

Maximum Likelihood Inference for the Band-Read Error Model for Capture-Recapture Data with Misidentification

Wei Zhang¹, Steven J. Price², Simon J. Bonner^{1,3}

¹Department of Statistical and Actuarial Sciences
University of Western Ontario, London, ON N6A 5B7, Canada

²Department of Forestry and Natural Resources, University of Kentucky
Lexington, KY 40546-0073, USA

³Department of Biology, University of Western Ontario
London, ON N6A 5B7, Canada

Correspondence

Wei Zhang

Department of Statistical and Actuarial Sciences

University of Western Ontario, London, ON N6A 5B7, Canada

Email: wzhan592@uwo.ca

Running headline

Maximum likelihood for the band-read error model

Abstract

1. Capture-recapture is a widely-used method in ecology for estimating the abundance or demographic parameters of wildlife populations. Misidentification of animals is a common problem for many capture-recapture experiments. Considerably misleading inference may be obtained when traditional models are used for capture-recapture data with misidentification.
2. One proposed model for modeling misidentification is the so-called band-read error model, which assumes that it is possible to identify one marked individual as another on each capture occasion. Currently, fitting this model relies primarily on a Bayesian Markov chain Monte Carlo approach, while maximum likelihood is difficult because there is not a computationally efficient likelihood function available. The Bayesian method is exact but requires expert implementation and considerable computation time.
3. In this paper, we propose an approximate model for modeling misidentification and then develop a fast maximum-likelihood approach for the approximate model using likelihood constructed by the saddlepoint approximation method. Compared to the Bayesian approach, the proposed approximate method yields similarly reliable estimation results but is much faster and easier to implement.
4. We demonstrate the promising performance of our proposed method by simulation and by comparisons with the Bayesian inference under the original model. We apply the method to analyze capture-recapture data from a population of Northern Dusky Salamanders (*Desmognathus fuscus*) collected in North Carolina, USA.

Keywords: Band-read error model, Capture-recapture, Latent multinomial model, Misidentification, Saddlepoint approximation, Salamander

Introduction

When applying conventional capture-recapture methods to estimate demographic parameters of animal populations, one may run into problems if misidentification occurs. Misidentification of animals is widespread in capture-recapture experiments, especially when individual identification is accomplished by recognizing natural features of animals through, for example, DNA samples (e.g., Wright *et al.*, 2009) and photographs (e.g., Morrison *et al.*, 2011), or by using some non-natural marks such as visible implant elastomer tags (e.g., Curtis, 2006). It has been shown that even for a low rate of misidentification inferences on the parameters of interest can be considerably misleading under capture-recapture models that ignore misidentification (e.g., Bonner *et al.*, 2016; Vale *et al.*, 2014; Link *et al.*, 2010).

Link *et al.* (2010) proposed model $M_{t,\alpha}$ for modeling misidentification. Key assumptions of model $M_{t,\alpha}$ include that a specific identification error never occurs twice, and that an individual is never misidentified as another captured individual. These assumptions simplify the model, but might be questionable for many data sets in practice (Link *et al.*, 2010). Bonner *et al.* (2016) further investigated the problem of misidentification under a different model, termed as the band-read error (BRE) model, assuming that identification errors cause one individual to be misidentified as another marked individual. Thus, choosing appropriate models for misidentification in practice depends on whether identification errors tend to match existing individuals or create new identities.

Both model $M_{t,\alpha}$ and the BRE model fall into a more general class of hierarchical models that can be described using $\mathbf{Y} = \mathbf{A}\mathbf{X}$, where observed data \mathbf{Y} are linearly related to a latent (unobservable) vector \mathbf{X} by a known non-invertible matrix \mathbf{A} . The latent vector \mathbf{X} , a vector of counts of true capture histories for our misidentification problem, is assumed to arise from a probability distribution with unknown parameters $\boldsymbol{\theta}$. The distribution may or may not have probability mass functions in closed forms. For example, \mathbf{X} follows a multinomial distribution for model $M_{t,\alpha}$, while we do not know exactly the distribution of \mathbf{X} for the BRE model (Bonner *et al.*,

2016). We wish to estimate the parameters $\boldsymbol{\theta}$ from the data \mathbf{Y} . A simple and immediate likelihood function here is $f(\mathbf{y}|\boldsymbol{\theta}) = \sum_{\mathbf{x} \in \mathcal{A}} f(\mathbf{x}|\boldsymbol{\theta})$, where $\mathcal{A} = \{\mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{y}\}$. However, it is computationally infeasible to use this likelihood for inference in practice because there are excessively many elements in set \mathcal{A} .

To fit model $M_{t,\alpha}$, Link *et al.* (2010) proposed a Bayesian Markov chain Monte Carlo (MCMC) method that aims to sample from the joint posterior distribution of \mathbf{x} and $\boldsymbol{\theta}$ instead of specifying set \mathcal{A} . Schofield & Bonner (2015) showed that the MCMC algorithm of Link *et al.* (2010) cannot be employed to obtain inference for the BRE model because it may produce non-irreducible Markov chains. Bonner *et al.* (2016) extended the more general MCMC algorithm of Schofield & Bonner (2015) to deal with capture-recapture models such as the BRE model with more complicated patterns of identification errors than model $M_{t,\alpha}$. These Bayesian MCMC algorithms are general, but require considerable computation time, especially for fitting the complicated BRE model. For one typical data set with four capture occasions, fitting time can be as long as 24 hours on a 2.3 GHz laptop. This is rather slow because sampling from the full conditional distribution $[\mathbf{x}|\mathbf{y}, \boldsymbol{\theta}]$ is difficult in this context (Bonner *et al.*, 2016; Schofield & Bonner, 2015).

