_0^{-1}$, $V_{41} = V_{14} = \alpha_0^{-1}$, $V_{44} = \varphi - \alpha_0^{-1}$, $V_{45} = V_{54} = (1-\alpha_0)^2\varphi$, $V_{55} = (1-\alpha_0)^4\varphi - (1-\alpha_0)^3$, and:

$$\begin{aligned} \mathbf{V}_{22} &= \mathbb{E}\left[\left\{\frac{k_0\phi(\mathbf{X};\boldsymbol{\beta}_0,k_0)\mu(\mathbf{X};\boldsymbol{\beta}_0)}{1-\phi(\mathbf{X};\boldsymbol{\beta}_0,k_0)} - k_0 - \mu(\mathbf{X};\boldsymbol{\beta}_0,k_0)\right\}\frac{k_0\mu(\mathbf{X};\boldsymbol{\beta}_0)\mathbf{X}^{\otimes 2}}{\{k_0 + \mu(\mathbf{X};\boldsymbol{\beta}_0)\}^2}\right], \\ \mathbf{V}_{23} &= \mathbf{V}_{32}^{\top} = -\mathbb{E}\left[\frac{k_0\phi(\mathbf{X};\boldsymbol{\beta}_0,k_0)\psi(\mathbf{X};\boldsymbol{\beta}_0,k_0)\mu(\mathbf{X};\boldsymbol{\beta}_0)\mathbf{X}}{\{1-\phi(\mathbf{X};\boldsymbol{\beta}_0,k_0)\}\{k_0 + \mu(\mathbf{X};\boldsymbol{\beta}_0)\}}\right], \\ V_{33} &= \mathbb{E}\left[\frac{\phi(\mathbf{X};\boldsymbol{\beta}_0,k_0)\psi^2(\mathbf{X};\boldsymbol{\beta}_0,k_0)}{1-\phi(\mathbf{X};\boldsymbol{\beta}_0,k_0)} + \frac{k_0\mu(\mathbf{X};\boldsymbol{\beta}_0) + \mu^2(\mathbf{X};\boldsymbol{\beta}_0)}{k_0\{k_0 + \mu(\mathbf{X};\boldsymbol{\beta}_0)\}^2}\right] \\ &\quad + \mathbb{E}\{S_2(Y + k_0 - 1, Y)\}, \\ \mathbf{V}_{24} &= \mathbb{E}\left[\frac{\phi(\mathbf{X};\boldsymbol{\beta}_0,k_0)}{1-\phi(\mathbf{X};\boldsymbol{\beta}_0,k_0)} \cdot \frac{k_0\mu(\mathbf{X};\boldsymbol{\beta}_0)\mathbf{X}}{k_0 + \mu(\mathbf{X};\boldsymbol{\beta}_0)}\right], \quad \mathbf{V}_{25} = \mathbf{V}_{52}^{\top} = (1-\alpha_0)^2\mathbf{V}_{24}, \\ V_{34} &= V_{43} = -\mathbb{E}\left\{\frac{\phi(\mathbf{X};\boldsymbol{\beta}_0,k_0)\psi(\mathbf{X};\boldsymbol{\beta}_0,k_0)}{1-\phi(\mathbf{X};\boldsymbol{\beta}_0,k_0)}\right\}, \quad V_{35} = V_{53} = (1-\alpha_0)^2V_{34}. \end{aligned}$$

The following theorem presents the large-sample properties of the maximum penalized empirical likelihood estimator $(\widehat{N}_p, \widehat{\boldsymbol{\beta}}_p, \widehat{k}_p, \widehat{\alpha}_p)$ associated with the penalized empirical likelihood ratio statistic of N.

Theorem 1. *Suppose that $k_0 > 0$, $0 < \alpha_0 < 1$, and the tuning parameter satisfies $C = O(N^{-2})$. If $\mathbf{W}$ is nonsingular and $\varphi < \infty$, as $N_0 \to \infty$:*

(a) $\sqrt{N_0}\{\log(\widehat{N}_p) - \log(N_0), (\widehat{\boldsymbol{\beta}}_p - \boldsymbol{\beta}_0)^{\top}, \widehat{k}_p - k_0, \widehat{\alpha}_p - \alpha_0\}^{\top} \xrightarrow{d} \mathrm{N}(\mathbf{0}, \mathbf{W}^{-1})$;

(b) $\sqrt{N_0}(\widehat{N}_p / N_0 - 1) \xrightarrow{d} N(0, \sigma^2)$, *where*

$$\sigma^2 = \varphi - 1 - \begin{bmatrix} \mathbf{V}_{42} & V_{43} \end{bmatrix}\begin{bmatrix} \mathbf{V}_{22} & \mathbf{V}_{23} \\ \mathbf{V}_{32} & V_{33} \end{bmatrix}^{-1}\begin{bmatrix} \mathbf{V}_{24} \\ V_{34} \end{bmatrix};$$

(c) $2\{\ell_p(\widehat{N}_p, \widehat{\boldsymbol{\beta}}_p, \widehat{k}_p, \widehat{\alpha}_p) - \max_{(\boldsymbol{\beta},k,\alpha)} \ell_p(N, \boldsymbol{\beta}, k, \alpha)\} \xrightarrow{d} \chi_1^2$, *where χ_1^2 denotes the standard chi-square distribution.*

Proof. As the proposed semiparametric empirical likelihood approach can be seen as an extension of the EL method of ref. [16] to the negative binomial regression model, the proof of Theorem 1 is very similar to those of Theorem 1 and Corollary 1 in ref. [16]. Here, we only highlight the difference and the formulae of $\mathbf{V}$ and $\boldsymbol{\Sigma}$ in our framework.

We first argue that $\xi_0 = -(1-\alpha_0)^{-1}$ is the limit of $\widehat{\xi}$, where $\widehat{\xi}$ is the solution to:

$$\sum_{i=1}^{n} \frac{\phi(\mathbf{x}_i; \widehat{\boldsymbol{\beta}}_p, \widehat{k}_p) - \widehat{\alpha}_p}{1 + \xi\{\phi(\mathbf{x}_i; \widehat{\boldsymbol{\beta}}_p, \widehat{k}_p) - \widehat{\alpha}_p\}} = 0.$$

For this purpose, we define a function:

$$\hbar(N, \boldsymbol{\beta}, k, \alpha, \xi) = \log \binom{N}{n} + (N-n)\log(\alpha) - \sum_{i=1}^{n} \log[1 + \xi\{\phi(\boldsymbol{x}_i; \boldsymbol{\beta}, k) - \alpha\}]$$
$$+ \sum_{i=1}^{n} \left[\log \binom{y_i + k - 1}{y_i} + k\log\left\{ \frac{k}{k + \mu(\boldsymbol{x}_i; \boldsymbol{\beta})} \right\} + y_i \log\left\{ \frac{\mu(\boldsymbol{x}_i; \boldsymbol{\beta})}{k + \mu(\boldsymbol{x}_i; \boldsymbol{\beta})} \right\} \right]$$
$$- n\log(n) + f_p(N).$$

It can be seen that $\ell_p(N, \boldsymbol{\beta}, k, \alpha) = \hbar(N, \boldsymbol{\beta}, k, \alpha, \xi_*)$, where ξ_* is the solution to $\partial\hbar/\partial\xi = 0$. The first partial derivatives of $\hbar(N, \boldsymbol{\beta}, k, \alpha, \xi)$ with respect to α and ξ are:

$$\frac{\partial\hbar}{\partial\alpha} = \frac{N-n}{\alpha} + \xi \sum_{i=1}^{n} \frac{1}{1 + \xi\{\phi(\boldsymbol{x}_i; \boldsymbol{\beta}, k) - \alpha\}},$$
$$\frac{\partial\hbar}{\partial\xi} = -\sum_{i=1}^{n} \frac{\phi(\boldsymbol{x}_i; \boldsymbol{\beta}, k) - \alpha}{1 + \xi\{\phi(\boldsymbol{x}_i; \boldsymbol{\beta}, k) - \alpha\}}.$$

Setting the above equations to zero gives $\widehat{\xi} = -(\widehat{N}_p - n)/(n\widehat{\alpha}_p)$. Since α_0 is the probability of never being captured, n follows a Binomial distribution $\mathrm{Bi}(N_0, 1 - \alpha_0)$. When $(\widehat{N}_p, \widehat{\alpha}_p)$ is consistent, it follows from the strong law of large numbers that as $N_0 \to \infty$:

$$\widehat{\xi} = -\frac{\widehat{N}/N_0 - n/N_0}{(n/N_0)\widehat{\alpha}_p} \xrightarrow{p} -\frac{1 - (1 - \alpha_0)}{(1 - \alpha_0)\alpha_0} = -\frac{1}{1 - \alpha_0} = \xi_0,$$

where $\xrightarrow{p}$ denotes convergence in probability.

Below, we derive the formulae of $\mathbf{V}$ and $\boldsymbol{\Sigma}$. Let $\boldsymbol{\gamma} = (\gamma_1, \boldsymbol{\gamma}_2^\top, \gamma_3, \gamma_4, \gamma_5)^\top$, with $\gamma_1 = \sqrt{N_0}(N/N_0 - 1)$, $\boldsymbol{\gamma}_2 = \sqrt{N_0}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)$, $\gamma_3 = \sqrt{N_0}(k - k_0)$, $\gamma_4 = \sqrt{N_0}(\alpha - \alpha_0)$, $\gamma_5 = \sqrt{N_0}(\xi - \xi_0)$. Define

$$\mathcal{H}(\boldsymbol{\gamma}) = \hbar(N_0 + N_0^{1/2}\gamma_1, \boldsymbol{\beta}_0 + N_0^{-1/2}\boldsymbol{\gamma}_2, k_0 + N_0^{-1/2}\gamma_3, \alpha_0 + N_0^{-1/2}\gamma_4, \xi_0 + N_0^{-1/2}\gamma_5).$$

According to Lemma 2 in the Supplementary Material of [16], deriving the formula of $\mathbf{V}$ is equivalent to calculating the first two partial derivatives of $\mathcal{H}(\boldsymbol{\gamma})$ with respect to $\boldsymbol{\gamma}$ at $\mathbf{0}$. It follows from the law of large numbers and the central limit theorem that:

$$\frac{\partial\mathcal{H}(\mathbf{0})}{\partial\gamma_1} = N_0^{1/2}\{S_1(N_0, n) + \log(\alpha_0) + f_p'(N_0)\} = N_0^{1/2}\left(\frac{n/N_0 - 1}{\alpha_0} + 1\right) + O_p(N_0^{-1/2}),$$
$$\frac{\partial\mathcal{H}(\mathbf{0})}{\partial\boldsymbol{\gamma}_2} = N_0^{-1/2} \sum_{i=1}^{n} \left[\left\{ -\frac{\mu(\boldsymbol{x}_i; \boldsymbol{\beta}_0)}{1 - \phi(\boldsymbol{x}_i; \boldsymbol{\beta}_0, k_0)} + y_i \right\} \frac{k_0 \boldsymbol{x}_i}{k_0 + \mu(\boldsymbol{x}_i; \boldsymbol{\beta}_0)} \right],$$
$$\frac{\partial\mathcal{H}(\mathbf{0})}{\partial\gamma_3} = N_0^{-1/2} \sum_{i=1}^{n} \left\{ \frac{\psi(\boldsymbol{x}; \boldsymbol{\beta}_0, k_0)}{1 - \phi(\boldsymbol{x}_i; \boldsymbol{\beta}_0, k_0)} - \frac{y_i}{k_0 + \mu(\boldsymbol{x}_i; \boldsymbol{\beta}_0)} + S_1(y_i + k_0 - 1, y_i) \right\},$$
$$\frac{\partial\mathcal{H}(\mathbf{0})}{\partial\gamma_4} = N_0^{-1/2} \left\{ \frac{N_0 - n}{\alpha_0} - \sum_{i=1}^{n} \frac{1}{1 - \phi(\boldsymbol{x}_i; \boldsymbol{\beta}_0, k_0)} \right\},$$
$$\frac{\partial\mathcal{H}(\mathbf{0})}{\partial\gamma_5} = -N_0^{-1/2}(1 - \alpha_0) \sum_{i=1}^{n} \frac{\phi(\boldsymbol{x}_i; \boldsymbol{\beta}_0, k_0) - \alpha_0}{1 - \phi(\boldsymbol{x}_i; \boldsymbol{\beta}_0, k_0)},$$

where the first equation uses the result (a) of Lemma A1 and Equation (A1) of Lemma A2 in the Appendix A. Using the result (b) of Lemma A1 and Equation (A2) of Lemma A2, we have:

