BAYES ESTIMATION OF POPULATION SIZE FROM CAPTURE-RECAPTURE MODELS WITH TIME VARIATION AND BEHAVIOR RESPONSE

Shen-Ming Lee[1], Wen-Han Hwang[1] and Li-Hui Huang[2]

[1]*Feng-Chia University and* [2]*the Overseas Chinese Institute of Technology*

Abstract: This paper considers size estimation of a closed population using capture-recapture models when the capture probabilities vary with time (or trapping occasion) and behavior response. A unified approach via the Bayesian framework is proposed to make inferences about the population size for four specific models. Based on data augmentation considerations, we show how Gibbs sampling associated with an adaptive rejection sampling technique can be applied to calculate Bayes estimates in our setting. The prior distributions that we have chosen are all noninformative except as regards the behavior response parameter. A simulation study investigates the performance of the proposed procedure and compares it with the maximum likelihood estimates derived by Chao, Chu and Hsu (2000). The estimates are also applied to capture data of deer mice discussed in the literature. The results show that Gibbs sampling provides a useful inference procedure for estimating population size, particularly when the capture probability is high or the amount of recapture information is sufficient.

Key words and phrases: Bayes estimation, behavior response, capture-recapture model, Gibbs sampling, Markov chain Monte Carlo method, population size, time variation.

1. Introduction

For ecological studies, the first problem usually encountered is the need to know the sizes of various wildlife populations in a given study area. In order to make precise inference of population sizes, a variety of capture-recapture sampling methods are widely used. In this paper, we focus on the problem of size estimation of a closed population using capture-recapture models when capture probabilities vary with time (or trapping occasion) and behavior response. This model is known as M_{tb}. There are three submodels of M_{tb} — namely, models M_t, M_b and M_0. Models M_t and M_b take into account capture probabilities varying with time and behavior response, respectively. Model M_0 assumes every animal has the same capture probability on every capture occasion. These capture-recapture models have been extensively discussed; for review see Otis,

Burnham, White and Anderson (1978), White, Anderson, Burnham and Otis (1982), Seber (1982, 1992), and Pollock (1991).

Within the literature analyzing M_0, M_t and M_b, maximum likelihood estimation (MLE) techniques have been derived; see Darroch (1958) for M_0 and M_t and Zippin (1958) for M_b. Seber and Whale (1970) show that for M_b the MLE of population sizes exists only under special conditions. However, because animals frequently exhibit a behavioral response to being captured, models M_b and M_{tb} are both practical and important in biological and ecological applications. Unfortunately, there is relatively little literature on M_{tb}. Papers by Lloyd (1994), Rexstad and Burnham (1991), and Lee (1996) are exceptions. Otis, Burnham, White and Anderson (1978) shows that the MLE of M_{tb} does not exist since it is unidentifiable without further assumptions on the parameters. A popular assumption was considered by Otis, Burnham, White and Anderson (1978) and Lloyd (1994); namely that the ratio of the recapture probability to the initial capture probability is a constant. Moreover, Lloyd (1994) developed an estimating function to solve the problem. Recently, Chao, Chu and Hsu (2000) derived the MLE under the same assumption. However, the MLE solution often fails to converge when the data is sparse. Therefore, we concentrate on behavior response and present the Bayesian approach using Gibbs sampling, a Markov chain Monte Carlo method.

For the Bayesian approach, Roberts (1967) was the pioneer of population size inference for M_0. Castledine (1981) and Smith (1988, 1991) dealt with the problem for model M_t. They developed numerical methods to calculate the posterior mean of the population size and used normal approximation to obtain the creditable interval. George and Robert (1992) were the first to use the modern Bayesian technique of Gibbs sampling to estimate population size for M_t. Lee and Chen (1998) applied Gibbs sampling to M_b and M_{tb}. Their approach has the advantage of using the Bayesian framework to overcome the identification problem for M_{tb}. However, their method does not use recapture information and leads to unstable estimates. In other words, the performance of their estimates often depends on the chosen prior distributions. In this paper, a Bayesian framework using recapture information for M_{tb} is set up, and Gibbs sampling is used for inference of the population parameters. When there is no behavioral response by the captured animals, the proposed setup can be reduced to the work of Castledine (1981) for M_0 and George and Robert (1992) for M_t. For related work based on the Bayesian approach, see Fienberg, Johnson and Junker (1999) and Basu and Ebrahimi (2001). The major advantage of the Bayesian method via the Gibbs sampling technique is that it can provide sufficient information about the parameters in addition to the point and interval estimates. This is because the posterior distribution samples are available to the user. For

details of Gibbs sampling and discussion of the estimation procedures, see Geman and Geman (1984), Tanner and Wong (1987), Gelfand, Hills, Rancine-Poon and Smith (1990), Tanner (1994), and Gilks, Richardson and Spiegelhalter (1996).

Section 2 formulates the time variation and behavior response models and the Bayesian framework. Section 3 examines real capture-recapture data as an illustrative example. Simulation results are presented to show the performance of the proposed estimators. We give a brief conclusion in Section 4.

2. Bayes Estimates for Models M_{tb}, M_b, M_t and M_0

Let $i = 1, \ldots, N$ index the animals in a closed population, $j = 1, \ldots, t$ index the trapping samples, and P_{ij} be the capture probability of the ith animal in the jth trapping sample. Animals are assumed to act independently. If the animals exhibit a behavior response, then P_{ij} depends on the capture history of the first $j - 1$ samples and can be expressed as

$$P_{ij} = \begin{cases} P_{ij}^* & \text{if the } i\text{th animal has not been caught before the } j\text{th sample;} \\ b_{ij}^* & \text{if the } i\text{th animal has been caught before the } j\text{th sample.} \end{cases} \quad (1)$$

Let X_{ij} equal 1 if the ith animal is caught in the jth sample and 0 otherwise. The likelihood function of the capture-recapture experiment is as follows:

$$
\begin{aligned}
L(N, \boldsymbol{P}|\mathcal{D}) &= \prod_{i=1}^{N} \prod_{j=1}^{t} P_{ij}^{X_{ij}} (1 - P_{ij})^{1-X_{ij}} \\
&= \prod_{i=1}^{N} \prod_{j=1}^{t} P_{ij}^{*X_{ij} I[(\sum_{k=1}^{j-1} X_{ik})=0]} b_{ij}^{*X_{ij} I[(\sum_{k=1}^{j-1} X_{ik})>0]} \\
&\quad \times (1 - P_{ij}^*)^{(1-X_{ij})I[(\sum_{k=1}^{j-1} X_{ik})=0]} (1 - b_{ij}^*)^{(1-X_{ij})I[(\sum_{k=1}^{j-1} X_{ik})>0]}, \quad (2)
\end{aligned}
$$

where $I(\cdot)$ is the usual indicator function, $\boldsymbol{P} = (P_{ij}, i = 1, \ldots, N; j = 1, \ldots, t)$, and $\mathcal{D} = \{X_{ij}, i = 1, \ldots, N; j = 1, \ldots, t\}$.