Zhang *et al.* (2019) proposed a much faster maximum-likelihood approach based on the saddlepoint approximation method (Daniels, 1954), and demonstrated the promising performance of the method under a general class of latent multinomial models (LMMs), including model $M_{t,\alpha}$ as a specific case. Key to the method of Zhang *et al.* (2019) is finding the moment generating function of \mathbf{X} , which is straightforward when \mathbf{X} follows multinomial or many other known distributions such as multivariate Poisson. For the BRE model, the vector \mathbf{X} follows a very complicated distribution whose moment generating function cannot be obtained easily. Thus it is not trivial to implement the saddlepoint method of Zhang *et al.* (2019) to the BRE model.

Motivated by the promising efficiency of the method of Zhang *et al.* (2019), we show in this paper how the saddlepoint approximation method can be implemented for the BRE model. This paper is organized as follows. We first introduce the CJS/BRE model for open-population

capture-recapture data with misidentification. Then we propose a product-multinomial model to approximate the distribution of \mathbf{X} under the CJS/BRE model so that the saddlepoint approximation can be applied. In the following sections, we describe the saddlepoint method and methods for variance estimation. We assess the performance of the proposed method by simulation. Compared with the Bayesian approach of Bonner *et al.* (2016), the saddlepoint method is much faster. We illustrate the proposed method using data from a study of Northern Dusky Salamanders (*Desmognathus fuscus*) conducted in North Carolina, USA.

Materials and Methods

CJS/BRE model

Models accounting for identification errors in capture-recapture generally consider the events in two stages: the first describing the dynamics of population and the observation process, and the second considering the identification of individuals to construct capture histories. For example, model $M_{t,\alpha}$ combines the closed population model with time-dependent capture probabilities of Otis *et al.* (1978) with the alpha error process in which individuals are correctly identified with probability α each time they are captured and every error leads to the apparent observation of a new individual that is captured exactly one time. Following Bonner *et al.* (2016), we consider the so-called CJS/BRE model which combines the dynamic and the observation process of the Cormack-Jolly-Seber (CJS) model (Cormack, 1964; Jolly, 1965; Seber, 1965) and the error process of the BRE model.

We first consider population dynamics under the CJS model. The set of parameters for the CJS model with T capture occasions includes $\boldsymbol{\phi} = (\phi_1, \dots, \phi_{T-1})$ and $\mathbf{p} = (p_2, \dots, p_T)$, where ϕ_t denotes the probability of each individual surviving to occasion $t + 1$, and p_t denotes the probability of each individual being captured on occasion t , given that the individual is alive on occasion t . The CJS model assumes that all captured individuals are identified without error,

and that captures of different individuals or the same individual on different occasions are all independent. Each individual has two possible outcomes on each capture occasion: non-capture (code 0) and capture (code 1). For example, an individual has capture history 1001 if the individual is captured on occasions 1 and 4 while not captured on occasions 2 and 3. It is straightforward that for a T -occasion study, there are 2^T possible capture histories. Here we exclude the null history and the history with a single capture on the last occasion, which do not contribute to the likelihood of the CJS model. It follows that we have $I = 2^T - 2$ capture histories, denoted by $\boldsymbol{\xi}_i = (\xi_{i1}, \dots, \xi_{iT}), i = 1, \dots, I$. We call these histories latent capture histories, each of which represents the true encounter records of an individual without misidentification.

Next we consider identification errors under the BRE model. Key assumptions of the BRE model include: (1) all individuals are correctly identified when they are first captured and marked, (2) individuals are correctly identified with probability α on subsequent recaptures, (3) identification errors cause an individual recaptured on one occasion to be identified as another marked individual that is not resighted on the same occasion, and (4) one individual can only be involved in one event on each occasion. Given these assumptions, there are four possible events for each marked individual on each occasion:

- 0: it is not captured;
- 1: it is captured and correctly identified;
- 2: it is captured but misidentified as another marked individual;
- 3: it is not captured but another captured individual is misidentified as it.

For an experiment with T capture occasions total of 4^T histories can be constructed from these four events. We call these histories latent error histories, which describe an individual's true encounter records as well as misidentification. We first exclude the four histories with event 0 only for the first $T - 1$ occasions, because individuals captured only on the final occasion do not contribute likelihood of the CJS model. Also, the BRE model assumes that all individuals captured for the

first time are correctly identified, so we only need consider histories whose first non-zero entry is 1. As a result, we only consider $J = (4^T - 4) / 3$ latent error histories which we denote by $\mathbf{v}_j = (v_{j1}, \dots, v_{jT})$ for $j = 1, \dots, J$. The counts of these histories constitute the latent vector \mathbf{X} in the CJS/BRE model. The vector \mathbf{X} is not observable because any latent error history with misidentification (i.e. event 2 or 3) cannot be observed directly, and will generate one observable history instead.

Observable histories for the CJS/BRE model belong to the same set as the latent capture histories. We denote the observable histories by $\boldsymbol{\omega}_i = (\omega_{i1}, \dots, \omega_{iT})$ for $i = 1, \dots, I$. The counts of these histories constitute an observed data vector \mathbf{Y}_1 .