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

Similarly, it can be verified that $\partial^2 \mathcal{H}(\mathbf{0})/(\partial \gamma_4 \partial \gamma_1) = \partial^2 \mathcal{H}(\mathbf{0})/(\partial \gamma_1 \partial \gamma_4) = \alpha_0^{-1} + O_p(N_0^{-1/2})$ and:

$$\frac{\partial^2 \mathcal{H}(\mathbf{0})}{\partial \gamma_2 \partial \gamma_2^\top} = N_0^{-1} \sum_{i=1}^{n} \left(\left[\frac{k_0 \phi(\mathbf{x}_i; \boldsymbol{\beta}_0, k_0) \mu(\mathbf{x}_i; \boldsymbol{\beta}_0)}{\{1 - \phi(\mathbf{x}_i; \boldsymbol{\beta}_0, k_0)\}^2} - \frac{k_0}{1 - \phi(\mathbf{x}_i; \boldsymbol{\beta}_0, k_0)} - y_i \right] \frac{k_0 \mu(\mathbf{x}_i; \boldsymbol{\beta}_0) \mathbf{x}_i^{\otimes 2}}{\{k_0 + \mu(\mathbf{x}_i; \boldsymbol{\beta}_0)\}^2} \right),$$

$$\frac{\partial^2 \mathcal{H}(\mathbf{0})}{\partial \gamma_2 \partial \gamma_3} = \left\{ \frac{\partial^2 \mathcal{H}(\mathbf{0})}{\partial \gamma_3 \partial \gamma_2^\top} \right\}^\top = N_0^{-1} \sum_{i=1}^{n} \left[- \frac{k_0 \phi(\mathbf{x}_i; \boldsymbol{\beta}_0, k_0) \psi(\mathbf{x}_i; \boldsymbol{\beta}_0, k_0)}{\{1 - \phi(\mathbf{x}_i; \boldsymbol{\beta}_0, k_0)\}^2} + \frac{y_i}{k_0 + \mu(\mathbf{x}_i; \boldsymbol{\beta}_0)} \right.$$

$$\left. - \frac{1}{1 - \phi(\mathbf{x}_i; \boldsymbol{\beta}_0, k_0)} \cdot \frac{\mu(\mathbf{x}_i; \boldsymbol{\beta}_0)}{k_0 + \mu(\mathbf{x}_i; \boldsymbol{\beta}_0)} \right] \frac{\mu(\mathbf{x}_i; \boldsymbol{\beta}_0) \mathbf{x}_i}{k_0 + \mu(\mathbf{x}_i; \boldsymbol{\beta}_0)},$$

$$\frac{\partial^2 \mathcal{H}(\mathbf{0})}{\partial \gamma_2 \partial \gamma_4} = \left\{ \frac{\partial^2 \mathcal{H}(\mathbf{0})}{\partial \gamma_4 \partial \gamma_2^\top} \right\}^\top = N_0^{-1} \sum_{i=1}^{n} \frac{\phi(\mathbf{x}_i; \boldsymbol{\beta}_0, k_0)}{\{1 - \phi(\mathbf{x}_i; \boldsymbol{\beta}_0, k_0)\}^2} \cdot \frac{k_0 \mu(\mathbf{x}_i; \boldsymbol{\beta}_0) \mathbf{x}_i}{k_0 + \mu(\mathbf{x}_i; \boldsymbol{\beta}_0)},$$

$$\frac{\partial^2 \mathcal{H}(\mathbf{0})}{\partial \gamma_2 \partial \gamma_5} = \left\{ \frac{\partial^2 \mathcal{H}(\mathbf{0})}{\partial \gamma_5 \partial \gamma_2^\top} \right\}^\top = (1 - \alpha_0)^2 N_0^{-1} \sum_{i=1}^{n} \frac{\phi(\mathbf{x}_i; \boldsymbol{\beta}_0, k_0)}{\{1 - \phi(\mathbf{x}_i; \boldsymbol{\beta}_0, k_0)\}^2} \cdot \frac{k_0 \mu(\mathbf{x}_i; \boldsymbol{\beta}_0) \mathbf{x}_i}{k_0 + \mu(\mathbf{x}_i; \boldsymbol{\beta}_0)},$$

$$\frac{\partial^2 \mathcal{H}(\mathbf{0})}{\partial \gamma_3^2} = N_0^{-1} \sum_{i=1}^{n} \left[\frac{\phi(\mathbf{x}_i; \boldsymbol{\beta}_0, k_0) \psi^2(\mathbf{x}_i; \boldsymbol{\beta}_0, k_0)}{\{1 - \phi(\mathbf{x}_i; \boldsymbol{\beta}_0, k_0)\}^2} + \frac{1}{1 - \phi(\mathbf{x}_i; \boldsymbol{\beta}_0, k_0)} \cdot \frac{\mu^2(\mathbf{x}_i; \boldsymbol{\beta}_0)}{k_0 \{k_0 + \mu(\mathbf{x}_i; \boldsymbol{\beta}_0)\}^2} \right.$$

$$\left. + \frac{y_i}{\{k_0 + \mu(\mathbf{x}_i; \boldsymbol{\beta}_0)\}^2} + S_2(y_i + k_0 - 1, y_i) \right],$$

$$\frac{\partial^2 \mathcal{H}(\mathbf{0})}{\partial \gamma_3 \partial \gamma_4} = \frac{\partial^2 \mathcal{H}(\mathbf{0})}{\partial \gamma_4 \partial \gamma_3} = -N_0^{-1} \sum_{i=1}^{n} \frac{\phi(\mathbf{x}_i; \boldsymbol{\beta}_0, k_0) \psi(\mathbf{x}_i; \boldsymbol{\beta}_0, k_0)}{\{1 - \phi(\mathbf{x}_i; \boldsymbol{\beta}_0, k_0)\}^2},$$

$$\frac{\partial^2 \mathcal{H}(\mathbf{0})}{\partial \gamma_3 \partial \gamma_5} = \frac{\partial^2 \mathcal{H}(\mathbf{0})}{\partial \gamma_5 \partial \gamma_3} = -(1 - \alpha_0)^2 N_0^{-1} \sum_{i=1}^{n} \frac{\phi(\mathbf{x}_i; \boldsymbol{\beta}_0, k_0) \psi(\mathbf{x}_i; \boldsymbol{\beta}_0, k_0)}{\{1 - \phi(\mathbf{x}_i; \boldsymbol{\beta}_0, k_0)\}^2},$$

$$\frac{\partial^2 \mathcal{H}(\mathbf{0})}{\partial \gamma_4^2} = N_0^{-1} \sum_{i=1}^{n} \frac{1}{\{1 - \phi(\mathbf{x}_i; \boldsymbol{\beta}_0, k_0)\}^2} - \frac{1 - n/N_0}{\alpha_0^2},$$

$$\frac{\partial^2 \mathcal{H}(\mathbf{0})}{\partial \gamma_4 \partial \gamma_5} = \frac{\partial^2 \mathcal{H}(\mathbf{0})}{\partial \gamma_5 \partial \gamma_4} = (1 - \alpha_0)^2 N_0^{-1} \sum_{i=1}^{n} \frac{1}{\{1 - \phi(\mathbf{x}_i; \boldsymbol{\beta}_0, k_0)\}^2},$$

$$\frac{\partial^2 \mathcal{H}(\mathbf{0})}{\partial \gamma_5^2} = (1 - \alpha_0)^4 N_0^{-1} \sum_{i=1}^{n} \frac{1}{\{1 - \phi(\mathbf{x}_i; \boldsymbol{\beta}_0, k_0)\}^2} - (1 - \alpha_0)^3.$$

With the arguments of Lemma A3 in the Appendix A, we have that the leading term of $\partial^2 \mathcal{H}(\mathbf{0})/(\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^\top)$ is as follows:

$$\frac{\partial^2 \mathcal{H}(\mathbf{0})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^\top} = \mathbf{V} + O_p(N_0^{-1/2}), \quad \mathbf{V} = \begin{bmatrix} V_{11} & \mathbf{0}^\top & 0 & V_{14} & 0 \\ \mathbf{0} & \mathbf{V}_{22} & \mathbf{V}_{23} & \mathbf{V}_{24} & \mathbf{V}_{25} \\ 0 & \mathbf{V}_{32} & V_{33} & V_{34} & V_{35} \\ V_{41} & \mathbf{V}_{42} & V_{43} & V_{44} & V_{45} \\ 0 & \mathbf{V}_{52} & V_{53} & V_{54} & V_{55} \end{bmatrix},$$

where V_{ij}s are the same as those defined in Equation (5).

Next, we derive the formula of $\boldsymbol{\Sigma}$. Define $\mathbf{u}_n = (u_{n1}, \mathbf{u}_{n2}^\top, u_{n3}, u_{n4}, u_{n5})^\top$, where:

$$u_{n1} = N_0^{1/2} \left(\frac{n/N_0 - 1}{\alpha_0} + 1 \right), \quad \mathbf{u}_{n2} = \frac{\partial \mathcal{H}(\mathbf{0})}{\partial \gamma_2}, \quad u_{n3} = \frac{\partial \mathcal{H}(\mathbf{0})}{\partial \gamma_3}, \quad u_{n4} = \frac{\partial \mathcal{H}(\mathbf{0})}{\partial \gamma_4}, \quad u_{n5} = \frac{\partial \mathcal{H}(\mathbf{0})}{\partial \gamma_5}.$$

It can be verified that $\partial \mathcal{H}(\mathbf{0}) / \partial \boldsymbol{\gamma} = \mathbf{u}_n + O_p(N_0^{-1/2})$ and $\mathbb{E}(\mathbf{u}_n) = \mathbf{0}$. Here, we only verify that $\mathbb{E}(\mathbf{u}_{n2}) = \mathbf{0}$ and $\mathbb{E}(u_{n3}) = 0$. In fact, it follows Lemma A3 that:

$$
\begin{aligned}
\mathbb{E}(\mathbf{u}_{n2}) &= N_0^{1/2} \mathbb{E}\left\{ -\frac{k_0 \boldsymbol{X} \mu(\boldsymbol{X}; \boldsymbol{\beta}_0)}{k_0 + \mu(\boldsymbol{X}; \boldsymbol{\beta}_0)} + \frac{k_0 \boldsymbol{X} \mathbb{E}(Y \mid \boldsymbol{X})}{k_0 + \mu(\boldsymbol{X}; \boldsymbol{\beta}_0)} \right\} = \mathbf{0}, \\
\mathbb{E}(u_{n3}) &= N_0^{1/2} \mathbb{E}\left[\psi(\boldsymbol{X}; \boldsymbol{\beta}_0, k_0) - \frac{\mathbb{E}(Y \mid \boldsymbol{X})}{k_0 + \mu(\boldsymbol{X}; \boldsymbol{\beta}_0)} + \mathbb{E}\{S_1(Y + k_0 - 1, Y) \mid \boldsymbol{X}\} \right] \\
&= N_0^{1/2} \mathbb{E}\left[\psi(\boldsymbol{X}; \boldsymbol{\beta}_0, k_0) - \frac{\mu(\boldsymbol{X}; \boldsymbol{\beta}_0)}{k_0 + \mu(\boldsymbol{X}; \boldsymbol{\beta}_0)} - \log\left\{ \frac{k_0}{k_0 + \mu(\boldsymbol{X}; \boldsymbol{\beta}_0)} \right\} \right] = 0,
\end{aligned}
$$

where the last second equation uses Equation (A3) of Lemma A4 in the Appendix A.