In this model, information about N is difficult to extract from the data because there are so many parameters. Therefore, the parameter space must be restricted. The most common assumptions are $P_{ij} = P_j$ or $P_{ij} = P$, indicating no behavior response with or without time variation. These models are designated M_t or M_0, respectively, in Otis, Burnham, White and Anderson (1978). Furthermore if there is a behavior response, an intuitive restriction is that $P_{ij}^* = P_j$ and $b_{ij}^* = b_j$. However the parameters are unidentifiable in this situation, as indicated by Otis, Burnham, White and Anderson (1978). We therefore need additional restrictions in order to estimate the parameters. In this paper, we consider one of

$$P_{ij} = P_j \phi^{I(\sum_{k=1}^{j-1} X_{ik}>0)}, \quad (3)$$

$$P_{ij} = P\phi^{I(\sum_{k=1}^{j-1} X_{ik}>0)}, \tag{4}$$

where ϕ represents the behavior response effect. The definition in (3) is equivalent to $b_j = \phi P_j$ and (4) is equivalent to $P_{ij}^* = P$ and $b_{ij}^* = b = \phi P$. If $\phi > 1$ the response is trap-happy, if $\phi < 1$ the response is trap-shy; we denote (3) and (4) by M_{tb} and M_b, respectively. If $\phi = 1$, these models are M_t and M_0 as proposed by Darroch (1958). In the reminder of this section, we give a unified inference procedure for size estimation of a closed population using the Bayesian technique and Gibbs sampling.

2.1. Model M_{tb}

We consider the case that all animals which have not been caught before the jth sample have the same capture probability P_j in the jth sample. The recapture probability for all animals in the jth sample is ϕP_j, where ϕ is the behavior response effect. The structure of P_{ij} can be expressed as (3). The likelihood function for this model reduces to

$$L(N, \boldsymbol{P}, \phi|\mathcal{D}) \propto \frac{N!}{(N-M_{t+1})!} \prod_{j=1}^{t} P_j^{u_j}(1-P_j)^{N-M_{j+1}} \prod_{j=2}^{t} (\phi P_j)^{m_j}(1-\phi P_j)^{M_j-m_j}$$

$$= \frac{N!}{(N-M_{t+1})!} \left\{ \prod_{j=1}^{t} P_j^{n_j}(1-P_j)^{N-M_{j+1}}(1-\phi P_j)^{M_j-m_j} \right\} \phi^{m_\bullet}, \tag{5}$$

where $\boldsymbol{P} = (P_1, \ldots, P_t)$, $m_\bullet = m_2 + \cdots + m_t$, $M_{j+1} = u_1 + \cdots + u_j$ is the number of distinct animals captured prior to the $(j+1)$th sample, $n_j = u_j + m_j$ is the total caught in the jth sample, and m_j and u_j are the number of marked and unmarked animals captured in the jth sample, respectively.

For the Bayesian approach, we need to assign a joint prior distribution for the parameters. We consider an independent prior distribution of the form $\pi(N, \boldsymbol{P}, \phi) = (\prod_{j=1}^{t} \pi(P_j))\pi(\phi)\pi(N)$. Such a prior leads to conditional posteriors as follows:

$$\pi(N \mid \boldsymbol{P}, \phi, \mathcal{D}) \propto \frac{N!}{(N-M_{t+1})!} \left(\prod_{j=1}^{t} (1-P_j)^N \right) \pi(N), \tag{6}$$

$$\pi(P_j \mid \boldsymbol{P}_{(-j)}, N, \phi, \mathcal{D}) \propto P_j^{n_j}(1-P_j)^{N-M_{j+1}}(1-\phi P_j)^{M_j-m_j} \pi(P_j), \tag{7}$$

for $j = 1, \ldots, t$ and

$$\pi(\phi \mid N, \boldsymbol{P}, \mathcal{D}) \propto \phi^{m_\bullet} \prod_{j=2}^{t} (1-\phi P_j)^{M_j-m_j} \pi(\phi), \tag{8}$$

where $\boldsymbol{P}_{(-j)}$ denotes the vector $\boldsymbol{P}$ with P_j deleted. The conditional posterior distribution of N in (6) is the same as that proposed by George and Robert (1992) for M_t.

In this paper, we consider the prior distribution of $\boldsymbol{P}$ to be $\pi(\boldsymbol{P}) = \prod \pi(P_j)$, where $\pi(P_j) = U(0,1)$ and $U(0,1)$ denotes a uniform distribution. The prior distribution of ϕ is $\pi(\phi) = U(\alpha, \beta)$, where $U(\alpha, \beta)$ denotes a uniform distribution with range in (α, β) assumed to be specified in this study. It follows that (7) and (8) reduce to

$$\pi(P_j \mid \boldsymbol{P}_{(-j)}, N, \phi, \mathcal{D}) \propto P_j^{n_j}(1 - P_j)^{N - M_{j+1}}(1 - \phi P_j)^{M_j - m_j}, \qquad (9)$$

for $j = 1, \ldots, t$ and

$$\pi(\phi \mid N, \boldsymbol{P}, \mathcal{D}) \propto \phi^{m \bullet} \prod_{j=2}^{t}(1 - \phi P_j)^{M_j - m_j} I(\alpha < \phi < \beta). \qquad (10)$$