Finally we discuss the relationships between the three sorts of histories defined above. As an example, we consider an individual with latent capture history 1010. This individual would have a latent error history 1023 if it is misidentified as another individual on occasion 3 and another individual is misidentified as it on occasion 4. The latent error history 1023 will be observed as 1001. According to the relationship between latent error histories and observable histories, an $I \times J$ matrix \mathbf{A}_1 can be constructed such that $\mathbf{Y}_1 = \mathbf{A}_1 \mathbf{X}$. More specifically,

$$A_{1ij} = \begin{cases} 1, & \text{if } \omega_{it} = \mathcal{I}\{v_{jt} = 1\} + \mathcal{I}\{v_{jt} = 3\} \text{ for all } t = 1, \dots, T, \\ 0, & \text{otherwise,} \end{cases}$$

where $\mathcal{I}\{\cdot\}$ denotes the usual indicator function. Assumption (3) of the BRE model implies that the number of events 2 and 3 on occasion t must be the same for $t = 2, \dots, T$. Thus we can construct a $(T - 1) \times J$ matrix \mathbf{A}_2 such that $\mathbf{0} = \mathbf{A}_2 \mathbf{X}$, where

$$A_{2tj} = \begin{cases} -1, & \text{if } v_{jt+1} = 2, \\ 1, & \text{if } v_{jt+1} = 3, \\ 0, & \text{otherwise.} \end{cases}$$

Combining the two equations above, we get $\mathbf{Y} = \mathbf{A}\mathbf{X}$, where $\mathbf{A} = (\mathbf{A}_1^\top, \mathbf{A}_2^\top)^\top$ and $\mathbf{Y} = (\mathbf{Y}_1^\top, \mathbf{0}^\top)^\top$. The probability distribution of the latent vector \mathbf{X} (see Bonner *et al.*, 2016, for details) is considerably complicated, making it very difficult to compute the probability mass function of \mathbf{Y} explicitly and thus the likelihood for the CJS/BRE model.

Approximate model for \mathbf{X}

Since the BRE model requires the number of events 2 and 3 to be identical on each occasion, the vector \mathbf{X} satisfies $\mathbf{A}_2\mathbf{X} = \mathbf{0}$. For this reason, the vector \mathbf{X} under the CJS/BRE model does not follow a product-multinomial distribution as the standard CJS model. In order to apply the saddlepoint method of Zhang *et al.* (2019), we approximate the distribution of \mathbf{X} by a product-multinomial model and then use a product-LMM to approximate the likelihood of the CJS/BRE model. To derive the approximate product-multinomial distribution, we need to find the probabilities assigned to each cell of \mathbf{X} .

We first illustrate the relationship between latent capture and error histories, which is key to finding the cell probabilities of \mathbf{X} . Let π_j (or $\pi_{\mathbf{v}_j}$) denote the probability of latent error history \mathbf{v}_j , and let β_i (or $\beta_{\boldsymbol{\xi}_i}$) denote the probability of latent capture history $\boldsymbol{\xi}_i$, conditional on first capture. Since latent capture histories do not incorporate identification errors, we can readily obtain probabilities $\boldsymbol{\beta} = (\beta_1, \dots, \beta_I)$ in terms of $\boldsymbol{\phi}$ and \mathbf{p} under the standard CJS model. For example, the probability of latent capture history 101 is $\beta_{101} = \phi_1(1 - p_2)\phi_2p_3$. After taking into account identification errors, one latent capture history may lead to a set of several possible latent error histories. For convenience, the events 0 and 1 mentioned below are these appearing after the leading event 1 of a latent capture or error history. Event 0 in a latent capture history may generate event 0 or 3 in a resulting latent error history. Similarly, event 1 may generate event 1 or 2. For example, the history 101 may produce latent error histories 101, 102, 131, or 132. We define a set of derived parameters $\lambda_t, t = 2, \dots, T$ to denote the probabilities that event 0 on occasion t in a latent capture history leads to event 3 in a latent error history. These parameters can be

Table 1: An example of computing the probabilities of latent error histories for the CJS/BRE model with $T = 2$ capture occasions.

Latent capture history	β_i	Latent error history	π_j
10	$\beta_1 = 1 - \phi_1 p_2$	10	$\pi_1 = \beta_1 (1 - \lambda_2)$
		13	$\pi_2 = \beta_1 \lambda_2$
11	$\beta_2 = \phi_1 p_2$	11	$\pi_3 = \beta_2 \alpha$
		12	$\pi_4 = \beta_2 (1 - \alpha)$

expressed in terms of model parameters later. Recall that α denotes the probability that event 1 on any occasion in a latent capture history leads to event 1 in a latent error history.

Using the notation defined above, we assign a probability to each of the J latent error histories. Suppose ξ_i is the corresponding latent capture history of latent error history \mathbf{v}_j . It follows that

$$\pi_j = \beta_i \prod_{t=t_0+1}^T \left[(1 - \lambda_t)^{\mathcal{I}\{v_{jt}=0\}} \alpha^{\mathcal{I}\{v_{jt}=1\}} (1 - \alpha)^{\mathcal{I}\{v_{jt}=2\}} \lambda_t^{\mathcal{I}\{v_{jt}=3\}} \right],$$

where t_0 is the occasion on which the first event 1 of the history \mathbf{v}_j appears. Then the vector $\boldsymbol{\pi} = (\pi_1, \dots, \pi_J)^\top$ can be expressed in terms of $\boldsymbol{\beta}$, α , and $\boldsymbol{\lambda} = (\lambda_2, \dots, \lambda_T)$, for example, $\pi_{102} = \beta_{101} (1 - \lambda_2) (1 - \alpha)$.