For the covariance matrix of $\mathbf{u}_n$, it follows Lemma A4 that:

$$
\begin{aligned}
\mathbb{V}\mathrm{ar}(u_{n1}) &= \alpha_0^{-1} - 1, \quad \mathbb{C}\mathrm{ov}(u_{n4}, u_{n1}) = -\alpha_0^{-1}, \\
\mathbb{C}\mathrm{ov}(\mathbf{u}_{n2}, u_{n1}) &= \mathbb{C}\mathrm{ov}(\mathbf{u}_{n2}, u_{n4}) = \mathbb{C}\mathrm{ov}(\mathbf{u}_{n2}, u_{n5}) = \mathbf{0}, \\
\mathbb{C}\mathrm{ov}(u_{n3}, u_{n1}) &= \mathbb{C}\mathrm{ov}(u_{n3}, u_{n4}) = \mathbb{C}\mathrm{ov}(u_{n3}, u_{n5}) = \mathbb{C}\mathrm{ov}(u_{n5}, u_{n1}) = 0, \\
\mathbb{V}\mathrm{ar}(\mathbf{u}_{n2}) &= \mathbb{E}\left[\frac{1}{N_0} \sum_{i=1}^{N_0} \left\{ -\frac{\mu(\boldsymbol{X}_i; \boldsymbol{\beta}_0)}{1 - \phi(\boldsymbol{X}_i; \boldsymbol{\beta}_0, k_0)} + Y_i \right\} \frac{k_0 \boldsymbol{X}_i I(Y_i > 0)}{k_0 + \mu(\boldsymbol{X}_i; \boldsymbol{\beta}_0)} \right]^{\otimes 2} \\
&= \mathbb{E}\left[\left\{ -\frac{\mu(\boldsymbol{X}; \boldsymbol{\beta}_0)}{1 - \phi(\boldsymbol{X}; \boldsymbol{\beta}_0, k_0)} + Y \right\}^2 \frac{k_0^2 I(Y > 0) \boldsymbol{X}^{\otimes 2}}{\{k_0 + \mu(\boldsymbol{X}; \boldsymbol{\beta}_0)\}^2} \right] \\
&= \mathbb{E}\left[\left\{ \frac{\mu^2(\boldsymbol{X}; \boldsymbol{\beta}_0)}{\{1 - \phi(\boldsymbol{X}; \boldsymbol{\beta}_0, k_0)\}^2} + Y^2 - \frac{2Y\mu(\boldsymbol{X}; \boldsymbol{\beta}_0)}{1 - \phi(\boldsymbol{X}; \boldsymbol{\beta}_0, k_0)} \right\} \frac{k_0^2 I(Y > 0) \boldsymbol{X}^{\otimes 2}}{\{k_0 + \mu(\boldsymbol{X}; \boldsymbol{\beta}_0)\}^2} \right] \\
&= \mathbb{E}\left[\left\{ -\frac{k_0 \mu(\boldsymbol{X}; \boldsymbol{\beta}_0) \phi(\boldsymbol{X}; \boldsymbol{\beta}_0, k_0)}{1 - \phi(\boldsymbol{X}; \boldsymbol{\beta}_0, k_0)} + k_0 + \mu(\boldsymbol{X}; \boldsymbol{\beta}_0) \right\} \frac{k_0 \mu(\boldsymbol{X}; \boldsymbol{\beta}_0) \boldsymbol{X}^{\otimes 2}}{\{k_0 + \mu(\boldsymbol{X}; \boldsymbol{\beta}_0)\}^2} \right] \\
&= -\mathbf{V}_{22}, \\
\mathbb{C}\mathrm{ov}(\mathbf{u}_{n2}, u_{n3}) &= \mathbb{E}\left[\left\{ -\frac{\mu(\boldsymbol{X}; \boldsymbol{\beta}_0)}{1 - \phi(\boldsymbol{X}; \boldsymbol{\beta}_0, k_0)} + Y \right\} \left\{ \frac{\psi(\boldsymbol{x}; \boldsymbol{\beta}_0, k_0)}{1 - \phi(\boldsymbol{X}; \boldsymbol{\beta}_0, k_0)} - \frac{Y}{k_0 + \mu(\boldsymbol{X}; \boldsymbol{\beta}_0)} \right. \right. \\
&\qquad \left. \left. + S_1(Y + k_0 - 1, Y) \right\} \cdot \frac{k_0 \boldsymbol{X} I(Y > 0)}{k_0 + \mu(\boldsymbol{X}; \boldsymbol{\beta}_0)} \right] \\
&= \mathbb{E}\left[\left\{ \frac{\phi(\boldsymbol{X}; \boldsymbol{\beta}_0, k_0) \psi(\boldsymbol{x}; \boldsymbol{\beta}_0, k_0)}{1 - \phi(\boldsymbol{x}; \boldsymbol{\beta}_0, k_0)} - \frac{1}{k_0} + \frac{\mathbb{E}[YS_1(Y + k_0 - 1, Y) \mid \boldsymbol{X}]}{\mu(\boldsymbol{X}; \boldsymbol{\beta}_0)} \right. \right. \\
&\qquad \left. \left. + \log \frac{k_0}{k_0 + \mu(\boldsymbol{X}; \boldsymbol{\beta}_0)} \right\} \frac{k_0 \mu(\boldsymbol{X}; \boldsymbol{\beta}_0) \boldsymbol{X}}{k_0 + \mu(\boldsymbol{X}; \boldsymbol{\beta}_0)} \right] = -\mathbf{V}_{23}, \\
\mathbb{V}\mathrm{ar}(u_{n3}) &= \mathbb{E}\left[\left\{ \frac{\psi(\boldsymbol{X}; \boldsymbol{\beta}_0, k_0)}{1 - \phi(\boldsymbol{X}; \boldsymbol{\beta}_0, k_0)} - \frac{Y}{k_0 + \mu(\boldsymbol{X}; \boldsymbol{\beta}_0)} + S_1(Y + k_0 - 1, Y) \right\}^2 I(Y > 0) \right] \\
&= \mathbb{E}\left[-\frac{\{\psi(\boldsymbol{X}; \boldsymbol{\beta}_0, k_0)\}^2}{1 - \phi(\boldsymbol{X}; \boldsymbol{\beta}_0, k_0)} + \frac{(k_0 + 1)\{\mu(\boldsymbol{X}; \boldsymbol{\beta}_0, k_0)\}^2 + k_0 \mu(\boldsymbol{X}; \boldsymbol{\beta}_0, k_0)}{k_0 \{k_0 + \mu(\boldsymbol{X}; \boldsymbol{\beta}_0, k_0)\}^2} \right. \\
&\qquad \left. + \mathbb{E}\left[\{S_1(Y + k - 1, Y)\}^2 \mid \boldsymbol{X} \right] - \frac{2\mathbb{E}\{YS_1(Y + k - 1, Y) \mid \boldsymbol{X}\}}{k + \mu(\boldsymbol{X}; \boldsymbol{\beta}_0, k_0)} \right] \\
&= -\mathbf{V}_{33}, \\
\mathbb{V}\mathrm{ar}(u_{n4}) &= N_0^{-1} \mathbb{E}\left[\sum_{i=1}^{N_0} \left\{ \frac{I(Y_i = 0)}{\alpha_0} - \frac{I(Y_i > 0)}{1 - \phi(\boldsymbol{X}_i; \boldsymbol{\beta}_0, k_0)} \right\} \right]^2 \\
&= \mathbb{E}\left\{ \frac{I(Y = 0)}{\alpha_0} - \frac{I(Y > 0)}{1 - \phi(\boldsymbol{X}; \boldsymbol{\beta}_0, k_0)} \right\}^2 = \alpha_0^{-1} + \mathbb{E}\left\{ \frac{1}{1 - \phi(\boldsymbol{X}; \boldsymbol{\beta}_0, k_0)} \right\},
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{C}\mathrm{ov}(u_{n4}, u_{n5}) &= \mathbb{E}\left[-\frac{1-\alpha_0}{N_0} \sum_{i=1}^{N_0} \left\{ \frac{I(Y_i = 0)}{\alpha_0} - \frac{I(Y_i > 0)}{1 - \phi(\boldsymbol{X}_i; \boldsymbol{\beta}_0, k_0)} \right\} \right. \\
&\qquad \left. \cdot \frac{\{\phi(\boldsymbol{X}_i; \boldsymbol{\beta}_0, k_0) - \alpha_0\} I(Y_i > 0)}{1 - \phi(\boldsymbol{X}_i; \boldsymbol{\beta}_0, k_0)} \right] \\
&= -(1-\alpha_0) + (1-\alpha_0)^2 \mathbb{E}\left\{ \frac{1}{1 - \phi(\boldsymbol{X}; \boldsymbol{\beta}_0, k_0)} \right\}, \\
\mathbb{V}\mathrm{ar}(u_{n5}) &= N_0^{-1}(1-\alpha_0)^2 \mathbb{E}\left[\sum_{i=1}^{N_0} \frac{\{\phi(\boldsymbol{X}_i; \boldsymbol{\beta}_0, k_0) - \alpha_0\}^2 I(Y_i > 0)}{\{1 - \phi(\boldsymbol{X}_i; \boldsymbol{\beta}_0, k_0)\}^2} \right] \\
&= (1-\alpha_0)^2 \mathbb{E}\left[\frac{\{\phi(\boldsymbol{X}; \boldsymbol{\beta}_0, k_0) - \alpha_0\}^2}{1 - \phi(\boldsymbol{X}; \boldsymbol{\beta}_0, k_0)} \right] = V_{55}.
\end{aligned}
$$

By the central limit theorem, as $N_0 \to \infty$ we have $\mathbf{u}_n \xrightarrow{d} \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma})$, where:

$$
\boldsymbol{\Sigma} = \begin{pmatrix}
-V_{11} & \mathbf{0}^\top & 0 & -V_{14} & 0 \\
\mathbf{0} & -\mathbf{V}_{22} & -\mathbf{V}_{23} & \mathbf{0} & \mathbf{0} \\
0 & -\mathbf{V}_{32} & -\mathbf{V}_{33} & 0 & 0 \\
-V_{41} & \mathbf{0}^\top & 0 & 2V_{45}(1-\alpha_0)^{-2} - V_{44} & V_{55}(1-\alpha_0)^{-2} \\
0 & \mathbf{0}^\top & 0 & V_{55}(1-\alpha_0)^{-2} & V_{55}
\end{pmatrix}.
$$

Since the covariance matrix $\boldsymbol{\Sigma}$ has the same form as the $\boldsymbol{\Sigma}$ in Lemma 3 of the Supplementary Material in ref. [16], so does the matrix $\mathbf{W}$. The rest of the proof is similar and omitted. This completes the proof of Theorem 1. $\square$

When $C = 0$, there is no penalty term and the likelihood functions with and without penalty coincide. This implies that the asymptotic results in Theorem 1 hold for the empirical likelihood estimators without penalty. Utilizing the result (c), a penalized empirical likelihood ratio interval estimator can be constructed, namely:

$$
\mathcal{I}_p = \left\{ N : 2\left\{ \ell_p(\widehat{N}_p, \widehat{\boldsymbol{\beta}}_p, \widehat{k}_p, \widehat{\alpha}_p) - \max_{(\boldsymbol{\beta}, k, \alpha)} \ell_p(N, \boldsymbol{\beta}, k, \alpha) \right\} \leq \chi_1^2(1-a) \right\},
$$

where $\chi_1^2(1-a)$ stands for the $(1-a)$th quantile of χ_1^2. Correspondingly, the empirical likelihood ratio interval estimator derived without penalty is as follows:

$$
\mathcal{I} = \left\{ N : 2\left\{ \ell(\widehat{N}, \widehat{\boldsymbol{\beta}}, \widehat{k}, \widehat{\alpha}) - \max_{(\boldsymbol{\beta}, k, \alpha)} \ell(N, \boldsymbol{\beta}, k, \alpha) \right\} \leq \chi_1^2(1-a) \right\}.
$$

Despite both interval estimators asymptotically yielding correct coverage probability $(1-a)$, our simulation studies indicate that $\mathcal{I}_p$ generally outperforms $\mathcal{I}$ in terms of interval width.