Taking the prior distribution of N as a Jeffrey's prior $\pi(N) = 1/N$, the conditional posterior distribution of N is

$$P(N = n \mid \boldsymbol{P}, \phi, \mathcal{D}) = \binom{n-1}{M_{t+1}-1}\left(1 - \prod_{j=1}^{t}(1 - P_j)\right)^{M_{t+1}}\left(\prod_{j=1}^{t}(1 - P_j)\right)^{N - M_{t+1}}, \quad (11)$$

where $n = M_{t+1}, M_{t+1} + 1, \ldots$. It is easy to recognize that the conditional posterior of N is a negative binomial with parameter $(M_{t+1}, 1 - \prod(1 - P_j))$. Alternatively, for the constant prior of N, the conditional posterior of N is a negative binomial with parameter $(M_{t+1} + 1, 1 - \prod(1 - P_j))$. Notice that the prior distributions of all parameters that we have chosen here are all noninformative except in the case of ϕ. If we have more prior information about the parameters, we can make a more informed choice of prior distributions. For example, we may take $\pi(P_j)$ to be a Beta distribution with parameters γ_1 and γ_2 when we have more information about the capture probability. In this situation, the conditional posterior distribution of P_j becomes

$$\pi(P_j \mid \boldsymbol{P}_{(-j)}, N, \phi, \mathcal{D}) \propto P_j^{n_j + \gamma_1 - 1}(1 - P_j)^{N - M_{j+1} + \gamma_2 - 1}(1 - \phi P_j)^{M_j - m_j}. \quad (12)$$

It is trivial that (9) is a special case of the above equation since $U(0,1)$ is a Beta distribution with $\gamma_1 = 1$ and $\gamma_2 = 1$. Interested readers may refer to Smith (1991) to find examples of other choices.

The prior distribution of ϕ that we have chosen is a constant prior distribution, but it is not a noninformative prior since we need to specify its range. The motivation for this choice was twofold. First, it is theoretically justifiable to take a noninformative prior distribution, a constant prior, for ϕ; however, this

leads to numerical complications in the implementation of the Gibbs sequence. Thus, in practice, it is necessary to restrict the range of ϕ to be between α to β in advance. Such restrictions are not unsual, and there is a common technique for dealing with this limitation. For example, Norris and Pollock (1996) took twice the true population size as the upper limit of the population size estimate in their simulation study. In other words, by taking α small and β large enough, the prior of ϕ can be regarded as noninformative in practice. Second, we also tried a lognormal distribution as the prior of ϕ in the simulation in Section 3. The performance of estimates was slightly inferior to those under the uniform prior given the same prior mean and variance. The advantage of assigning a lognormal prior distribution is that no restriction of ϕ is necessary; however, at least to us, it is a little unnatural to specify the prior mean and variance of ϕ.

We now state the Gibbs sampling procedure in the capture-recapture model. We first generate the initial values of $\boldsymbol{P}^{(0)}$ and $\phi^{(0)}$ from their prior distributions. Then we generate a value of $N^{(0)}$ from (11) with $\boldsymbol{P}$ and ϕ replaced by $\boldsymbol{P}^{(0)}$ and $\phi^{(0)}$. Starting with an initial value of $N^{(0)}$, we produce a 'Gibbs sequence' $\{\boldsymbol{P}^{(k)}, \phi^{(k)}, N^{(k)}\}$, $k = 0, 1, \ldots$, with iteratively simulated sampling from (9), (10) and (11). However, the conditional densities of P_j and ϕ are not identified. We therefore employ adaptive rejection sampling (see Gilks and Wild (1992)) to generate $\boldsymbol{P}$ and ϕ. We only state how to generate ϕ for simplicity. Let $f(\phi)$ denote the conditional densities in (10), suppressing the conditioning variables for simplicity. It is easy to check that $f(\phi)$ is a log-concave function in the range (a, b) where $a = \alpha$ and $b = \min\{\frac{1}{P_1}, \ldots, \frac{1}{P_t}, \beta\}$. The range is defined by the prior limit of ϕ and the restriction that $\phi P_j < 1$ for $j = 1, \ldots, t$. Let $S_n = \{\phi_1, \ldots, \phi_n\}$, where $\phi_1 < \cdots < \phi_n$ denotes a current set of abscissa in the range (a, b) (see below for an example). We then define a piecewise linear function $g_n(\phi)$ from the tangent at S_n to $\ln f(\phi)$, that is, $g_n(\phi) = \min\{\ln f(\phi_j) + \frac{f'(\phi_j)}{f(\phi_j)}(\phi - \phi_j), j = 1, \ldots, n\}$. Because $\ln f(\phi)$ is a concave function, this implies that g_n is a convex envelope of $\ln f(\phi)$. The adaptive rejection sampling employed works as follows.

STEP 1. Specify initial n and S_n.

STEP 2. Generate a point ϕ^* from the distribution proportional to $\exp(g_n(\phi))$.

STEP 3. Accept ϕ^* with probability $f(\phi^*) / \exp(g_n(\phi^*))$; otherwise let $S_{n+1} = S_n \bigcup \{\phi^*\}$ and go back to STEP 2.

In the example and simulation study of this paper, we take the initial value of n as 5 and $S_5 = \{\phi_3 - \frac{1}{10}(b-a), \phi_3 - \frac{3}{100}(b-a), \phi_3, \phi_3 + \frac{3}{100}(b-a), \phi_3 + \frac{1}{10}(b-a)\}$, where $\phi_3 = \frac{m_\bullet}{\sum_{j=2}^t M_j P_j}$. The initial value of ϕ_3 is chosen because $E(m_\bullet) = \phi E(\sum_{j=2}^t M_j P_j)$ and this value is close to the mode of f. Given this setup, the iteration steps for the adaptive rejection procedure usually take one or two steps. We employ a similar method for the parameters P_j. Note that the distribution

proportional to $\exp(g_n(\phi))$ is a piecewise exponential distribution and can be easily simulated.

The Gibbs sequence forms a Markov chain. According to the Ergodic Theorem, if the chain length is sufficiently large, then the sample mean will converge to the posterior mean. Hence, we can estimate the population size and other parameters by the sample mean. We can also use the quantile interval of the simulated data to estimate the creditable intervals of interesting parameters. The initial process is generally unstable due to the wide range of choices for the starting values. In order to avoid the influence of the starting value, we discard the first k terms as the burn-in period and use the remaining terms to estimate the parameters. Therefore, we need to determine the lengths of the process and burn-in period. This is also called the convergence diagnostic, and there are many methods available to achieve this (see Cowles and Bradley (1996)). In this study, we use the multiple chains method developed by Gelman and Rubin (1992). This method, both quite reasonable and easy to handle, is as follows.