We now consider how to express the derived parameters $\boldsymbol{\lambda}$ in terms of model parameters. For convenience, we rewrite the vector \mathbf{X} as $\mathbf{X} = (\mathbf{X}_1^\top, \dots, \mathbf{X}_{T-1}^\top)^\top$, where \mathbf{X}_t is a vector of counts of all latent error histories whose first event 1 appears on occasion t . Accordingly, we write $\boldsymbol{\pi}$ as $\boldsymbol{\pi} = (\boldsymbol{\pi}_1^\top, \dots, \boldsymbol{\pi}_{T-1}^\top)^\top$, where $\boldsymbol{\pi}_t$ contains probabilities of latent error histories whose counts are in \mathbf{X}_t . Let N_t denote the number of individuals that are first captured and marked on occasion $t = 1, \dots, T - 1$. It follows that the vector of expected counts of all latent error histories is given by $\mathbf{X}^* = (N_1 \boldsymbol{\pi}_1^\top, \dots, N_{T-1} \boldsymbol{\pi}_{T-1}^\top)^\top$. The CJS/BRE model requires the number of events 2 and 3 on each occasion to be the same, thus the expected count of events 2 and 3 on each occasion must be the same. Then the vector \mathbf{X}^* satisfies the following system of $T - 1$ linear equations:

$$\mathbf{A}_2 \mathbf{X}^* = \mathbf{0},$$

from which we can express $\boldsymbol{\lambda}$ in terms of α , $\boldsymbol{\beta}$, and $\mathbf{N} = (N_1, \dots, N_{T-1})$. Consequently, we can express $\boldsymbol{\pi}$ in terms of parameters α , $\boldsymbol{\phi}$, \mathbf{p} , and \mathbf{N} , because $\boldsymbol{\beta}$ is a function of $\boldsymbol{\phi}$ and \mathbf{p} . We illustrate this procedure using an example with $T = 2$ as shown in Table 1. In the example, the system $\mathbf{A}_2 \mathbf{X}^* = \mathbf{0}$ contains only one equation $\pi_2 = \pi_4$, which yields $\lambda_2 = (1 - \alpha) \beta_2 / \beta_1$.

Using the cell probabilities obtained, we approximate the distribution of each \mathbf{X}_t by a multinomial model:

$$\mathbf{X}_t \sim \text{Multinomial}(N_t; \boldsymbol{\pi}_t), \quad (1)$$

where $\boldsymbol{\pi}_t = (\pi_{t1}, \dots, \pi_{tH_t})^\top$ and H_t denotes the dimension of \mathbf{X}_t . Then we obtain a product-multinomial model and a product-LMM to approximate the distributions of the latent vector \mathbf{X} and the observed vector \mathbf{Y} .

Saddlepoint approximation

The saddlepoint approximation was initially introduced to statistics by Daniels (1954), and has become a powerful tool for approximating the probability distribution of a random variable from its moment generating function (MGF). For details about the mathematical derivation of the method, see Daniels (1954), Lugannani & Rice (1980), and Barndorff-Nielsen & Cox (1989). See Butler (2007) for applications of the saddlepoint approximation in a wide range of fields.

The MGF of the variable \mathbf{X} is defined as

$$M_{\mathbf{X}}(\mathbf{r}) = \text{E} \{ \exp(\mathbf{r}^\top \mathbf{X}) \},$$

where \mathbf{r} takes values from \mathbb{R}^J , enabling the expectation of $\exp(\mathbf{r}^\top \mathbf{X})$ to be finite. Since the components \mathbf{X}_t of \mathbf{X} are all independent, we have

$$M_{\mathbf{X}}(\mathbf{r}) = M_{\mathbf{X}_1}(\mathbf{r}_1) \cdots M_{\mathbf{X}_{T-1}}(\mathbf{r}_{T-1}),$$

where $\mathbf{r} = (\mathbf{r}_1^\top, \dots, \mathbf{r}_{T-1}^\top)^\top$ and $M_{\mathbf{X}_t}$ denotes the MGF of \mathbf{X}_t . Let $\mathbf{r}_t = (r_{t1}, \dots, r_{tH_t})^\top$. Note that

\mathbf{X}_t is assumed to follow the multinomial distribution (1), whose MGF is

$$M_{\mathbf{X}_t}(\mathbf{r}_t) = \left\{ \sum_{h=1}^{H_t} \pi_{th} \exp(r_{th}) \right\}^{N_t}.$$

Following Zhang *et al.* (2019), the MGF of $\mathbf{Y} = \mathbf{A}\mathbf{X}$ can be obtained from that of \mathbf{X} by

$$M_{\mathbf{Y}}(\mathbf{s}) = \mathbb{E} \{ \exp(\mathbf{s}^\top \mathbf{Y}) \} = \mathbb{E} \left[\exp \left\{ (\mathbf{A}^\top \mathbf{s})^\top \mathbf{X} \right\} \right] = M_{\mathbf{X}}(\mathbf{A}^\top \mathbf{s}),$$

where $\mathbf{s} \in \mathbb{R}^L$ with $L = I + T - 1$ denoting the dimension of the variable \mathbf{Y} . We denote by $\boldsymbol{\theta}$ all parameters underlying the distribution of \mathbf{Y} . The saddlepoint approximation to the probability mass function of \mathbf{Y} is

$$\tilde{f}_{\mathbf{Y}}(\mathbf{y}; \boldsymbol{\theta}) = \frac{1}{(2\pi)^{L/2} |K_{\mathbf{Y}}''(\hat{\mathbf{s}}; \boldsymbol{\theta})|^{1/2}} \exp \{ K_{\mathbf{Y}}(\hat{\mathbf{s}}; \boldsymbol{\theta}) - \hat{\mathbf{s}}^\top \mathbf{y} \}, \quad (2)$$

where $K_{\mathbf{Y}}(\mathbf{s}; \boldsymbol{\theta}) = \log \{ M_{\mathbf{Y}}(\mathbf{s}; \boldsymbol{\theta}) \}$ denotes the cumulant generating function of \mathbf{Y} , $|\cdot|$ denotes the determinant, and $\hat{\mathbf{s}} = \hat{\mathbf{s}}(\mathbf{y}, \boldsymbol{\theta})$ solves the saddlepoint equation

$$K_{\mathbf{Y}}'(\mathbf{s}; \boldsymbol{\theta}) = \mathbf{y}, \quad (3)$$

which usually does not have an analytic solution, and is treated as an optimization problem of minimizing $K_{\mathbf{Y}}(\mathbf{s}; \boldsymbol{\theta}) - \mathbf{s}^\top \mathbf{y}$ with respect to \mathbf{s} .