Remark 1. *One might question whether overdispersion exists, or equivalently, whether the zero-truncated Poisson regression model adequately fit the data. Various methods have been proposed to address this question. See, for instances, refs. [8,11,25,26].*

2.4. Numerical Implementation

In this section, we aim to develop an EM algorithm to facilitate the proposed estimation method described in Section 2.3. For better presentation, we begin by considering a special case when N is fixed. Our primary objective is to maximize the profile penalized empirical log-likelihood function for a given N, as specified in Equation (4). In other words, we shall design an EM algorithm to calculate the maximum penalized empirical likelihood estimator of $(\boldsymbol{\beta}, k, \alpha)$ when N is fixed.

In this case, the observed data can be represented as $O = \{(y_1, x_1), (y_2, x_2), \ldots, (y_n, x_n)\}$, where each y_i is positive. Additionally, the observed data include the counts $(y_{n+1}, y_{n+2}, \ldots, y_N)$, all of which are zero. For these individuals not captured, their covariate information is missing and represented as $O^* = (x^*_{n+1}, x^*_{n+2}, \ldots, x^*_N)$. According to the principle of empirical likelihood, the potential values of the x^*_i's are drawn from $(x_1, x_2, \ldots, x_n)$, where the associated probabilities are $(p_1, p_2, \ldots, p_n)$.

The observed and missing data constitute the complete data. The likelihood is as follows:

$$\prod_{i=1}^{n}\{P(X = x_i)P(Y = y_i \mid X = x_i)\} \times \prod_{i=n+1}^{N}\{P(X = x^*_i)P(Y = 0 \mid X = x^*_i)\}.$$

Correspondingly, the log-likelihood of $\theta = (\beta^\top, k, \alpha, \{p_i\})^\top$ becomes:

$$\ell_c(\theta) = \sum_{i=1}^{n}[\log\{f(y_i, x_i; \beta, k)\} + \log(p_i)]$$
$$+ \sum_{j=n+1}^{N}\sum_{i=1}^{n}I(x^*_j = x_i)[\log\{f(0, x_i; \beta, k)\} + \log(p_i)].$$

The core of the EM-algorithm is its iterative process, which consists of an expectation step (E-step) followed by a maximization step (M-step) in each iteration. Before these two steps, we use $\theta^{old} = (\beta^{old}, k^{old}, \alpha^{old}, \{p^{old}_i\})$ to denote the current value of parameters. In the E-step, we need to compute the expectation of $\ell_c(\theta)$ conditional on O and θ^{old}. For this purpose, we calculate the conditional expectation of the indicator $I(x^*_j = x_i)$, which is equal to:

$$P(X = x_i \mid Y = 0, \theta = \theta^{old}) = \frac{\phi(x_i; \beta^{old}, k^{old})p^{old}_i}{\alpha^{old}},$$

where $\alpha^{old} = \sum_{i=1}^{n}\phi(x_i; \beta^{old}, k^{old})p^{old}_i$ denotes the current value of α. Correspondingly, the conditional expectation of the log-likelihood $\ell_c(\theta)$ is equal to:

$$Q\left(\theta \mid \theta^{old}\right) = \sum_{i=1}^{n}\left[\log\{f(y_i, x_i; \beta)\} + u^{old}_i \log\{f(0, x_i; \beta, k)\}\right]$$
$$+ \sum_{i=1}^{n}\left(1 + u^{old}_i\right)\log(p_i)$$
$$=: \ell_1(\beta, k) + \ell_2(\{p_i\}).$$

where $u^{old}_i = (N - n)\phi(x_i; \beta^{old}, k^{old})p^{old}_i/\alpha^{old}$ represents the weight for $i = 1, \ldots, n$.

The M-step consists of maximizing the function $Q(\theta \mid \theta^{old})$. The separation of parameters (β, k) and p_i's makes the maximization procedure much more elegant, which can be implemented using the following steps.

Step 1. Update (β, k) to (β^{new}, k^{new}) by maximizing $\ell_1(\beta, k)$. Given that $\ell_1(\beta, k)$ can be interpreted as a weighted log-likelihood function, we propose that maximizing $\ell_1(\beta, k)$ is analogous to fitting a negative binomial regression model to the observed counts $(y_1, y_2, \ldots, y_n)$ and the n-dimensional zero vector with covariates $(x_1, x_2, \ldots, x_n, x_1, x_2, \ldots, x_n)$ and weights $(1, 1, \ldots, 1, u^{old}_1, u^{old}_2, \ldots, u^{old}_n)$. This step can be readily implemented through the `glm.nb()` function from the `MASS` package in `R`.

Step 2. Update p_i values by maximizing $\ell_2(p_1, \ldots, p_n)$ under the positive and sum-to-one constraints. This step yields a closed form, namely, $p^{new}_i = (u^{old}_i + 1)/\sum_{i=1}^{n}(u^{old}_i + 1)$ for $i = 1, \ldots, n$.

Step 3. Update α by calculating $\alpha^{new} = \sum_{i=1}^{n}\phi(x_i; \beta^{new}, k^{new})p^{new}_i$.

The E- and M-steps are repeated until the sequence of $(\boldsymbol{\beta}^{new}, k^{new}, \alpha^{new}, \{p_i^{new}\})$ or $\widetilde{\ell}_p(N, \boldsymbol{\beta}^{new}, k^{new}, \alpha^{new}, \{p_i^{new}\})$ converges. The EM algorithm outlined above exhibits a desirable property under very general circumstances: the penalized empirical likelihood does not decrease with successive iterations. Given that the penalized empirical log-likelihood is bounded above by zero, the convergence of the sequence $(\boldsymbol{\beta}^{new}, k^{new}, \alpha^{new}, \{p_i^{new}\})$ to a local maximum of $\widetilde{\ell}_p(N, \boldsymbol{\beta}, k, \alpha, \{p_i\})$ is always guaranteed.

To compute the maximum penalized empirical likelihood estimator $(\widehat{N}_p, \widehat{\boldsymbol{\beta}}_p, \widehat{k}_p, \widehat{\alpha}_p)$, the aforementioned EM algorithm remains applicable after some modifications. In this scenario, the current parameter is denoted by $\boldsymbol{\theta}^{old} = (N^{old}, \boldsymbol{\beta}^{old}, k^{old}, \alpha^{old}, \{p_i^{old}\})$ and the weight is $u_i = (N^{old} - n)\phi(\boldsymbol{x}_i; \boldsymbol{\beta}^{old}, k^{old})p_i^{old}/\alpha^{old}$ in the E-step. In addition, the M-step incorporates a maximization step for the population size parameter.

Step 4. Calculate the updated value N^{new}, by maximizing the partial log-likelihood function relavent on N, expressed as $\log \binom{N}{n} + (N - n)\log(\alpha^{new}) + f_p(N)$. This optimization can be efficiently performed using the `optimize()` function available in the R software (version 4.3.1, https://www.r-project.org/).

The penalized empirical likelihood ratio confidence interval for N is computed by identifying the two zeros of the modified penalized likelihood ratio function:

$$\ell_m(N) = 2\left\{\ell_p(\widehat{N}_p, \widehat{\boldsymbol{\beta}}_p, \widehat{k}_p, \widehat{\alpha}_p) - \max_{(\boldsymbol{\beta}, k, \alpha)} \ell_p(N, \boldsymbol{\beta}, k, \alpha)\right\} - \chi_1^2(1 - a), \tag{6}$$

where the search for these zeros is conducted within the intervals $[n, \widehat{N}_p]$ and $[\widehat{N}_p, M]$, and M is a sufficiently large user-specified value ensuring that $\ell_m(M) > 0$. This can be implemented via the `uniroot()` function available in the R software. In summary, the pseudocodes outlined in Algorithms A1–A3 (Appendix B) offer the procedures for calculating both the maximum penalized empirical likelihood estimator and the corresponding penalized empirical likelihood ratio confidence interval for N.

3. Simulations

To demonstrate the efficiency of penalized empirical likelihood estimators, several simulations are conducted and multiple synthetic datasets are analyzed. In the simulation settings, we fix the abundance N_0 at 250, 500, and 1000. We consider two different scenarios for generating the covariate X:

(A) A binary variable, $X \sim \mathrm{Bi}(1, 0.3)$, is used to represent a discrete-valued covariate, as in the case study presented in Section 4.
(B) Alternatively, a continuous variable, $X \sim U(0, 1)$, is considered.

Given $\boldsymbol{X} = (1, X)^\top$, we simulate the count response Y from a negative binomial regression model (1), where the regression coefficient $\boldsymbol{\beta}_0$ is set at $(-0.5, 0.3)^\top$ or $(0.1, 0.3)^\top$ and the dispersion parameter k_0 is set at 0.5, 1, or 5.

Under each of the 18 ($3 \times 2 \times 3$) parameter combinations, we simulate 2000 random samples for each scenario. Subsequently, we calculate the maximum empirical likelihood estimators ($\widehat{N}$ and $\widehat{N}_p$) as well as the empirical likelihood ratio interval estimators ($\mathcal{I}$ and $\mathcal{I}_p$), where the estimators with the subscript p are derived via the penalized method. When the penalty is applied, we recommend an adaptive tuning parameter value of $C = \{2n(\nu - n)^2\}^{-1}$, which is proven effective in our numerical studies.

3.1. Evaluation of Point Estimators

We evaluate the performance of two point estimators $\widehat{N}$ and $\widehat{N}_p$, by assessing their relative bias in percent (%Rbias) and relative mean squared error (RMSE). For a generic estimator $\check{N}$, the %Rbias and RMSE are as follows:

$$\%\text{Rbias} = \frac{1}{2000} \sum_{j=1}^{2000} \frac{\check{N}_j - N_0}{N_0} \times 100 \quad \text{and} \quad \text{RMSE} = \frac{1}{2000} \sum_{j=1}^{2000} \frac{(\check{N}_j - N_0)^2}{N_0},$$

respectively, where $\check{N}_j$ represents the estimate derived from the jth random sample. Table 1 reports the simulated %Rbiases and RMSEs of both estimators $\widehat{N}$ and $\widehat{N}_p$.