First we simulate m parallel chains, each with length $2k$, from different starting points. Then we discard the first k numbers of each chain as the burn-in period and use the remaining part to estimate R, the ratio of between-chain variation to within-chain variation. If the ratio $\hat{R}^{\frac{1}{2}}$ is high, then we believe that further iteration will be necessary. When $\hat{R}^{\frac{1}{2}}$ approaches one, some authors agree that we can continue untill $\hat{R}^{\frac{1}{2}}$ is less than 1.1 (see Gelman (1996)), and we can then terminate the iteration. Finally, we use the sample mean of the remaining part to estimate the parameters and the quantile interval of the remaining part to estimate the creditable interval.

2.2. Model M_b

In this model, we assume that all animals have the same capture probability P in the first capture and the same recapture probability ϕP after the first capture. The likelihood function becomes

$$L(N, P, \phi | \mathcal{D}) \propto \frac{N!}{(N - M_{t+1})!} P^{n_\bullet} (1 - P)^{tN - M_\bullet - M_{t+1}} \phi^{m_\bullet} (1 - \phi P)^{M_\bullet - m_\bullet}, \quad (13)$$

where $M_\bullet = M_2 + \cdots + M_t$ and $n_\bullet = n_1 + \cdots + n_t$. Taking N, P and ϕ to be a priori independent, the conditional posterior distributions are

$$\pi(N \mid P, \phi, \mathcal{D}) \propto \frac{N!}{(N - M_{t+1})!} (1 - P)^{tN} \pi(N), \quad (14)$$

$$\pi(P \mid N, \phi, \mathcal{D}) \propto P^{n_\bullet} (1 - P)^{tN - M_\bullet - M_{t+1}} (1 - \phi P)^{M_\bullet - m_\bullet} \pi(P), \quad (15)$$

$$\pi(\phi \mid N, P, \mathcal{D}) \propto \phi^{m_\bullet} (1 - \phi P)^{M_\bullet - m_\bullet} \pi(\phi). \quad (16)$$

We choose $\pi(P) = U(0,1)$ and $\pi(\phi) = U(\alpha, \beta)$. Subsequently, (15) and (16) reduce to

$$\pi(P|N, \phi, \mathcal{D}) \propto P^{n_\bullet}(1 - P)^{tN - M_\bullet - M_{t+1}}(1 - \phi P)^{M_\bullet - m_\bullet}, \tag{17}$$

$$\pi(\phi|N, P, \mathcal{D}) \propto \phi^{m_\bullet}(1 - \phi P)^{M_\bullet - m_\bullet}. \tag{18}$$

Taking the prior of N to be a constant or a Jeffrey's prior in (14), the conditional posterior of N follows a negative binomial distribution with parameters $(M_{t+1} + 1, 1 - (1 - P)^t)$ or $(M_{t+1}, 1 - (1 - P)^t)$, respectively. Moreover, the logit model on P and $b = \phi P$ can be described as $\alpha = \ln(P/(1 - P)) \sim N(\mu, \sigma^2)$ and $\beta = \ln(b/(1 - b)) \sim N(\nu, \sigma^2)$. In this case, the conditional posterior of α is

$$\pi(\alpha \mid N, \beta, \mathcal{D}) \propto \frac{\exp(\alpha M_{t+1} - \frac{1}{2}(\frac{\alpha - \mu}{\sigma})^2)}{(1 + e^\alpha)^{tN - M_\bullet}}, \tag{19}$$

a log-concave function of α. Therefore, we can also simulate α via adaptive rejection sampling. Based on these conditional posterior distributions, Gibbs sampling can be readily implemented.

2.3. Model M_t

Suppose all animals in the population have the same capture probability P_j in the jth sample. This is model M_t, which has been studied by Darroch (1958), Otis, Burnham, White and Anderson (1978), and George and Robert (1992). In our model, we take $\phi = 1$ in (5) and the likelihood function becomes

$$L(N, \boldsymbol{P}|\mathcal{D}) \propto \frac{N!}{(N - M_{t+1})!} \prod_{j=1}^{t} P_j^{n_j}(1 - P_j)^{N - n_j}. \tag{20}$$

Using the uniform prior $\pi(P_j) = U(0,1)$, the conditional posterior of P_j given N is

$$\pi(\boldsymbol{P}|N, \mathcal{D}) = \prod_{j=1}^{t} Beta(n_j + 1, N - n_j + 1). \tag{21}$$

We emphasize that the above result can also be obtained from (9) by taking $\phi = 1$. The conditional posteriors of N given $\boldsymbol{P}$ under a Jeffery's prior or a constant prior are negative binomial distributions with parameter $(M_{t+1}, 1 - \prod(1 - P_j))$ or $(M_{t+1} + 1, 1 - \prod(1 - P_j))$, respectively. Given the initial values of $\boldsymbol{P}^{(0)}$ and $N^{(0)}$, the Gibbs sequence $\{N^{(k)}, \boldsymbol{P}^{(k)}\}$, $k = 1, 2, \ldots$, can be obtained from the Beta and the negative binomial distribution. If we choose $\pi(P_j) = Beta(\gamma_1, \gamma_2)$ then the conditional posterior of P_j given N is

$$\pi(\boldsymbol{P}|N, \mathcal{D}) = \prod_{j=1}^{t} Beta(n_j + \gamma_1, N - n_j + \gamma_2). \tag{22}$$

Note that the above results are the same as those found by George and Robert (1992), and can be derived from our proposed M_{tb}.

2.4. Model M_0

Model M_0 assumes that all animals in the population have the same capture probability P for all sampling occasions. The model has been studied by Darroch (1958), Otis, Burnham, White and Anderson (1978), and Castledine (1981). We take $\phi = 1$ in (15) and use a Beta prior for P; the conditional posterior of P given N is

$$\pi(P|N, \mathcal{D}) = Beta(n_. + \gamma_1, tN - n_. + \gamma_2). \tag{23}$$

The conditional posterior of N given P is the same as that for M_b assuming a Jeffrey's prior or a constant prior. Therefore P and N can be easily generated.

3. Example and Simulation

In this section, we illustrate the proposed procedure with a real example and a brief simulation study, focusing on inference about the population size.

3.1. Real example

We consider deer mouse data as an illustrative example. These data were collected by S. Hoffman and analyzed by Otis, Burnham, White and Anderson (1978), Chao, Chu and Hsu (2000), and Huggins and Yip (2001). As shown in Table 1, there are $t = 5$ capture occasions; the total number of distinct mice captured is $M_{t+1} = 110$ and the total number captured is $n_. = 283$. We show the results of the likelihood and the Bayesian approaches in Table 2.