Maximum likelihood estimates (MLEs) of the parameters $\boldsymbol{\theta}$ are obtained by minimizing the negative logarithm of the saddlepoint likelihood function (2). This minimization is not trivial because associated with each evaluation of the saddlepoint likelihood is an inner optimization, i.e., solving the saddlepoint equation (3). See the Supporting Information of Zhang *et al.* (2019) for an efficient and accurate method of accomplishing this calculation using the R package TMB (Kristensen *et al.*, 2016).

Here we have only introduced the basic idea of how to apply the saddlepoint method to our model, while omitting further details about the implementation of the method which can be found in Section 2.2 of Zhang *et al.* (2019). For example, we have to factorize the original likelihood function of the model to enable the saddlepoint approximation method to work properly if \mathbf{Y}_1 , a subvector of \mathbf{Y} , contains variables that have zero count. In addition, some tuning is needed for the observed data vector and the matrix \mathbf{A}_1 if \mathbf{A}_1 is not of full row rank.

Variance estimation

Wald confidence intervals (CIs) for the model parameters can be readily constructed via the asymptotic distribution of the maximum likelihood estimator under the approximate model. Simulation studies (results shown below) indicate that this approach is fast and works well in many situations, but may sometimes result in significantly inflated CIs. For example, for some data sets the Wald CIs for some parameters are unreasonably too close to $(0, 1)$ — the whole parameter space. Compared with the corresponding credible intervals produced by the Bayesian method of Bonner *et al.* (2016), those Wald CIs are not reliable because the variances for the estimators produced by the approximate model are too large to be accurate. Alternatively, a more robust way for variance estimation is to generate CIs via a parametric bootstrap method. The bootstrap method works well in all situations we have tested but is slower than the Wald approach.

Results

Simulation study

We conducted simulations to assess the performance of the saddlepoint approximation method applied to the proposed approximate model. We simulated data under the original CJS/BRE model, and then fit the approximate model to the data. Note that estimating α from the data is difficult because the data contain very limited information about the parameter (Link *et al.*,

Table 2: Estimation results obtained using the saddlepoint approximation method based on simulations in two scenarios each with 100 replicates. We assume $T = 4$ capture occasions, $\alpha = 0.9$, and $N_t = 100$ for $t = 1, \dots, T - 1$.

Parameter	True value	Bias%	(MLE > 0.9)%	Wald CI		Bootstrap CI	
				Width	Coverage	Width	Coverage
ϕ_1	0.70	-1.1	2.0	0.41	0.91	0.40	0.94
ϕ_2	0.70	-2.0	4.0	0.43	0.96	0.40	0.96
p_2	0.50	3.7	0.0	0.35	0.95	0.37	0.92
p_3	0.50	0.5	0.0	0.31	0.98	0.32	0.95
$\phi_3 p_4$	0.35	1.5	0.0	0.16	0.95	0.16	0.96
ϕ_1	0.90	-0.6	54.0	0.51	0.88	0.25	0.98
ϕ_2	0.90	0.9	59.0	0.66	0.93	0.26	0.98
p_2	0.50	1.8	0.0	0.26	0.93	0.26	0.94
p_3	0.50	0.0	0.0	0.22	0.94	0.23	1.00
$\phi_3 p_4$	0.45	-1.1	0.0	0.15	0.91	0.15	0.97

2010; Bonner *et al.*, 2016). One common way to handle this is to collect more data (Link *et al.*, 2010). In practice, information about α can be obtained using double observers or double tags (Bonner *et al.*, 2016). For convenience, we set α to be known in our simulations, which does not make a difference to our main points in this paper. Estimating α together with other parameters is possible but may lead to wider confidence intervals for the parameters. Similar to the CJS model, parameters p_T and ϕ_{T-1} cannot be estimated separately under the CJS/BRE model with T capture occasions, but we can estimate the product $p_T \phi_{T-1}$.

As examples, Table 2 presents estimation results from simulations in two scenarios each with 100 replicates. In both scenarios, fitting the approximate model to the simulated data produces almost unbiased inference for all estimable parameters, with approximately nominal coverage of 95% (both Wald and bootstrap) CIs. Note that for the first simulation in Table 2, the differences between the Wald and bootstrap CI widths for all the parameters are negligible, while for the second the Wald CI widths for parameters ϕ_1 and ϕ_2 are at least two times as large as the bootstrap CI widths. We found that this also happened for simulations in some other settings. Inspection of numerous simulation results indicated this happened in cases where the MLE of one or more of

Table 3: Comparisons of the saddlepoint approximation method and the Bayesian approach for the CJS/ BRE model based on simulations in four scenarios each with 50 replicates. The true parameter values for simulations are: $\phi_1 = \phi_2 = \phi_3 = 0.7$ and $p_2 = p_3 = p_4 = 0.6$. Values before and within the parentheses in the fourth and fifth columns represent results for the Wald and bootstrap CIs respectively.