We first examine the simulation results when $\boldsymbol{\beta}_0 = (0.1, 0.3)^{\top}$ with the average capture probability of 56%. From the last two columns of Table 1, we see that the %Rbiases and RMSEs of both estimators are comparable when $k_0 = 5$, where the variance-to-mean ratio (VMR) is as low as 1.3. However, the results differ significantly when k_0 decreases to 1 and further to 0.5, with VMRs of 2.3 and 3.5, respectively. As the degree of overdispersion increases, the maximum empirical likelihood estimator $\widehat{N}$ uniformly exhibits larger biases and RMSEs than the maximum penalized empirical likelihood estimator $\widehat{N}_p$. For example, in Scenario A when $k_0 = 0.5$ and $N_0 = 250$, the relative bias of $\widehat{N}$ is as large as 29.8%, and the RMSE of $\widehat{N}$ is about 44 times (1768/40) the RMSE of $\widehat{N}_p$.

Table 1. Relative biases in percent (%Rbiases) and relative mean squared errors (RMSEs) in simulation studies. The RMSE values are rounded to the nearest integer.

| | | $\boldsymbol{\beta_0 = (-0.5, 0.3)^{\top}}$ | | | | | | $\boldsymbol{\beta_0 = (0.1, 0.3)^{\top}}$ | | | | | |
| | | $k_0 = 0.5$ | | $k_0 = 1$ | | $k_0 = 5$ | | $k_0 = 0.5$ | | $k_0 = 1$ | | $k_0 = 5$ | |
	N_0	$\widehat{N}$	$\widehat{N}_p$	$\widehat{N}$	$\widehat{N}_p$	$\widehat{N}$	$\widehat{N}_p$	$\widehat{N}$	$\widehat{N}_p$	$\widehat{N}$	$\widehat{N}_p$	$\widehat{N}$	$\widehat{N}_p$
Scenario A													
%Rbias	250	82.6	−2.92	34.9	4.24	16.7	10.1	29.8	0.59	5.64	1.86	0.46	0.44
	500	37.8	3.81	14.6	5.53	4.31	4.02	6.65	2.08	1.36	1.14	0.24	0.24
	1000	19.8	7.90	3.27	2.73	1.44	1.44	1.97	1.62	0.55	0.53	−0.06	−0.06
RMSE	250	4006	73	1567	73	330	57	1768	40	136	24	5	4
	500	2558	130	881	114	51	41	271	53	25	23	4	4
	1000	2336	246	123	96	28	28	61	54	17	16	4	4
Scenario B													
%Rbias	250	83.14	−1.75	30.38	4.17	15.08	9.02	24.23	1.00	6.07	2.20	0.28	0.25
	500	33.18	4.30	11.01	4.66	3.72	3.57	6.48	2.03	1.01	0.86	0.19	0.19
	1000	15.12	6.95	3.02	2.56	1.19	1.18	1.55	1.24	0.48	0.47	−0.07	−0.07
RMSE	250	4222	73	1228	68	297	56	1143	38	128	24	4	4
	500	2474	121	720	96	34	31	571	48	19	17	4	4
	1000	1168	210	115	84	24	24	55	48	15	15	3	3

Next, we examine the results when $\boldsymbol{\beta}_0 = (-0.5, 0.3)^{\top}$. In this scenario, the %Rbiases and RMSEs of both estimators are larger than those when $\boldsymbol{\beta}_0 = (0.1, 0.3)^{\top}$. This is expected because the average capture probability reduces from 56% to 40%, making the sample provide less information about the model parameters. As in the aforementioned scenario, similar conclusions can be drawn, with the advantages of $\widehat{N}_p$ over $\widehat{N}$ being even clear. All relative biases of $\widehat{N}_p$ are smaller than 10%. In contrast, $\widehat{N}$ generally overestimates the population size, with the largest relative bias approaching 83% in Scenario B when $k_0 = 0.5$ and $N_0 = 250$. In this case, its RMSE is about 58 times (4222/73) the RMSE of $\widehat{N}_p$.

3.2. Evaluation of Interval Estimators

We evaluate and contrast the performance of two empirical likelihood ratio interval estimators: $\mathcal{I}_p$ (with penalty) and $\mathcal{I}$ (without penalty). This comparison focuses on their

coverage probabilities and interval widths. As discussed in the introduction, the interval estimators may have unbounded upper limits. In the simulation, the interval estimator's upper limit was set to the minimum of the right endpoint and 100 times the sample size.

Table 2 presents the simulation results at the 95% confidence level. Overall, the coverage accuracy of the two interval estimators is similar, with coverage probabilities either matching or slightly exceeding the nominal 95% level. However, the widths of $\mathcal{I}_p$ are always shorter than those of $\mathcal{I}$, indicating that the penalized empirical likelihood ratio interval estimator offers greater precision. Specifically, the width of $\mathcal{I}_p$ is 12% (605/5112) of that of $\mathcal{I}$ in Scenario B when $N_0 = 250$, $k_0 = 0.5$, and $\boldsymbol{\beta}_0 = (0.1, 0.3)^\top$.

Table 2. Simulated coverage probabilities (CPs, unit: %) and average widths (AWs) of interval estimators $\mathcal{I}$ and $\mathcal{I}_p$ at the 95% level, along with the proportion of bounded cases (PBC, unit: %) whose upper limits are less than 100 times the observed sample sizes. The AW and PBC values are rounded to the nearest integer.

| | | $\boldsymbol{\beta}_0 = (-0.5, 0.3)^\top$ | | | | | | $\boldsymbol{\beta}_0 = (0.1, 0.3)^\top$ | | | | | |
| | | $k_0 = 0.5$ | | $k_0 = 1$ | | $k_0 = 5$ | | $k_0 = 0.5$ | | $k_0 = 1$ | | $k_0 = 5$ | |
	N_0	$\mathcal{I}$	$\mathcal{I}_p$	$\mathcal{I}$	$\mathcal{I}_p$	$\mathcal{I}$	$\mathcal{I}_p$	$\mathcal{I}$	$\mathcal{I}_p$	$\mathcal{I}$	$\mathcal{I}_p$	$\mathcal{I}$	$\mathcal{I}_p$
Scenario A													
CP	250	94.00	93.90	93.05	93.10	97.75	97.80	93.45	93.60	94.35	94.35	95.60	95.60
	500	94.70	94.70	94.00	94.05	97.15	97.15	93.75	93.90	94.55	94.55	95.25	95.25
	1000	94.60	94.80	95.25	95.25	95.80	95.80	94.65	94.65	95.20	95.20	96.10	96.10
AW	250	742	290	836	363	647	410	776	317	577	330	185	171
	500	1532	745	1294	833	929	745	1104	696	576	498	200	200
	1000	2721	1763	2046	1584	971	909	1426	1152	591	582	258	258
PBC	250	33	100	46	100	67	100	58	100	86	100	100	100
	500	46	100	66	100	90	100	80	100	98	100	100	100
	1000	67	100	89	100	99	100	97	100	100	100	100	100
Scenario B													
CP	250	93.85	93.6	93.50	93.55	97.15	97.20	93.95	93.65	93.90	93.95	95.15	95.15
	500	94.85	94.8	94.60	94.60	96.65	96.65	94.30	94.40	94.85	94.90	95.10	95.10
	1000	94.15	94.2	95.35	95.35	96.10	96.10	94.85	94.85	95.65	95.65	95.75	95.75
AW	250	6002	820	5646	973	3654	947	5112	605	2122	458	174	154
	500	9773	1804	6999	1744	2462	1075	4817	1010	887	504	182	181
	1000	12250	3343	4981	2157	896	820	2084	1195	545	536	237	237
PBC	250	34	100	48	100	73	100	60	100	88	100	100	100
	500	48	100	70	100	93	100	83	100	99	100	100	100
	1000	70	100	92	100	100	100	98	100	100	100	100	100

Finally, we explore the possible explanations for the significant differences in the widths of the confidence intervals. For this purpose, Table 2 also reports the proportion of bounded cases, in which the upper limits are less than 100 times the observed sample sizes. In the other cases, the upper limits can be seen unbounded. From Table 2, it is clear that there are many unbounded cases for $\mathcal{I}$. The smaller N_0 or k, the more likely this boundary problem occurs. In contrast, the upper limits of $\mathcal{I}_p$ are always bounded. This demonstrates that the penalized empirical likelihood method effectively reduces the occurrence of boundary problems, leading to more precise intervals.

4. Case Study

Among the various bear species found in North America, the black bear (*Ursus americanus*) stands out as the most widely distributed game species. However, their populations have become fragmented, owing to over-harvesting. Understanding black bear demography is crucial for informing effective management and conservation strategies [27].

This study aims to make statistical inferences for the abundance of black bears at the U.S. Army's Fort Drum Military Installation in northern New York. For this purpose, the capture–recapture experiment was conducted using an array of 38 baited "hair snares" during June and July 2006. The study area and the locations of the 38 hair snares are illustrated in Figure 1.4 in ref. [28]. Each week, for eight consecutive weeks, barbed wire traps were baited and checked for hair samples. Over the 8-week period, a total of 47 black bears were captured at least once, with their sex recorded upon capture. The original data structure also includes the trap locations and whether each bear was captured at each location. Although capture status was recorded for each bear across all 38 hair snares, we treat this as a standard capture–recapture dataset. Specifically, a bear was considered captured if it was caught at least once in any given week. The analyzed data include the number of weeks that bears were captured and the sex of each bear; see Table 3 for the frequencies.

Table 3. Description of the black bear data.

Number of Captures	1	2	3	4	5	6	7
Male	11	7	5	1	2	1	1
Female	8	4	2	1	0	1	3

We first examine the estimation results of abundance in the framework of the Poisson regression model. With this model, we find that the maximum empirical likelihood estimate equals 52. This point estimate is smaller than Chao lower bound estimate of 63, indicating that the Poisson regression model may be not adequate to fit this dataset. This inadequacy is further illustrated via the ratio plot (see Figure 1A), indicating a nonconstant relationship between the ratios of frequencies and the number of captures.
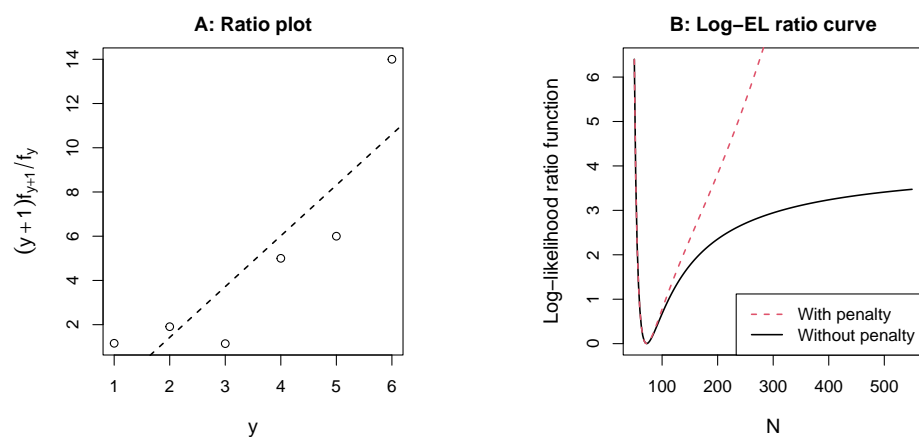

Figure 1. (**A**) The ratio plot depicts the relationship between the frequency (f_y) of capture counts (y), where the dashed line represents the fitted linear model. (**B**) The log-EL ratio curve shows the profile empirical log-likelihood ratio functions with penalty (red dashed line) and without penalty (black solid line) of the abundance.

Now, we examine the estimation results in the framework of the negative binomial regression model. Applying the maximum empirical likelihood methods with and without penalty, we find that both approaches yield a point estimate of 72 with a standard error

of 8.9 for the abundance, significantly higher than the estimate of 52 derived from the Poisson regression model. This difference can be attributed to overdispersion present in the negative binomial regression model, as indicated by an overdispersion parameter estimate of 1.2 with a standard error of 0.2. For completeness, we also calculate the ratio regression estimate proposed in ref. [12], which is 74 with a standard error of 10.2. Our point estimate is close to the ratio regression estimate but more efficient, as indicated by the smaller standard error.