The upper part of Table 2 presents the results for the likelihood approach with two rows (UMLE and QMLE) derived directly from Chao, Chu and Hsu (2000). First, the unconditional maximum likelihood estimate (UMLE) is around 161 with a 95% bootstrap confidence interval from 121 to 283. Second, the conditional maximum likelihood estimate (CMLE) is 174 with a 95% confidence interval from 124 to 289. Finally, the quasi maximum likelihood estimate (QMLE) is 152 with a 95% confidence interval from 119 to 229. Note that although the above three estimators have been shown to be asymptotically equivalent when the population size is large enough, their performance is still quite different in this example.

The lower part of Table 2 presents the results for the Bayesian approach. There are four different priors for ϕ chosen here. The prior for N is a Jeffery's prior and the prior for $\boldsymbol{P}$ is an i.i.d. uniform distribution. For each case, we generate five parallel chains from different randomly selected starting points of $\boldsymbol{P}$, ϕ and N. Then, we compute $\hat{R}^{\frac{1}{2}}$ (with respect to N and ϕ) to determine the

burn-in period cutoff k. As we can see from Figure 1, $\hat{R}^{\frac{1}{2}}$ becomes smaller than 1.1 after 100, 500, 500 and 1500 iterations in each case. We therefore fix k at 1500 for simplicity and record the remaining 1500 values in each chain. This way the recorded 7500 values mimic the posterior distribution of N. Our estimate depends on the choice of the prior for ϕ, which ranges from 127 to 155.

Table 1. Capture-recapture counts of deer mice.

occasion		1	2	3	4	5
had been caught	(M_j)	0	37	68	77	98
newly caught	(u_j)	37	31	9	21	12
marked caught	(m_j)	0	23	49	44	57
total caught	(n_j)	37	54	58	65	69

Table 2. Estimates of deer mouse population size.

method	$\hat{N}$	$\hat{SE}$	95% CI	$\hat{\phi}$	95% CI of ϕ
UMLE*	161	42.8	(121, 283)	3.19	
CMLE	174	45.3	(124, 289)	3.63	(1.67, 7.37)
QMLE*	152	29.9	(119, 229)	2.87	
(α, β)					
(0.5, 2)	127	6.0	(116, 139)	1.76	(1.36, 1.99)
(0.8, 3)	138	11.6	(119, 164)	2.25	(1.54, 2.94)
(1, 5)	150	21.4	(121, 203)	2.61	(1.64, 4.51)
(1, 8)	155	32.3	(122, 251)	2.99	(1.65, 6.16)

* indicates results derived from Chao, Chu and Hsu (2000) directly.

It is not easy to chose a prior distribution if no other information is known. Here we provide a naive trial-and-error procedure. When we adopt the Bayes method to estimate the population size N and behavior response ϕ, we first limit ϕ to range from 0.5 to 2 (a reasonable range in general situations) and we get a 95% creditable interval for ϕ ranging from 1.36 to 1.99. Therefore the upper limit of the creditable interval for ϕ (1.99) is too close to the value presumed (2). Next, we adjust the prior for ϕ from 0.8 to 3 and the creditable interval is found to range from 1.54 to 2.94. Again it is too close to the prior upper limit of 3. We repeat this procedure until the range of the creditable interval is not too close to either side of the prior limits of ϕ. It turn out this occurs with the third choice of limits (ϕ between 1 and 5). If we release the prior limit of ϕ to be from 1 to 8 (a very wide range for ϕ), we find that the outcome is similar but the creditable interval is wider. Hence, for the deer mouse data, we believe the population size is around 150 with a 95% creditable interval from 120 to 200. Our result is similar to QMLE but with a narrower interval estimate.

Note that such a trial-and-error method has no theoretical justification; the judgement of whether or not the creditable interval limits are too close to the
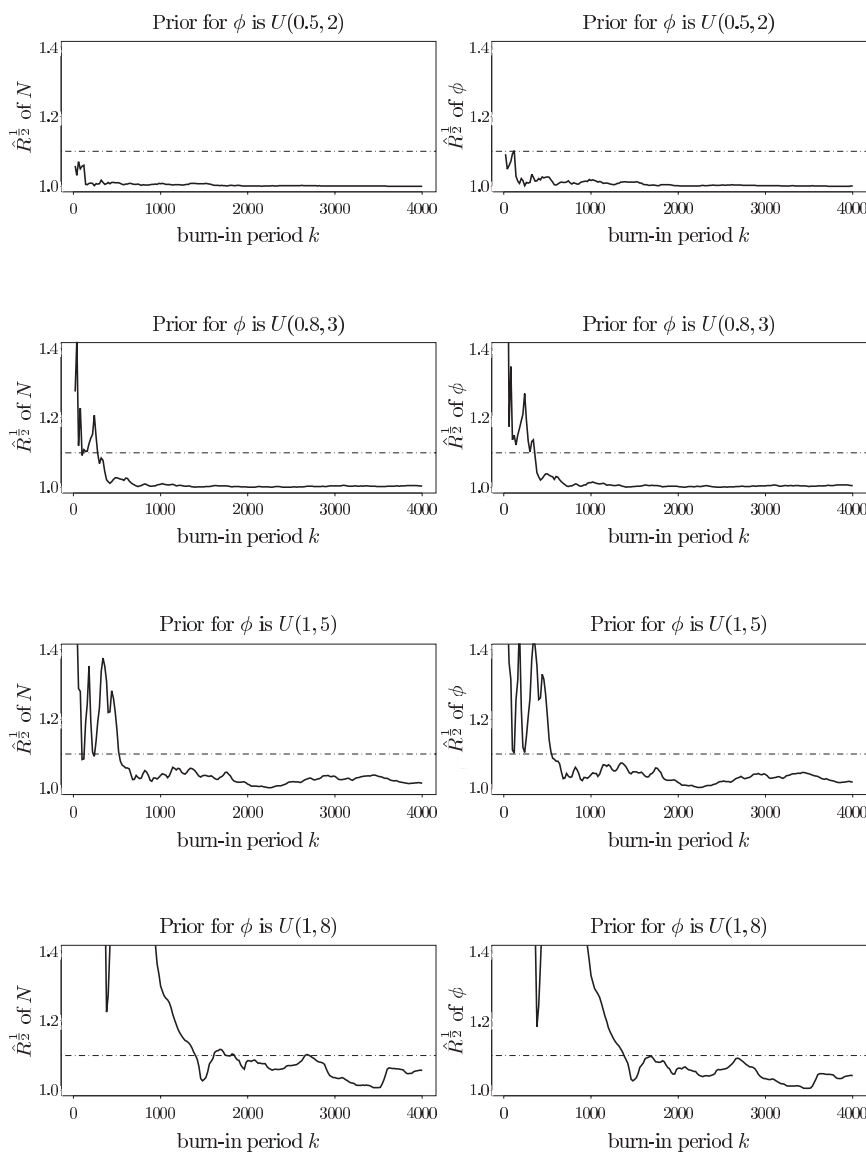