(N_t, α)	Parameter	Saddlepoint			Bayesian		
		Bias%	CI width	CI coverage	Bias%	CI width	CI coverage
(100, 0.8)	ϕ_1	-0.6	0.51 (0.44)	0.90 (0.90)	-2.1	0.41	0.96
	ϕ_2	0.5	0.49 (0.44)	0.92 (0.96)	-0.9	0.40	0.98
	p_2	3.7	0.47 (0.44)	0.90 (0.92)	4.1	0.40	0.94
	p_3	2.1	0.40 (0.42)	1.00 (1.00)	4.6	0.37	0.96
	$\phi_3 p_4$	-0.8	0.19 (0.19)	0.98 (1.00)	-0.1	0.18	0.98
(200, 0.8)	ϕ_1	1.1	0.35 (0.35)	0.92 (0.94)	0.0	0.32	0.90
	ϕ_2	-0.1	0.34 (0.34)	0.98 (0.96)	2.4	0.31	0.96
	p_2	-0.2	0.31 (0.33)	0.96 (0.94)	2.8	0.29	0.94
	p_3	1.2	0.29 (0.30)	1.00 (0.98)	0.5	0.26	0.98
	$\phi_3 p_4$	-0.6	0.14 (0.14)	0.94 (0.94)	-1.0	0.13	0.98
(100, 0.9)	ϕ_1	-1.8	0.37 (0.36)	0.96 (0.92)	-1.8	0.34	0.98
	ϕ_2	0.8	0.35 (0.36)	0.98 (0.98)	2.2	0.34	0.98
	p_2	2.0	0.35 (0.37)	1.00 (0.98)	2.4	0.34	1.00
	p_3	0.7	0.31 (0.32)	1.00 (0.96)	0.0	0.31	0.98
	$\phi_3 p_4$	-1.2	0.16 (0.16)	0.96 (0.94)	-2.3	0.16	0.98
(200, 0.9)	ϕ_1	2.5	0.26 (0.27)	0.94 (0.94)	1.4	0.26	0.90
	ϕ_2	-0.8	0.24 (0.25)	0.94 (0.98)	1.4	0.26	0.96
	p_2	-1.4	0.25 (0.25)	0.92 (0.92)	0.1	0.24	0.98
	p_3	0.1	0.22 (0.23)	0.92 (0.96)	-1.5	0.22	0.94
	$\phi_3 p_4$	-0.3	0.12 (0.12)	0.98 (0.98)	-3.0	0.11	0.90

the parameters was close to the upper boundary of the parameter space (say above 0.9). The percentages of MLEs > 0.9 for ϕ_1 and ϕ_2 are 54% and 59% for the second simulation in Table 2, while the percentages for the first simulation are 2% and 4%. Thus, the MLEs can be regarded as an empirical indicator for us in practice to determine if the bootstrap method is needed for variance estimation.

We also conducted simulations to investigate the influence of fitting the approximate model on parameter estimation. For each simulated data set, we fit our approximate model using the saddlepoint approximation method, and fit the original model using the Bayesian approach of

Bonner *et al.* (2016). Then we compared estimation results from the two methods. Table 3 shows four examples of such comparisons. We can see that both methods generate almost unbiased inference with roughly nominal coverage of 95% confidence (credible) intervals for all estimable parameters. The bootstrap and Wald CI widths are similar for these simulations because the percentages of MLEs > 0.9 are low for all the parameters. The only concern for our proposed method is that in some cases it might produce slightly wider CIs than the Bayesian approach does under the exact model. One such example is the first simulation study shown in Table 3. We find that the performance of the saddlepoint approximation depends on the value of parameter α . Given that other parameters are fixed, the method works better for a larger value of α . For example, the percentage differences between confidence (credible) interval widths obtained from the two methods turn out to be less than 10% for all the parameters in the third simulation study shown in Table 3 when α increases from 0.8 to 0.9. This point can be further illustrated by the second and fourth simulation studies in Table 3. The performance of the approximate method also depends on the components of \mathbf{N} . More individuals captured on each occasion can enhance the performance of the approximate method, which, for example, can be seen from the first and second simulations in Table 3.

A key motivation of developing the approximate model in this paper is that computation times using the saddlepoint approximation for the model can be substantially shorter. On average, it cost less than one second on a 2.3 GHz laptop to fit the approximate model to one simulated data set in the settings of Tables 3, provided that the Wald CIs were generated. It cost about three minutes to generate bootstrap CIs from 1,000 replicates for one data set. In contrast, for the same data set about 24 hours was needed using the Bayesian MCMC algorithm of Bonner *et al.* (2016) under the original CJS/BRE model. This was rather slow because we set the length of the chain to be one million to ensure the effective sample size was at least 1,000 for all parameters.

Example: salamanders data

We consider data from a study of Northern Dusky Salamanders (*Desmognathus fuscus*) conducted at Stephens Road Nature Preserve in Huntersville, North Carolina, USA (Price *et al.*, 2012). Salamanders were sampled twice per month within a 100m section of a stream using capture-recapture methods from October 2005 to November 2010. Sampling methods included turning over cover objects such as rocks and logs, searching leaf litter, and checking beneath coverboards along the stream bank. Each captured individual was uniquely marked by subcutaneous injections of a liquid polymer known as visible implant elastomer (Northwest Marine Technologies, Shaw Island, WA). Visible implant elastomer is widely used and considered a reliable technique in capture-recapture studies of amphibians (Marold, 2001; Bailey *et al.*, 2004); however, some researchers have noted that this technique occasionally results in misidentification of individuals. Misidentification of individuals may occur if marks migrate from the injection area (Marold, 2001; Grant, 2008), or if observers fail to see marks through heavily pigmented skin or misidentify the color of the elastomer (Bailey, 2004; Heemeyer *et al.*, 2007).

Price *et al.* (2012) investigated the effect of drought conditions on the temporary emigration patterns and survival of salamanders using the monthly data. Here, we use yearly samples from the study to illustrate the proposed method for capture-recapture data with misidentification. The numbers of salamanders first captured and marked on occasions 1 to 5 (years 2005 – 2009) form the vector $\mathbf{N} = (62, 377, 361, 244, 495)$. Within the study period 81% of these 1,539 salamanders were captured once only, 18% twice and 1% three times or more. For those captured more than once, 79% were recaptured in the year after they were marked and not seen again. Only 8% were recaptured two or more years after they were marked. In the analysis of Price *et al.* (2012), no misidentification was considered. However, there is some evidence of misidentification in the study; for example one individual was captured in the first two years, not captured in the next three, and then captured again. This is unusual because only 1.5% ($19\% \times 8\%$) of all the marked salamanders were recaptured more than two years after they were initially marked.