Applying the empirical likelihood method with and without penalty, we obtain significantly different interval estimates of the abundance. At the 95% confidence level, the empirical likelihood ratio interval estimate ($\mathcal{I}$) is [52, 1343], whereas the penalized empirical likelihood ratio interval estimate ($\mathcal{I}_p$) is [52, 201]. Clearly, $\mathcal{I}$ exhibits a significantly larger upper limit in comparison to $\mathcal{I}_p$. In order to understand this discrepancy, we plot the profile empirical log-likelihood ratio functions with and without penalty of the abundance (see Figure 1B). The two functions are very close when the abundance N is less than 100 but diverge when $N > 100$. The empirical log-likelihood ratio function increases gradually without penalty, flattening as N increases. In contrast, the penalized log-likelihood ratio function exhibits a rapid ascent with increasing N, thereby enhancing the precision of the confidence interval. This demonstrates how the penalized empirical likelihood method effectively mitigates the boundary problem caused by overdispersion.

As suggested by one peer reviewer, we also applied the classical Lincoln–Petersen method [1,2] to the black bear data. For performing the Lincoln–Petersen method, the eight weeks of data were divided into two samples: the initial weeks constituted sample 1, while the remaining weeks formed sample 2. The Lincoln–Petersen estimator is calculated as $n_1 n_2 / n_{12}$, where n_1, n_2, and n_{12} represent the number of individuals in sample 1, sample 2, and the overlap between the two samples, respectively. The variance of this estimator is calculated as:

$$\frac{(n_1 + 1)(n_2 + 1)(n_1 - n_{12})(n_2 - n_{12})}{(n_{12} + 1)^2 (n_{12} + 2)}.$$

Table 4 presents the Lincoln–Petersen estimates along with their standard errors. It is clear that all estimates are consistently lower than the maximum penalized empirical likelihood estimate of 72. This discrepancy is likely due to the assumption of sample independence inherent in the Lincoln–Petersen method, which may not hold true in this context. Specifically, bears captured in the first sample may be more likely to be recaptured due to the lure, leading to a positive correlation between the two samples and an overestimation of capture probability. Additionally, the standard errors for the Lincoln–Petersen estimates are uniformly smaller than the standard error of 8.9 obtained from our proposed method. This difference might arise because the Lincoln–Petersen estimator assumes equal capture probability across all individuals, failing to account for heterogeneity, such as sex-based variation. In contrast, our proposed method explicitly incorporates sex as a factor, addressing this limitation.

Table 4. Lincoln–Petersen estimates and standard errors (SEs).

Sample 1	Sample 2	n_1	n_2	n_{12}	Estimate	SE
week 1	weeks 2–8	9	45	7	58	7.8
weeks 1–2	weeks 3–8	14	43	10	60	7.7
weeks 1–3	weeks 4–8	27	38	18	57	5.2
weeks 1–4	weeks 5–8	33	38	24	52	3.2
weeks 1–5	weeks 6–8	38	31	22	54	3.8
weeks 1–6	weeks 7-8	40	31	22	54	3.8
weeks 1–7	week 8	45	11	9	55	6.0

Given that the true population size is unknown, we aim to validate the reliability of our method's conclusions by assessing whether the negative binomial regression model accurately fits the observed frequency data. To this end, we have included a goodness-of-fit test in the revision. Specifically, for the black bear data, we define the χ^2 test statistic as:

$$\chi^2 = \frac{(e_1 - m_1)^2}{e_1} + \frac{(e_2 - m_2)^2}{e_2} + \frac{(e_3 - m_3)^2}{e_3} + \frac{(e_4 - m_4)^2}{e_4} + \frac{(e_5 - m_5)^2}{e_5},$$

where $m_1 = 19$, $m_2 = 11$, $m_3 = 7$, $m_4 = 2$, and $m_5 = 8$ represent the observed frequencies of captures occurring once, twice, three times, four times, and at least five times, respectively. The expected frequencies are computed as:

$$e_k = n \frac{\sum_{i=1}^n \widehat{p}_i f(k, \mathbf{x}_i; \widehat{\boldsymbol{\beta}}_p, \widehat{k}_p)}{\sum_{i=1}^n \widehat{p}_i \{1 - \phi(\mathbf{x}_i; \widehat{\boldsymbol{\beta}}_p, \widehat{k}_p)\}}, \quad k = 1, 2, 3, 4,$$

with $e_5 = n - \sum_{k=1}^4 e_k$, where the $\widehat{p}_i$'s are those convergence values of the p_i's obtained from the EM algorithm. The χ^2 test statistic, calibrated with a χ_4^2 distribution, yields a value of 1.66 with a p-value of 0.8. This result suggests that the negative binomial regression model fits the black bear data well, thereby supporting the reliability of our estimation results with this model.

5. Conclusions and Discussion

The use of the negative binomial regression model is prevalent to address heterogeneity and dispersion related to capture–recapture frequency data. As noted in refs. [10,11], fitting this model often leads to identification and boundary problems for the dispersion parameter and then yields unbounded population size estimates. To address this boundary problem, we proposed imposing a half-normal prior on the population size or equivalently decreasing the empirical log-likelihood function for large population sizes by adding a suitable penalty function. This penalized technique could improve robustness of the maximum penalized empirical likelihood estimator, ensuring the derivation of a consistently bounded interval estimator. This penalized empirical likelihood approach constitutes a significant contribution of our paper. Additionally, we introduced an efficient EM algorithm to maximize the penalized empirical likelihood function. Unlike the classical Newton-type optimization method, the EM algorithm guarantees local convergence of the numerical procedure and yields stable estimates of population size. Compared to estimators without penalty, the proposed maximum penalized empirical likelihood estimator exhibits a higher efficiency and the penalized empirical likelihood ratio interval estimator is more precise.

This paper assumes that the capture–recapture experiment is conducted at a single site. However, in ecological studies, experiments are often conducted at multiple sites or follow a spatial pattern [29]. For example, the case study in Section 4 belongs to this case. To accommodate these general application scenarios, the count data can be considered as copies of $(Y_{(1)}, Y_{(2)}, \dots, Y_{(J)})$, where, for $j = 1, 2, \dots, J$, $Y_{(j)}$ represents the number of times a generic individual was captured at the jth site. In such cases, the multivariate mixed negative binomial regression model used in ref. [30] can be adopted to fit the count data. Specifically, the conditional probability model (1) becomes:

$$P(Y_{(1)} = y_{(1)}, Y_{(2)} = y_{(2)}, \dots, Y_{(J)} = y_{(J)} \mid \boldsymbol{X} = \boldsymbol{x})$$

$$= \int_0^\infty \cdots \int_0^\infty \prod_{j=1}^J \frac{\Gamma(y_{(j)} + k_j)}{\Gamma(y_{(j)} + 1)\Gamma(k_j)} \left\{ \frac{k_j}{k_j + \lambda_j \mu(\boldsymbol{x}; \boldsymbol{\beta})} \right\}^{k_j} \left\{ \frac{\lambda_j \mu(\boldsymbol{x}; \boldsymbol{\beta})}{k_j + \lambda_j \mu(\boldsymbol{x}; \boldsymbol{\beta})} \right\}^{y_{(j)}}$$

$$h(\lambda_1, \dots, \lambda_J; \gamma) d\lambda_1 \dots d\lambda_J,$$

where $(\lambda_1, \dots, \lambda_J)$ describes the sites' random effects and $h(\lambda_1, \dots, \lambda_J; \gamma)$ describes the spatial variation of capture intensities across sites. Following the approach used in ref. [30],

one can opt for a gamma distribution as the mixing distribution, with a probability density function described by:

$$h(\lambda_1, \ldots, \lambda_J; \gamma) = \frac{\gamma^\gamma}{\Gamma(\gamma)} \exp(-\gamma\lambda)\lambda^{\gamma-1} I(\lambda_1 = \cdots = \lambda_J = \lambda),$$

where $\gamma > 0$. When spatial auto-correlation arises, indicating that random effects are partially correlated across different sites, $(\log(\lambda_1), \ldots, \log(\lambda_J))$ can be modeled by a Gaussian random field. The covariance matrix of this field can be determined by flexible spatial covariance functions [31].

Throughout the paper, the conditional mean of the count data are assumed to have a log-linear relationship with individual covariates, as depicted in model (1). In practice, the relationship between the conditional mean and covariates may be more complex. To adapt our approach for practical application scenarios, one can employ the negative binomial additive model with flexible smoothing functions for each covariate [32]. In addition, applying the proposed methods to a more generalized Poisson mixture model, where the Tweedie distribution is used as the mixing distribution [33], is a straightforward yet challenging task.

Author Contributions: Conceptualization, Y.J. and Y.L.; methodology, Y.J. and Y.L.; software, Y.J. and Y.L.; validation, Y.L.; formal analysis, Y.J.; data curation, Y.J.; visualization, Y.J.; writing—original draft preparation, Y.J.; writing—review and editing, Y.L.; supervision, Y.L.; project administration, Y.L.; funding acquisition, Y.L. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The code and data required to reproduce analyses in the case study are available on Github at https://github.com/ecnuliuyang/AbunNB (accessed on 4 June 2024).

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Some Preparation for the Proof of Theorem 1

The following lemmas facilitate the proof of Theorem 1. Specifically, Lemma A1 establishes the magnitude bounds on the first two derivatives of the penalty function. Lemma A2 identifies the leading terms of the logarithmic derivatives of the gamma functions. Lemma A3 presents fundamental results regarding the expectations of sample-based functions. Finally, Lemma A4 provides key results for conditional expectations within the framework of the negative binomial regression model (1).

Lemma A1. *Suppose that* $f_p(N) = -C(N - \nu)^2 I(N > \nu)$, $C = O_p(N_0^{-2})$. *Then (a)* $f_p'(N_0) = O_p(N_0^{-1})$; *and (b)* $f_p''(N_0) = O_p(N_0^{-3/2})$.

Proof. It suffices to show that $\nu = O_p(N_0)$, which holds because ν is the lower bound of N_0. $\square$

Lemma A2. *Let* $\Gamma(x)$ *be the gamma function. Define:*

$$\begin{aligned} S_1(N, n) &= \frac{\partial \log\{\Gamma(N+1)/\Gamma(N-n+1)\}}{\partial N}, \\ S_2(N, n) &= \frac{\partial^2 \log\{\Gamma(N+1)/\Gamma(N-n+1)\}}{\partial N^2}. \end{aligned}$$

We have:

$$S_1(N_0, n) \ = \ -\log(\alpha_0) + \frac{(n/N_0) - 1 + \alpha_0}{\alpha_0} + O_p(N_0^{-1}), \tag{A1}$$

$$S_2(N_0, n) \ = \ -\frac{1 - \alpha_0}{N_0 \alpha_0} + O_p(N_0^{-3/2}). \tag{A2}$$

Proof. We refer readers to pages 6 and 9 of the Supplementary Material of ref. [16]. □

Lemma A3. *Suppose $r(x)$ is a nonzero function of x and $g(y)$ is a function of y. Then the following results hold.*

(a) If $\mathbb{E}[r(X)\{1 - \phi(X; \boldsymbol{\beta}_0, k_0)\}] < \infty$, we have:

$$\mathbb{E}\left\{\frac{1}{N_0} \sum_{i=1}^{n} r(x_i)\right\} = \mathbb{E}[r(X)\{1 - \phi(X; \boldsymbol{\beta}_0, k_0)\}].$$

(b) If $\mathbb{E}[\{r(X)\}^2\{1 - \phi(X; \boldsymbol{\beta}_0, k_0)\}] < \infty$, we have:

$$\frac{1}{N_0} \sum_{i=1}^{n} r(x_i) - \mathbb{E}[r(X)\{1 - \phi(X; \boldsymbol{\beta}_0, k_0)\}] = O_p(N_0^{-1/2}).$$

(c) If $\mathbb{E}[r(X)\mathbb{E}\{g(Y) \mid X\}] < \infty$ and $g(0) = 0$, we have:

$$\mathbb{E}\left\{\frac{1}{N_0} \sum_{i=1}^{n} g(y_i) r(x_i)\right\} = \mathbb{E}[r(X)\mathbb{E}\{g(Y) \mid X\}].$$

Specifically, the above equation equals $\mu(X; \boldsymbol{\beta}_0, k_0)$ if $g(y) = y$.