Figure 1. Plot of $\hat{R}^{\frac{1}{2}}$ with respect to the Gibbs sequences of N (left hand side) and ϕ (right hand side) under different prior distributions for ϕ. The baseline (dotted line) is 1.1.

prior limits is highly subjective. However it seems to work well when the data is rich enough (e.g., there is a high capture probability or wealth of recapture in-

formation). In this situation, the data indicate the right direction of incremental experimentation with prior limits. This phenomenon of subjective but educated guesswork is not uncommon in empirical analyses.

3.2. Simulation study

In this subsection, we carry out a simulation study to observe the performance of the proposed method and to compare it with the likelihood approach. Here we focus on M_{tb} because the other three models are special cases of it. The true population size N is fixed at 200 and there are $t = 5$ capture occasions. We consider the following eight combinations of four time effects of $\boldsymbol{P}$ and two values of ϕ; the four time effects of $\boldsymbol{P}$ are $(0.17, 0.15, 0.12, 0.1, 0.12)$, $(0.2, 0.17, 0.25, 0.15, 0.12)$, $(0.3, 0.25, 0.15, 0.2, 0.3)$ and $(0.38, 0.22, 0.34, 0.3, 0.44)$, in which the capture probabilities range from low to high and $E(M_{t+1})$ ranges from 100 to 175, with approximate increments of 25 animals. The two values of ϕ are 1.5 and 0.8 which represent trap-happy and trap-shy responses, respectively. These eight combinations were taken from Chao, Chu and Hsu (2000). The prior for N is a Jeffery's prior and the prior for $\boldsymbol{P}$ is the i.i.d. $U(0, 1)$. The prior for ϕ is a $U(\alpha, \beta)$ distribution where (α, β) is selected as follows:

$$(\alpha, \beta) = \begin{cases} (0.2, 1), (0.2, 2), (0.5, 1), (0.5, 2) & \text{for} \quad \phi = 0.8; \\ (0.5, 2), (1, 2), (1, 3), (1, 5) & \text{for} \quad \phi = 1.5. \end{cases}$$

Here we compare the likelihood approach studied by Chao, Chu and Hsu (2000) with the Bayesian estimating procedure. Note that there are three estimates (CMLE, UMLE and QMLE) derived in Chao, Chu and Hsu (2000). From the extensive simulation study of Chao, Chu and Hsu (1998), CMLE had the largest divergence rate among the three estimates. However, CMLE also had the smallest bias and appropriate coverage probability in the case of low capture probability. Moreover, all three estimates are similar in the case of high capture probability. Hence we only continue comparison with CMLE in this study.

For each combination of time effect $\boldsymbol{P}$ and behavior response ϕ, 200 data sets were generated. Then for each generated data set, the CMLE and the Bayes estimates under four different sets of priors were calculated. The Bayes estimates were given by Gibbs sampling of five parallel chains with 1000 iterations generated but only the last 500 iterations kept. Here we fixed the burn-in period cutoff at 500 for simplicity after observation of the performance of $\hat{R}^{\frac{1}{2}}$. The resulting 200 estimates of population size were averaged to give the "Average estimate" in Tables 3-10. Based on the 200 estimates, the sample s.e., the sample root mean squared error (RMSE), and the coverage percentage of the true N were calculated. We also calculated the average of the estimated s.e. and 95%

confidence intervals (CI) for CMLE based on 200 bootstrap replications. The average s.e. and 95% quantile creditable intervals (CI) of the marginal posterior distribution for Gibbs sampling are presented. In addition, the CMLE divergence rate is shown in the last column. Note that, when the CMLE fails to converge (the solution is larger than $7M_{t+1}$ or it doesn't reach a stable value after 1000 iterations), then we generate another data set until 200 successful estimates are completed. Tables 3-6 and Tables 7-10 show the simulation results for $\phi = 0.8$ and $\phi = 1.5$, respectively. In both series of tables, we also list the average number of distinct individuals captured ($\bar{M}_{t+1}$) and the total number of individuals captured ($\bar{n}$) in the experiment.

Table 3 considers the case of a trap-shy response and a low capture probability. There are 102 distinct animals out of 127 captures on average. All estimates underestimate the true population size because there is little recapture information in this case. The first two estimates using the Bayes method $((\alpha, \beta) = (0.2, 1), (0.2, 2))$ are poor since the prior left hand limit of ϕ is too small. As we increase the prior left hand limit of ϕ to 0.5, the performance improves. As we see in Table 3, the last two Bayes estimates have both smaller RMSEs and a more nearly accurate coverage probability than CMLE. Note that the CMLE divergence rate in this case is over 40%.

Table 3. $N = 200$, $P = (0.17\ 0.15\ 0.12\ 0.10\ 0.12)$, $\phi = 0.8$, $\bar{M}_{t+1} = 102$, $\bar{n}_. = 127$.

Method	Average estimate	Sample SE	Average* SE	Sample RMSE	Coverage (%)	95%CI^{**}	Divergence rate (%)
CMLE	156	48.31	39.73	64.98	69.5	101-226	43.18
BAYES							
(0.2, 1)	126	11.57	19.16	74.40	17.0	105 - 177	-
(0.2, 2)	129	12.09	23.95	71.50	40.5	105 - 194	-
(0.5, 1)	164	20.09	25.95	40.99	77.5	127 - 228	-
(0.5, 2)	171	21.60	36.63	35.38	96.0	128 - 269	-

   * indicates the average bootstrap s.e. for CMLE and the average sample posterior s.e. for the Bayesian approach.
  ** indicates the average 95% bootstrap confidence interval for CMLE and the average 95% equal-tails creditable interval for the Bayesian approach.