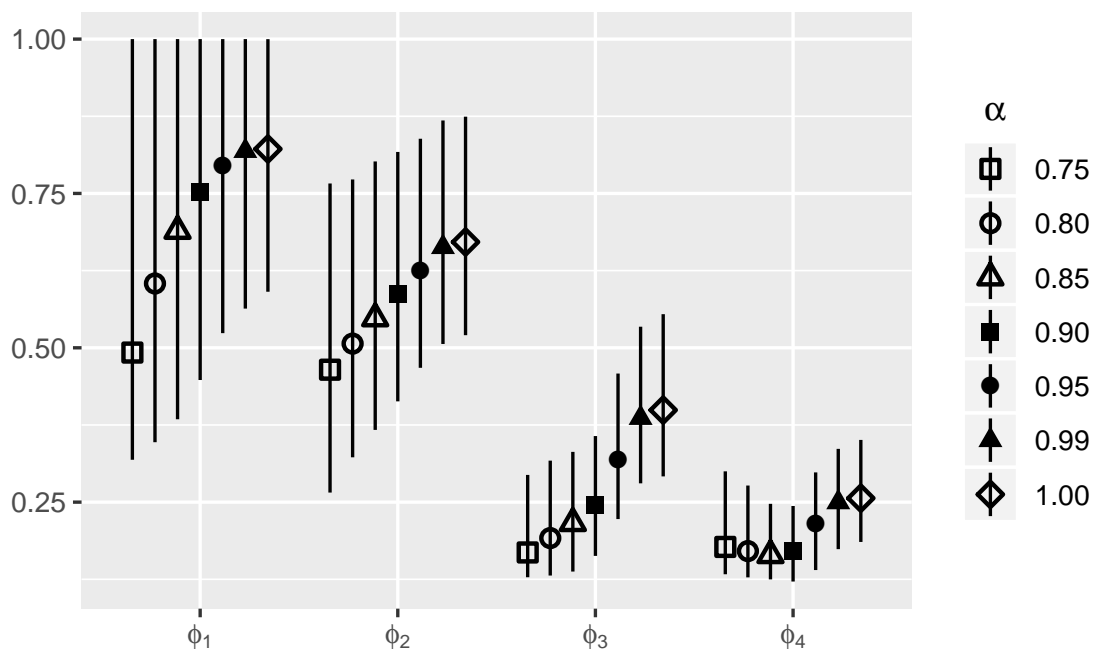


Figure 1: MLEs and 95% bootstrap confidence intervals (profile intervals for $\alpha = 1$) of annual survival probabilities for the salamanders data. When $\alpha = 1$ we fit the exact CJS model using the program MARK (White & Burnham, 1999) implemented via the R package RMark.

There is no direct information available for estimating the parameter α for this data set. To explore the possible effects of misidentification in this example, we fit the approximate model for different values of α and compare the resulting estimates of survival. Here we need to obtain reliable CIs for the parameters by bootstrapping. Figure 1 shows estimation results for survival probabilities. Regardless of the rate of misidentification, there is clear evidence of a decrease in the survival probability based on these models. For all the seven values of α we have explored, the results indicate that the survival probability decreased on average by 17.4% from year 2005 to year 2006, 55.9% from year 2006 to year 2007, and 21.4% from year 2007 to year 2008. It is also obvious from the figure that misidentification has two significant effects on the estimation of survival for the salamanders. First, a lower rate of misidentification (i.e., a larger value of α) increases the estimates of the survival probabilities. The estimates of the survival probabilities for $\alpha = 1$ are between 1.5 and 2.4 times as large as those for $\alpha = 0.75$. Second, a lower rate

of misidentification may also decrease the CI widths for these parameters. This effect is more significant for the initial survival probabilities ϕ_1 and ϕ_2 . The CI widths for parameters ϕ_1 and ϕ_2 for $\alpha = 1$ are 0.60 and 0.71 times as large as those for $\alpha = 0.75$. The CI widths for ϕ_3 and ϕ_4 remain almost the same for different values of α in this example. These results clearly indicate the necessity of taking into account misidentification in analysis of capture-recapture data if animal identification is error-prone. This observation coincides with the conclusions of previous authors such as Link *et al.* (2010) and Bonner *et al.* (2016).

Discussion

In this paper, we have developed a fast approximate method for fitting open-population capture-recapture data with misidentification under the CJS/BRE model. We have demonstrated the promising performance of the proposed method empirically by simulation and by comparisons with the Bayesian approach of Bonner *et al.* (2016). Based on the simulations, we have confidence that the approximation method works well for parameter estimation when most components of \mathbf{N} are large (say 200 or above). In this case, we need not be concerned about whether α is large or not. When it is not the case, α might be regarded as an indicator of the performance of the proposed method. If α is large, the approximation method can still work well when \mathbf{N} consists of small counts. Otherwise quite wide credible intervals will be obtained even if we fit the CJS/BRE model using the Bayesian method. Thus the approximate method works in most realistic situations.

We approximate the true distribution of the latent vector \mathbf{X} by a product-multinomial model, which shares exactly the same mean vector as \mathbf{X} but has a different variance-covariance matrix. While losing some information about the exact distribution of \mathbf{X} , this approximation enables the application of the saddlepoint approximation method to the CJS/BRE model. Our simulations show that misspecifying the variance-covariance matrix does not lead to obvious bias to the maximum likelihood estimators of the parameters. However, the Wald approach performs poorly for variance estimation in some situations as shown in the simulation study. Thus, we recommend the

more robust bootstrap method for variance estimation under the proposed approximate model. As far as we can tell, the only drawback is that it is slower than the Wald approach, though it is still much faster than the Bayesian method. Note that the computational speed can be easily improved, since R code implementing the bootstrap method can be parallelized without much effort. If one uses a computer with multiple cores or can get access to computer clusters, computation times using our approximate method can be further reduced. This is particularly beneficial for fitting data from capture-recapture studies with more capture occasions, for which finding MLEs is slower because there are more parameters and the dimension of the optimization problem is higher.