Proof. Let $\{(X_i, Y_i) : i = 1, \ldots, N_0\}$ denote the independent and identically distributed (i.i.d) copies of (X, Y) for individuals in the population. Note that:

$$\frac{1}{N_0} \sum_{i=1}^{n} r(x_i) = \frac{1}{N_0} \sum_{i=1}^{N_0} r(X_i) I(Y_i > 0),$$

the right hand side of which is a summation of i.i.d random variables. Hence, Result (a) follows from the fact that

$$\mathbb{E}\{r(X_i) I(Y_i > 0)\} = \mathbb{E}[r(X_i)\mathbb{E}\{I(Y_i > 0)|X_i\}] = \mathbb{E}[r(X_i)\{1 - \phi(X; \boldsymbol{\beta}_0, k_0)\}].$$

where the last equality uses that $\phi(x; \boldsymbol{\beta}_0, k_0) = \mathrm{pr}(Y = 0 \mid X = x)$.

For Result (b), we first write:

$$\frac{1}{N_0} \sum_{i=1}^{n} r(x_i) - \mathbb{E}[r(X)\{1 - \phi(X; \boldsymbol{\beta}_0, k_0)\}]$$

$$= \ \frac{1}{N_0} \sum_{i=1}^{N_0} \left(r(X_i) I(D_i > 0) - \mathbb{E}[r(X)\{1 - \phi(X; \boldsymbol{\beta}_0, k_0)\}] \right).$$

Because $\mathbb{E}[\{r(X)\}^2\{1 - \phi(X; \boldsymbol{\beta}_0, k_0)\}] < \infty$ and $r(x)$ is nonzero, by the central limit theorem we have:

$$N_0^{1/2}\left\{\frac{1}{N_0} \sum_{i=1}^{n} r(x_i) - \mathbb{E}[r(X)\{1 - \phi(X; \boldsymbol{\beta}_0, k_0)\}]\right\} \xrightarrow{d} \mathrm{N}[0, \mathbb{V}\mathrm{ar}\{r(X) I(Y > 0)\}],$$

where

$$\mathbb{V}\text{ar}\{r(\boldsymbol{X})I(Y > 0)\} = \mathbb{E}[\{r(\boldsymbol{X})\}^2\{1 - \phi(\boldsymbol{X}; \boldsymbol{\beta}_0, k_0)\}] - (\mathbb{E}[r(\boldsymbol{X})\{1 - \phi(\boldsymbol{X}; \boldsymbol{\beta}_0, k_0)\}])^2 < \infty$$

and $\xrightarrow{d}$ stands for convergence in distribution. Hence, result (b) follows.

For result (c), it follows that:

$$\mathbb{E}\left\{\frac{1}{N_0}\sum_{i=1}^n g(y_i)r(\boldsymbol{x}_i)\right\} = \mathbb{E}[r(\boldsymbol{X})\mathbb{E}\{g(Y)I(Y > 0) \mid \boldsymbol{X}\}] = \mathbb{E}[r(\boldsymbol{X})\mathbb{E}\{g(Y) \mid \boldsymbol{X}\}].$$

□

Lemma A4. *Under the negative binomial regression model* (1)*, we have:*

$$\mathbb{E}[S_1(Y + k - 1, Y) \mid \boldsymbol{X}] = -\log\left\{\frac{k}{k + \mu(\boldsymbol{X}; \boldsymbol{\beta})}\right\}, \tag{A3}$$

$$\mathbb{E}[YS_1(Y + k - 1, Y) \mid \boldsymbol{X}] = \left[\frac{1}{k} - \log\left\{\frac{k}{k + \mu(\boldsymbol{X}; \boldsymbol{\beta})}\right\}\right]\mu(\boldsymbol{X}; \boldsymbol{\beta}), \tag{A4}$$

$$\mathbb{E}[S_2(Y + k - 1, Y) \mid \boldsymbol{X}] = \left[\log\left\{\frac{k}{k + \mu(\boldsymbol{X}; \boldsymbol{\beta})}\right\}\right]^2 - \mathbb{E}[\{S_1(Y + k - 1, Y)\}^2 \mid \boldsymbol{X}]. \tag{A5}$$

Proof. Consider the negative binomial distribution:

$$q(y; \mu, k) = \frac{\Gamma(y + k)}{\Gamma(y + 1)\Gamma(k)}\left(\frac{k}{k + \mu}\right)^k\left(\frac{\mu}{k + \mu}\right)^y, \quad y = 0, 1, \dots.$$

We use $\mathbb{E}_q$ to denote the expectation of Y with respect to the probability mass function $q(y; \mu, k)$. Under regular conditions that the expectation operation and the first two partial derivatives of $q(y; \mu, k)$ with respect to (μ, k) are exchangeable, we have:

$$\mathbb{E}_q\left[\frac{\partial \log\{q(Y; \mu, k)\}}{\partial k}\right] = 0, \tag{A6}$$

$$\mathbb{E}_q\left[\frac{\partial^2 \log\{q(Y; \mu, k)\}}{\partial \mu \partial k}\right] = -\mathbb{E}_q\left[\frac{\partial \log\{q(Y; \mu, k)\}}{\partial \mu}\frac{\partial \log\{q(Y; \mu, k)\}}{\partial k}\right], \tag{A7}$$

$$\mathbb{E}_q\left[\frac{\partial^2 \log\{q(Y; \mu, k)\}}{\partial k^2}\right] = -\mathbb{E}_q\left(\left[\frac{\partial \log\{q(Y; \mu, k)\}}{\partial k}\right]^2\right). \tag{A8}$$

Since:

$$\frac{\partial \log\{q(y; \mu, k)\}}{\partial k} = \log\left(\frac{k}{k + \mu}\right) - \frac{y - \mu}{k + \mu} + S_1(y + k - 1, y),$$

it follows from Equation (A6) and $\mathbb{E}_q(Y) = \mu$ that $\mathbb{E}_q\{S_1(Y + k - 1, Y)\} = -\log\{k/(k + \mu)\}$. and thus, Equation (A3) holds. Since:

$$\mathbb{E}_q\left[\frac{\partial^2 \log\{q(Y; \mu, k)\}}{\partial \mu \partial k}\right] = \mathbb{E}_q\left\{\frac{Y - \mu}{\mu(k + \mu)}\left(1 - \frac{1}{k + \mu}\right)\right\} = 0,$$

$$\mathbb{E}_q\left[\frac{\partial \log\{q(Y; \mu, k)\}}{\partial \mu} \cdot \frac{\partial \log\{q(Y; \mu, k)\}}{\partial k}\right]$$

$$= \mathbb{E}_q\left[\frac{k(Y - \mu)}{\mu(k + \mu)} \cdot \left\{\log\frac{k}{k + \mu} - \frac{Y - \mu}{k + \mu} + S_1(Y + k - 1, Y)\right\}\right]$$

$$= \frac{k}{\mu(k + \mu)}\left[\mathbb{E}_q\{YS_1(Y + k - 1, Y)\} - \frac{\mu}{k} + \mu\log\frac{k}{k + \mu}\right],$$

where the last equation uses $\mathbb{E}_q\{(Y - \mu)^2\} = \mu(k + \mu)/k$. It follows from Equation (A7) that $\mathbb{E}_q\{YS_1(Y + k - 1, Y)\} = [k^{-1} - \log\{k/(k + \mu)\}]\mu$, and thus, Equation (A4) holds.

It can be verified that:

$$\mathbb{E}_q\left[\frac{\partial^2 \log\{q(Y;\mu,k)\}}{\partial k^2}\right] = \frac{1}{k} - \frac{1}{k + \mu} + \mathbb{E}_q\{S_2(Y + k - 1, Y)\},$$

$$\mathbb{E}_q\left[\left\{\frac{\partial \log\{q(Y;\mu,k)\}}{\partial k}\right\}^2\right] = -\left\{\log\left(\frac{k}{k + \mu}\right)\right\}^2 + \mathbb{E}_q\left\{\left(\frac{\mu - Y}{k + \mu}\right)^2\right\}$$

$$+ \mathbb{E}_q[\{S_1(Y + k - 1, Y)\}^2] - \frac{2\mu}{k(k + \mu)}$$

$$= -\left\{\log\left(\frac{k}{k + \mu}\right)\right\}^2 - \frac{\mu}{k(k + \mu)} + \mathbb{E}_q[\{S_1(Y + k - 1, Y)\}^2].$$

Following Equation (A8), we have:

$$\mathbb{E}_q[S_2(Y + k - 1, Y)] + \mathbb{E}_q[\{S_1(Y + k - 1, Y)\}^2] = \left\{\log\left(\frac{k}{k + \mu}\right)\right\}^2,$$

and thus, Equation (A5) holds. □

Appendix B. Pseudocodes to Perform the Penalized Empirical Likelihood Method

Pseudocodes are provided for the implementation of the penalized empirical likelihood method. Algorithm A1 details the steps for maximizing $\ell_p(N, \boldsymbol{\beta}, k, \alpha)$ with respect to $(\boldsymbol{\beta}, k, \alpha)$ for a fixed N using the EM algorithm. Algorithm A2 presents the process for obtaining the maximum penalized empirical likelihood estimator $(\widehat{N}_p, \widehat{\boldsymbol{\beta}}_p, \widehat{k}_p, \widehat{\alpha}_p)$. Finally, Algorithm A3 describes the procedure for constructing the penalized empirical likelihood ratio confidence interval $\mathcal{I}_p$.