Table 4 considers the case of a moderately low capture probability. There are 126 distinct animals out of 167 captures on average. The CMLE is nearly unbiased but still has a divergence rate of nearly 20%. The Bayes estimates all have a negative bias but smaller RMSEs than CMLE. From the performances of the RMSE and the coverage probability, the last two Bayes estimates are superior to CMLE.

Tables 5 and 6 consider the case of high capture probabilities. The Bayes estimates tend to be a little negatively biased but they all have smaller RMSEs and shorter creditable intervals than the CMLE confidence intervals.

Table 4. $N = 200$, $P = (0.20\ 0.17\ 0.25\ 0.15\ 0.12)$, $\phi = 0.8$, $\bar{M}_{t+1} = 126$, $\bar{n}_{.} = 167$

Method	Average estimate	Sample SE	Average SE	Sample RMSE	Coverage (%)	95%CI	Divergence rate (%)
CMLE	199	60.03	48.76	59.89	90.0	128 - 285	19.03
BAYES							
(0.2, 1)	153	12.09	19.31	48.33	56.0	129 - 202	-
(0.2, 2)	158	15.00	26.62	44.28	75.0	129 - 231	-
(0.5, 1)	172	12.90	20.38	29.93	89.0	144 - 222	-
(0.5, 2)	181	15.00	31.10	23.72	98.0	145 - 265	-

Table 5. $N = 200$, $P = (0.30\ 0.25\ 0.15\ 0.20\ 0.30)$, $\phi = 0.8$, $\bar{M}_{t+1} = 150$, $\bar{n}_{.} = 223$.

Method	Average estimate	Sample SE	Average SE	Sample RMSE	Coverage (%)	95%CI	Divergence rate (%)
CMLE	211	55.77	46.43	56.74	98.0	151 - 324	10.31
BAYES							
(0.2, 1)	173	10.36	16.18	28.31	80.0	152 - 213	-
(0.2, 2)	179	14.25	24.61	24.65	90.0	152 - 246	-
(0.5, 1)	183	9.62	15.50	18.90	94.0	161 - 221	-
(0.5, 2)	193	13.11	26.25	14.79	99.0	162 - 264	-

Table 6. $N = 200$, $P = (0.38\ 0.22\ 0.34\ 0.30\ 0.44)$, $\phi = 0.8$, $\bar{M}_{t+1} = 175$, $\bar{n}_{.} = 303$.

Method	Average estimate	Sample SE	Average SE	Sample RMSE	Coverage (%)	95%CI	Divergence rate (%)
CMLE	209	35.32	33.42	36.59	96.0	176 - 301	0
BAYES							
(0.2, 1)	192	8.96	10.24	11.87	89.5	177 - 215	
(0.2, 2)	199	14.82	17.66	14.79	94.5	177 - 245	
(0.5, 1)	193	8.51	9.95	10.78	92.5	178 - 216	
(0.5, 2)	201	14.06	17.60	14.08	96.5	179 - 246	

As indicated by a referee, in Tables 3 and 4 the Bayes estimates for the priors $U(0.2, 1)$ and $U(0.5, 1)$ are significally different; that is also true for $U(0.2, 2)$ and $U(0.5, 2)$. In other words, a small change in the left hand prior limit leads to large discrepancies in the estimates. In the case of high capture probabilities (Tables 5 and 6) such discrepancies tend to diminish. This phenomenon again shows that the choice of priors has a large impact when there is a low capture probability or sparse data. We should be very careful about choosing prior distributions in

such situations.

We next consider the trap-happy case in Tables 7-10. The performance of CMLE in Tables 7-10 is better than in Tables 3-6 since there is more recapture information. However, the existence of more recapture information does not seem to improve the rate of divergence very much. Bayes estimates, on the other hand, perform well once we are sure that the behavior response is larger than 1. It does not seem difficult in practice to decide whether a capture-recapture experiment response should be considered trap-happy or trap-shy in models of M_b and M_{tb}. Having already noticed the diminished sensitivity to the choice of priors, the performance of estimates in the trap-happy case is otherwise similar to that in the trap-shy case.

Table 7. $N = 200$, $P = (0.17\ 0.15\ 0.12\ 0.10\ 0.12)$, $\phi = 1.5$, $\bar{M}_{t+1} = 102$, $\bar{n}_. = 149$.

Method	Average estimate	Sample SE	Average SE	Sample RMSE	Coverage (%)	95%CI	Divergence rate (%)
CMLE	163	45.42	41.27	58.02	89.5	101 - 255	38.46
BAYES							
(0.5, 2)	136	12.79	21.72	64.84	40.5	111 - 194	-
(1 , 2)	173	18.05	25.78	32.27	92.5	136 - 236	-
(1 , 3)	179	19.22	33.46	28.29	98.0	136 - 267	-
(1 , 5)	183	20.45	40.42	26.23	98.0	137 - 292	-

Table 8. $N = 200$, $P = (0.20\ 0.17\ 0.25\ 0.15\ 0.12)$, $\phi = 1.5$, $\bar{M}_{t+1} = 126$, $\bar{n}_. = 205$.

Method	Average estimate	Sample SE	Average SE	Sample RMSE	Coverage (%)	95%CI	Divergence rate (%)
CMLE	209	57.35	51.35	57.94	95.5	134-325	18.37
BAYES							
(0.5, 2)	163	14.42	21.58	39.30	75.0	135 - 217	-
(1 , 2)	182	14.70	21.32	22.59	93.5	152 - 234	-
(1 , 3)	189	16.24	28.64	19.63	98.0	153 - 264	-
(1 , 5)	193	17.37	35.53	18.62	99.5	153 - 290	-

Table 9. $N = 200$, $P = (0.30\ 0.25\ 0.15\ 0.20\ 0.30)$, $\phi = 1.5$, $\bar{M}_{t+1} = 150$, $\bar{n}_. = 285$.

Method	Average estimate	Sample SE	Average SE	Sample RMSE	Coverage (%)	95%CI	Divergence rate (%)
CMLE	211	48.46	43.82	49.62	96.5	154 - 318	10.31
BAYES							
(0.5, 2)	178	12.66	16.97	25.16	87.5	154 - 218	-
(1 , 2)	188	10.48	15.86	15.88	98.5	165 - 225	-
(1 , 3)	194	12.97	22.81	14.21	99.5	165 - 253	-
(1 , 5)	197	15.03	28.40	15.27	99.0	165 - 276	-

Table 10. $N = 200$, $P = (0.38\ 0.22\ 0.34\ 0.30\ 0.44)$, $\phi = 1.5$, $\bar{M}_{t+1} = 175$, $\bar{n}_{\cdot} = 415$.