Due to the salamanders data we are interested in, we consider the CJS model for population dynamics and the BRE model for misidentification in this paper. It is evident that the BRE model can be used with other capture-recapture models for open or closed populations such as model M_t , which forms the basis of model $M_{t,\alpha}$ (Link *et al.*, 2010). We believe that the approximate method proposed here should still be applicable at least in cases where the capture-recapture model is multinomial. However, one should check the performance of the method thoroughly by simulation before putting it into use for real data analysis.

Acknowledgements

We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC), application number 43024-2016.

Authors' Contributions

WZ and SJB conceived the ideas and designed methodology; SJP collected the data; WZ, SJP and SJB analysed the data; WZ led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

Data Availability

Need to confirm this with Steve. Shall we make the data available somewhere for interested readers?

References

- Bailey, L.L. (2004) Evaluating elastomer marking and photo identification methods for terrestrial salamanders: marking effects and observer bias. *Herpetological Review*, **35**, 38–41.
- Bailey, L.L., Simons, T.R. & Pollock, K.H. (2004) Estimating detection probability parameters for Plethodon salamanders using the robust capture-recapture design. *Journal of Wildlife Management*, **68**, 1–13.
- Barndorff-Nielsen, O.E. & Cox, D.R. (1989) *Asymptotic techniques for use in statistics*. London: Chapman and Hall.
- Bonner, S.J., Schofield, M.R., Noren, P. & Price, S.J. (2016) Extending the latent multinomial model with complex error processes and dynamic Markov bases. *The Annals of Applied Statistics*, **10**, 246–263.
- Butler, R.W. (2007) *Saddlepoint approximations with applications*. Cambridge: Cambridge University Press.
- Cormack, R.M. (1964) Estimates of survival from the sighting of marked animals. *Biometrika*, **51**, 429–438.
- Curtis, J.M.R. (2006) Visible implant elastomer color determination, tag visibility, and tag loss: potential sources of error for mark–recapture studies. *North American Journal of Fisheries Management*, **26**, 327–337.

- Daniels, H.E. (1954) Saddlepoint approximations in statistics. *The Annals of Mathematical Statistics*, **25**, 631–650.
- Grant, E.H.C. (2008) Visual implant elastomer mark retention through metamorphosis in amphibian larvae. *Journal of Wildlife Management*, **72**, 1247–1252.
- Heemeyer, J.L., Homyack, J.A. & Haas, C.A. (2007) Retention and readability of visible implant elastomer marks in Eastern Red-backed Salamanders (*Plethodon cinereus*). *Herpetological Review*, **38**, 425–428.
- Jolly, G.M. (1965) Explicit estimates from capture-recapture data with both death and immigration-stochastic model. *Biometrika*, **52**, 225–247.
- Kristensen, K., Nielsen, A., Berg, C.W., Skaug, H.J. & Bell, B.M. (2016) TMB: Automatic differentiation and Laplace approximation. *Journal of Statistical Software*, **70**, 1–21.
- Link, W.A., Yoshizaki, J., Bailey, L.L. & Pollock, K.H. (2010) Uncovering a latent multinomial: Analysis of mark–recapture data with misidentification. *Biometrics*, **66**, 178–185.
- Lugannani, R. & Rice, S. (1980) Saddlepoint approximation for the distribution of the sum of independent random variables. *Advances in Applied Probability*, **12**, 475–490.
- Marold, M.R. (2001) Evaluating visual implant elastomer polymer for marking small, stream-dwelling salamanders. *Herpetological Review*, **32**, 91–92.
- Morrison, T.A., Yoshizaki, J., Nichols, J.D. & Bolger, D.T. (2011) Estimating survival in photographic capture-recapture studies: Overcoming misidentification error. *Methods in Ecology and Evolution*, **2**, 454–463.
- Otis, D.L., Burnham, K.P., White, G.C. & Anderson, D.R. (1978) Statistical inference from capture data on closed animal populations. *Wildlife Monographs*, **62**, 3–135.

- Price, S.J., Browne, R.A. & Dorcas, M.E. (2012) Resistance and resilience of a stream salamander to suprasedasonal drought. *Herpetologica*, **68**, 312–323.
- Schofield, M.R. & Bonner, S.J. (2015) Connecting the latent multinomial. *Biometrics*, **71**, 1070–1080.
- Seber, G.A.F. (1965) A note on the multiple-recapture census. *Biometrika*, **52**, 249–259.
- Vale, R.T.R., Fewster, R.M., Carroll, E.L. & Patenaude, N.J. (2014) Maximum likelihood estimation for model $M_{t,\alpha}$ for capture–recapture data with misidentification. *Biometrics*, **70**, 962–971.
- White, G.C. & Burnham, K.P. (1999) Program MARK: survival estimation from populations of marked animals. *Bird Study*, **46**, S120–S139.
- Wright, J.A., Barker, R.J., Schofield, M.R., Frantz, A.C., Byrom, A.E. & Gleeson, D.M. (2009) Incorporating genotype uncertainty into mark–recapture-type models for estimating abundance using DNA samples. *Biometrics*, **65**, 833–840.
- Zhang, W., Bravington, M.V. & Fewster, R.M. (2019) Fast likelihood-based inference for latent count models using the saddlepoint approximation. *Biometrics*, **75**, 723–733.