Algorithm A1 Pseudocode to calculate $\max_{(\boldsymbol{\beta}, k, \alpha)} \ell_p(N, \boldsymbol{\beta}, k, \alpha)$ via the EM algorithm

---

1: **Input:** Observations $\{(y_1, \boldsymbol{x}_1), (y_2, \boldsymbol{x}_2), \ldots, (y_n, \boldsymbol{x}_n)\}$, fixed value N, initial parameter values $(\boldsymbol{\beta}^{\text{new}}, k^{\text{new}}, \alpha^{\text{new}}, \{p_i^{\text{new}}\})$, and convergence threshold $\epsilon = 10^{-5}$

2: **Output:** $\ell_p(N, \boldsymbol{\beta}^{\text{old}}, k^{\text{old}}, \alpha^{\text{old}})$

3: **do**

4:     Set $(\boldsymbol{\beta}^{\text{old}}, k^{\text{old}}, \alpha^{\text{old}}, \{p_i^{\text{old}}\}) = (\boldsymbol{\beta}^{\text{new}}, k^{\text{new}}, \alpha^{\text{new}}, \{p_i^{\text{new}}\})$

5:     **for** $i = 1, 2, \ldots, n$ **do**

6:         Compute $u_i^{\text{old}} = \frac{(N-n)\phi(\boldsymbol{x}_i; \boldsymbol{\beta}^{\text{old}}, k^{\text{old}})p_i^{\text{old}}}{\alpha^{\text{old}}}$

7:     **end for**

8:     Update parameters $(\boldsymbol{\beta}^{\text{new}}, k^{\text{new}})$ using the R function `glm.nb()` by fitting a negative binomial regression model to $(y_1, y_2, \ldots, y_n, \underbrace{0, 0, \ldots, 0}_{n})$ with covariates $(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n, \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)$ and weights $(1, 1, \ldots, 1, u_1^{\text{old}}, u_2^{\text{old}}, \ldots, u_n^{\text{old}})$

9:     **for** $i = 1, 2, \ldots, n$ **do**

10:         Update $p_i^{\text{new}} = \frac{u_i^{\text{old}} + 1}{\sum_{i=1}^{n}(u_i^{\text{old}} + 1)}$

11:     **end for**

12:     Compute $\alpha^{\text{new}} = \sum_{i=1}^{n} \phi(\boldsymbol{x}_i; \boldsymbol{\beta}^{\text{new}}, k^{\text{new}})p_i^{\text{new}}$

13:     Calculate the difference diff $= \ell_p(N, \boldsymbol{\beta}^{\text{new}}, k^{\text{new}}, \alpha^{\text{new}}) - \ell_p(N, \boldsymbol{\beta}^{\text{old}}, k^{\text{old}}, \alpha^{\text{old}})$

14: **while** $|\text{diff}| > \epsilon$

Algorithm A2 Pseudocode for estimating $(\widehat{N}_p, \widehat{\boldsymbol{\beta}}_p, \widehat{k}_p, \widehat{\alpha}_p)$ using the EM algorithm

---

1: **Input:** Observations $\{(y_1, \boldsymbol{x}_1), (y_2, \boldsymbol{x}_2), \ldots, (y_n, \boldsymbol{x}_n)\}$, initial parameter values $(N^{new}, \boldsymbol{\beta}^{new}, k^{new}, \alpha^{new}, \{p_i^{new}\})$, and threshold $\epsilon = 10^{-5}$

2: **Output:** Estimate $(N^{old}, \boldsymbol{\beta}^{old}, k^{old}, \alpha^{old})$

3: **do**

4:     Set $(N^{old}, \boldsymbol{\beta}^{old}, k^{old}, \alpha^{old}, \{p_i^{old}\}) = (N^{new}, \boldsymbol{\beta}^{new}, k^{new}, \alpha^{new}, \{p_i^{new}\})$

5:     **for** $i = 1, 2, \ldots, n$ **do**

6:         Compute $u_i^{old} = \frac{(N^{old} - n)\phi(\boldsymbol{x}_i; \boldsymbol{\beta}^{old}, k^{old})p_i^{old}}{\alpha^{old}}$

7:     **end for**

8:     Update parameters $(\boldsymbol{\beta}^{new}, k^{new})$ by fitting a negative binomial regression model to $(y_1, y_2, \ldots, y_n, \underbrace{0, 0, \ldots, 0}_{n})$ with covariates $(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n, \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)$ and weights $(1, 1, \ldots, 1, u_1^{old}, u_2^{old}, \ldots, u_n^{old})$ using the R function `glm.nb()`

9:     **for** $i = 1, 2, \ldots, n$ **do**

10:         Calculate $p_i^{new} = \frac{u_i^{old} + 1}{\sum_{i=1}^{n}(u_i^{old} + 1)}$

11:     **end for**

12:     Compute $\alpha^{new} = \sum_{i=1}^{n} \phi(\boldsymbol{x}_i; \boldsymbol{\beta}^{new}, k^{new})p_i^{new}$

13:     Update N^{new} by maximizing $\log\binom{N}{n} + (N - n)\log(\alpha^{new}) + f_p(N)$ with respect to N using the R function `optimize()`

14:     Calculate the difference diff $= \ell_p(N^{new}, \boldsymbol{\beta}^{new}, k^{new}, \alpha^{new}) - \ell_p(N^{old}, \boldsymbol{\beta}^{old}, k^{old}, \alpha^{old})$

15: **while** $|\text{diff}| > \epsilon$

Algorithm A3 Pseudocode for calculating $\mathcal{I}_p$ at the $(1 - a)$ confidence level

---

1: **Input:** Observations $\{(y_1, \boldsymbol{x}_1), (y_2, \boldsymbol{x}_2), \ldots, (y_n, \boldsymbol{x}_n)\}$ and significance level a

2: **Output:** Confidence interval $[\widehat{N}_{pl}, \widehat{N}_{pu}]$

3: Apply Algorithm A2 to compute the maximum penalized empirical likelihood estimate $(\widehat{N}_p, \widehat{\boldsymbol{\beta}}_p, \widehat{k}_p, \widehat{\alpha}_p)$

4: Use the R function `uniroot()` to find the lower bound $\widehat{N}_{pl}$ by searching the root of $\ell_m(N)$ in the interval $[n, \widehat{N}_p]$, where $\ell_m(N)$ is defined in Equation (6)

5: **for** $k = 1, 2, \ldots$ **do**

6:     Set $M = 2^k \widehat{N}_p$.

7:     **if** $\ell_m(M) > 0$ **then**

8:         **Break**

9:     **end if**

10: **end for**

11: Use the R function `uniroot()` to determine the upper bound $\widehat{N}_{pu}$ by searching the root of $\ell_m(N)$ within $[\widehat{N}_p, M]$, where $\ell_m(N)$ is defined in Equation (6)

References

1. Lincoln, F.C. *Calculating Waterfowl Abundance on the Basis of Banding Returns*; Number 118; U.S. Department of Agriculture: Washington, DC, USA, 1930.
2. Petersen, C.G.J. The yearly immigration of young plaice in the Limfjord from the German sea. *Rep. Dan. Biol. Stn.* **1896**, *6*, 1–77.
3. McCrea, R.S.; Morgan, B.J.T. *Analysis of Capture–Recapture Data*; Chapman & Hall/CRC: London, UK, 2014.
4. Corrao, G.; Bagnardi, V.; Vittadini, G.; Favilli, S. Capture-recapture methods to size alcohol related problems in a population. *J. Epidemiol. Community Health* **2000**, *54*, 603–610. [CrossRef] [PubMed]
5. Frischer, M.; Bloor, M.; Finlay, A.; Goldberg, D.; Green, S.; Haw, S.; McKeganey, N.; Platt, S. A new method of estimating prevalence of injecting drug use in an urban population: Results from a Scottish city. *Int. J. Epidemiol.* **1991**, *20*, 997–1000. [CrossRef] [PubMed]
6. Gallay, A.; Vaillant, V.; Bouvet, P.; Grimont, P.; Desenclos, J.C. How many foodborne outbreaks of Salmonella infection occurred in France in 1995? Application of the capture-recapture method to three surveillance systems. *Am. J. Epidemiol.* **2000**, *152*, 171–177. [CrossRef]
7. Lindén, A.; Mäntyniemi, S. Using the negative binomial distribution to model overdispersion in ecological count data. *Ecology* **2011**, *92*, 1414–1421. [CrossRef]

8. Cruyff, M.J.L.F.; van Der Heijden, P.G.M. Point and interval estimation of the population size using a zero-truncated negative binomial regression model. *Biom. J.* **2008**, *50*, 1035–1050. [CrossRef]

9. Stoklosa, J.; Blakey, R.V.; Hui, F.K. An overview of modern applications of negative binomial modelling in ecology and biodiversity. *Diversity* **2022**, *14*, 320. [CrossRef]

10. Anan, O. Capture-Recapture Modelling for Zero-Truncated Count Data Allowing for Heterogeneity. Ph.D. Thesis, University of Southampton, Southampton, UK, 2016.

11. Böhning, D. Power series mixtures and the ratio plot with applications to zero-truncated count distribution modelling. *Metron* **2015**, *73*, 201–216. [CrossRef]

12. Rocchetti, I.; Bunge, J.; Böhning, D. Population size estimation based upon ratios of recapture probabilities. *Ann. Appl. Stat.* **2011**, *5*, 1512–1533. [CrossRef]

13. Godwin, R.T. One-inflation and unobserved heterogeneity in population size estimation. *Biom. J.* **2017**, *59*, 79–93. [CrossRef]

14. van Der Heijden, P.G.M.; Bustami, R.; Cruyff, M.J.L.F.; Engbersen, G.; van Houwelingen, H.C. Point and interval estimation of the population size using the truncated Poisson regression model. *Stat. Model.* **2003**, *3*, 305–322. [CrossRef]

15. van Der Heijden, P.G.M.; Cruyff, M.J.L.F.; van Houwelingen, H.C. Estimating the size of a criminal population from police records using the truncated Poisson regression model. *Stat. Neerl.* **2003**, *57*, 289–304. [CrossRef]

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

17. Liu, Y.; Li, P.; Liu, Y.; Zhang, R. Semiparametric empirical likelihood inference for abundance from one-inflated capture–recapture data. *Biom. J.* **2022**, *64*, 1040–1055. [CrossRef] [PubMed]

18. Liu, Y.; Liu, Y.; Li, P.; Qin, J. Full likelihood inference for abundance from continuous time capture–recapture data. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **2018**, *80*, 995–1014. [CrossRef]

19. Liu, Y.; Liu, Y.; Li, P.; Zhang, R. Two-step semiparametric empirical likelihood inference from capture–recapture data with missing covariates. *Test* **2024**, in press.

20. Liu, Y.; Liu, Y.; Li, P.; Zhu, L. Maximum likelihood abundance estimation from capture-recapture data when covariates are missing at random. *Biometrics* **2021**, *77*, 1050–1060. [CrossRef]

21. Owen, A.B. Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **1988**, *75*, 237–249. [CrossRef]

22. Owen, A.B. Empirical likelihood ratio confidence regions. *Ann. Stat.* **1990**, *18*, 90–120. [CrossRef]

23. Chao, A. Estimating the population size for capture–recapture data with unequal catchability. *Biometrics* **1987**, *43*, 783–791. [CrossRef]

24. Böhning, D.; Vidal-Diez, A.; Lerdsuwansri, R.; Viwatwongkasem, C.; Arnold, M. A generalization of Chao's estimator for covariate information. *Biometrics* **2013**, *69*, 1033–1042. [CrossRef]

25. Gurmu, S. Tests for detecting overdispersion in the positive Poisson regression model. *J. Bus. Econ. Stat.* **1991**, *9*, 215–222. [CrossRef]

26. Yehia, E.G. Power of Overdispersion Tests in Zero-Truncated Negative Binomial Regression Model. *Am. J. Theor. Appl. Stat.* **2021**, *10*, 152–157. [CrossRef]

27. Beston, J.A. Variation in life history and demography of the American black bear. *J. Wildl. Manag.* **2011**, *75*, 1588–1596. [CrossRef]

28. Royle, J.A.; Chandler, R.B.; Sollmann, R.; Gardner, B. *Spatial Capture-Recapture*; Academic Press: Cambridge, MA, USA, 2013.

29. Tourani, M. A review of spatial capture–recapture: Ecological insights, limitations, and prospects. *Ecol. Evol.* **2022**, *12*, e8468. [CrossRef] [PubMed]

30. Tzougas, G.; di Cerchiara, A.P. The multivariate mixed negative binomial regression model with an application to insurance a posteriori ratemaking. *Insur. Math. Econ.* **2021**, *101*, 602–625. [CrossRef]

31. Schmidt, A.M.; Guttorp, P. Flexible spatial covariance functions. *Spat. Stat.* **2020**, *37*, 100416. [CrossRef]

32. Thurston, S.W.; Wand, M.; Wiencke, J.K. Negative binomial additive models. *Biometrics* **2000**, *56*, 139–144. [CrossRef]

33. Bonat, W.H.; Jørgensen, B.; Kokonendji, C.C.; Hinde, J.; Demétrio, C.G. Extended Poisson–Tweedie: Properties and regression models for count data. *Stat. Model.* **2018**, *18*, 24–49. [CrossRef]