Method	Average estimate	Sample SE	Average SE	Sample RMSE	Coverage (%)	95%CI	Divergence rate (%)
CMLE	207	28.87	28.81	29.79	94.0	179 - 286	0.00
BAYES							
(0.5, 2)	194	9.34	10.37	10.69	92.0	179 - 218	-
(1 , 2)	196	8.75	10.21	9.53	95.0	180 - 219	-
(1 , 3)	200	12.42	14.64	12.41	96.0	181 - 237	-
(1 , 5)	211	31.31	26.57	33.29	95.5	182 - 279	-

4. Concluding Remarks

We have compared the results of the Bayesian approach with the likelihood approach, a comparison not previously made for capture-recapture models. The advantages of the Bayesian approach are the following.

1. The estimates do not diverge, as may be occur using MLE in situations of either a low capture probability or sparse data.

2. As seen in the simulation study, our estimates performed well and required only minimal prior information about ϕ.

3. If the prior is chosen appropriately, we may obtain a smaller RMSE and a shorter CI than in found using MLE.

4. This method easily extends to more complex models, such as those considering the latent structure in the aforementioned models or the open model.

The main reasons that the Bayesian approach is criticized relates to how the prior distributions are selected. Opponents of the Bayesian approach argue that the method of selecting prior distributions is too subjective. Proponents of the Bayesian approach have proposed methods including the two-stage prior, empirical Bayes, Bayes empirical Bayes, and noninformative prior to avoid the problem of making a subjective choice. In the capture-recapture models considered in this paper, we select the noninformative prior approach. In fact, we merely assume that ϕ is within a reasonable scope which can be determined by biologists or from the experiences of experts. On the other hand, we may also employ the trial-and-error method adopted in our example, at least in the cases of a high capture probability or adequate recapture information, to discover a reasonable range for ϕ. We can then use the method in this paper to improve estimation of population sizes.

References

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

Castledine, B. (1981). A Bayesian analysis of multiple-recapture sampling for a closed population. *Biometrika* **46**, 167-174.

Chao, A., Chu, W. and Hsu, C. H. (1998). Capture-recapture when time and behavioral response affect capture probabilities. Technical report, Institute of Statistics, National Tsing Hua University, Hsin-Chu, Taiwan.

Chao, A., Chu, W. and Hsu, C. H. (2000). Capture-recapture when time and behavioral response affect capture probabilities. *Biometrics* **56**, 427-433.

Cowles, M. K. and Carlin, B. P. (1996). Markov Chain Monte Carlo convergence diagnostics: A comparative review. *J. Amer. Statist. Assoc.* **91**, 883-904.

Darroch, J. (1958). The multiple-recapture census, I: Estimation of a closed population. *Biometrika* **64**, 435-447.

Fienberg, S. E., Johnson, M. S. and Junker, B. W. (1999). Classical multilevel and Bayesian approaches to population size estimation using multiple lists. *J. Roy. Statist. Soc. Ser. A* **162**, 383-405.

Gelfand, A. E., Hills, S. E., Racine-Poon, A. and Smith, A. F. M. (1990). Illustration of Bayesian inference in normal data models using Gibbs sampling. *J. Amer. Statist. Assoc.* **85**, 972-985.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequence (with discussion). *Statist. Sci.* **13**, 163-185.

Gelman, A. (1996). *Inference and monitoring convergence.* In *Markov Chain Monte Carlo in Practice* (Edited by W. R. Gilks, S. Richardson and D. J. Spiegelhalter). Chapman and Hall, London.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. on Patt. Anal. and Mach. Intell.* **6**, 721-741.

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

Gilks, W. R. and Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Appl. Statist.* **41**, 337-348.

Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in Practice.* Chapman and Hall, London.

Huggins, R. M. and Yip, P. S. F. (2001). A note on nonparametric inference for capture-recapture experiments with heterogeneous probabilities. *Statist. Sinica* **11**, 843-853.

Lee, S. M. (1996). Estimating population size for capture-recapture data when capture probabilities vary by time, behavior and individual animal. *Comm. Statist. Simulation Comput.* **25**, 431-457.

Lee, S. M. and Chen, C. W. S. (1998). Bayesian inference of population size for behavioral response models. *Statist. Sinica* **8**, 1233-1247.

Lloyd, C. J. (1994). Efficiency of martingale methods in recapture studies. *Biometrika* **81**, 305-315.

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

Otis, D. L., Burnham, K. P., White, G. C. and Anderson, D. R. (1978). Statistical inference for capture data on closed animal populations. *Wildlife Monographs* **62**. University of Kentucky, Louisville.

Pollock, K. H. (1991). Modeling capture, recapture, and removal statistics for estimation of demographic parameters for fish and wildlife populations: past, present, and future. *J. Amer. Statist. Assoc.* **86**, 225-238.

Rexstad, E. and Burnham, K. P. (1991). *User's Guide for Interactive Program Caputre.* Colorado Cooperative Fish and Wildlife Research Unit, Fort Collins.

Roberts, H. V. (1967). Informative stopping rules and inferences about population size. *J. Amer. Statist. Assoc.* **62**, 763-775.

Seber, G. F. A. (1982). *The Estimation of Animal Abundance.* 2nd edition. Griffin, London.

Seber, G. F. A. (1992). A review of estimating animal abundance II. *Internat. Statist. Rev.* **60**, 129-166.

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

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

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

Tanner, M. (1994). *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions.* Springer-Verlag, New York.

Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation (with discussion). *J. Amer. Statist. Assoc.* **82**, 528-550.

White, G. C., Anderson, D. R., Burnham, K. P. and Otis, D. L. (1982). Capture-recapture and removal methods for sampling closed populations. Los Alamos National Lab. Publication LA-8787-NERP.

Zippin, C. (1958). The removal method of population estimation. *J. Wild. Man.* **22**, 82-90.

Feng Chia University, 100 Wenhwa Road, Seatwen, Taichung, Taiwan.

E-mail: smlee@stat.fcu.edu.tw

Department of Statistics, Feng-Chia University, Taichung, 40724, Taiwan.

E-mail: whhwang@fcu.edu.tw

Department of International Trade, the Overseas Chinese Institute of Technology, Taichung, Taiwan.