Acta Mathematicae Applicatae Sinica, English Series
Vol. 32, No. 3 (2016) 659–668
DOI: 10.1007/s10255-016-0611-8
http://www.ApplMath.com.cn & www.SpringerLink.com

Acta Mathematicae Applicatae Sinica, English Series

Population Size Estimation with Covariate Values Missing Non-ignorable

Li-ping LIU[1,†], Zhi-chao GUO[1], Xiao-gang DUAN[2]

[1]LMAM and CSS, School of Mathematical Sciences, Peking University, Beijing 100871, China
[2]Beijing Normal University, Beijing 100875, China (Email: liping@math.pku.edu.cn)

Abstract The main purpose of this paper is using capture-recapture data to estimate the population size when some covariate values are missing, possibly non-ignorable. Conditional likelihood method is adopted, with a sub-model describing various missing mechanisms. The derived estimate is proved to be asymptotically normal, and simulation studies via a version of EM algorithm show that it is approximately unbiased. The proposed method is applied to a real example, and the result is compared with previous ones.

Keywords capture-recapture; conditional likelihood; EM algorithm; missing non-ignorable
2000 MR Subject Classification 62N01; 62N02

1 Introduction

As a branch of statistics, capture-recapture has been widely used in biological and ecological sciences. Its primary interest is to estimate the population size of certain species of wildlife, but it is also widely used in the investigation of sensitive groups of human being, as well as in software reliability, epidemiology and other fields (see [1,9,10]).

It is well-known that heterogeneity between individuals in the population usually exists, and ignoring it can result biased estimation. Heterogeneity can partly be explained by the some observed individual covariate, such as sex, weight, etc, partly by other reasons[13]. But in practice, some covariate values are often found to be missing, due to the difficulty in recording the covariete, or an inappropriate experiment design.

Missing mechanism is categorized into 3 kinds[8]: missing completely at random (MCAR), missing at random (MAR), and missing non-ignorable (MNI), the last kind means that whether the covariate value is missing depends on itself. Simply discarding those individuals with missing covariate is not only a loss of information, but also making the estimator biased. In recent years there are lots of works concerning MCAR and MAR in statistical studies, and also some advances in capture-recapture[11,12], but very few is done towards MNI case, and existing data augmentation and imputing methods are suitable to MCAR and MAR only. The main reason is that, in the MNI case, the problem is often non-identifiable. But in capture-recapture studies, we find that it is identifiable, because for recaptured individuals, "repeated measurement" occurs, as if in longitudinal studies.

The main purpose of this work is to establish a sub-model which can deal with various missing mechanisms, and to embed it into existing capture-recapture model, so that all parameter estimates can be derived. The sub-model is wished to be natural and easily interpreted. Conditional likelihood method[5] is adopted to avoid the presence of un-observed individuals in the likelihood function. A version of EM Algorithm is given to handle the practical computa-

Manuscript received April 23, 2012. Revised October 29, 2014.
Supported in part by the National Natural Science Foundation of China under Grant No. 11171006.
†Corresponding author.

tion concerning missing covariate values and to get conditional maximum likelihood estimates (CMLE).

In the next section the model is introduced, estimates for model parameters, as well as for population size, based on conditional likelihood approach are derived, and asymptotic normality of the estimates are established. In Section 3 some simulation experiments are conducted to test the performance of the proposed method and to compare it with existing method. The method is applied to a real example of the Hong Kong Mai Po data in Section 4, and the paper is concluded with a short discussion in Section 5.

2 Inference Procedure

In this work we study the situation where only a bivariate covariate (sex) is involved for simplicity, although more complicated situations, such as multi-covariate cases, are possible.

2.1 Notations, Model Assumptions and Likelihoods

Consider a continuous time capture-recapture experiment conducted in a time interval $[0,\tau]$, captured animals are marked and released back to the population immediately. Suppose the population is closed, in which ν individuals behave independently. Denote by $\delta_i = 1$ if individual i is caught at least once, and 0 otherwise. Let Z_i indicate, by 1 versus 0, if individual i is male or female, and ω_i be its observational indicator, so $\omega_i = 1$ means Z_i is recorded, and $\omega_i = 0$ means Z_i is missing. Assume Z_i follows a binomial distribution $B(1,p)$, where p is a parameter representing the proportion of males in the population. If there are n animals with $\delta_i = 1$, among which m individuals have complete covariate values and the others do not, for convenience we relabel the first m individuals as the captured ones with Z_i recorded, the following $n-m$ captured but with Z_i missing, and the remaining $\nu - n$ individuals not observed. Finally, we use the counting process $N_i(\cdot)$ to describe the capture history of individual i, where $N_i(t)$ is the number of times that individual i be caught up to time t.

The well-known Cox regression model[2] is used to characterize the counting process $N_i(t)$. Assume that there is no trap-response, then the hazard function of it is given by

$$\lambda(Z_i) = \exp(\beta_0 + \beta_1 Z_i),$$

where $\exp(\beta_0)$ is the baseline hazard, and β_1 is the regression parameter for Z_i.

Because the missing mechanism might be non-ignorable, the missing probabilities for males and females could be different. Let q_m and q_f be the probabilities that the covariate values of male animals and female animals are missing, respectively, then ω_i follows a binomial distribution $B(1, 1 - Z_i q_m - (1-Z_i)q_f)$.

Let θ represents all the parameters above, $\theta = (\beta_0, \beta_1, p, q_m, q_f)^T$. For an observed individual i, if $i \le m$, i.e., its covariate value is not missing, its contribution to the likelihood is

$$\begin{aligned} L_i =& P(N_i(\cdot)|Z_i,\omega_i,\theta)\cdot P(Z_i,\omega_i|\theta) \\ =& \Big[\prod_{0\le t\le\tau} \lambda(Z_i)^{\Delta N_i(t)} \exp\Big(-\int_0^\tau \lambda(Z_i)dt\Big)\cdot p^{Z_i}(1-p)^{1-Z_i}(1-q_m)^{Z_i}(1-q_f)^{1-Z_i}\Big]^{\omega_i} \\ =& \big\{\exp(\beta_0+\beta_1 Z_i)^{N_i(\tau)}\exp[-\tau\exp(\beta_0+\beta_1 Z_i)]\cdot p^{Z_i}(1-p)^{1-Z_i}(1-q_m)^{Z_i}(1-q_f)^{1-Z_i}\big\}^{\omega_i}, \end{aligned}$$

but if $m < i \le n$, Z_i has to be integrated out, so its contribution is

$$
\begin{aligned}
L_i =& \sum_{Z_i=0}^{1} P(N_i(\cdot)|Z_i,\omega_i,\theta)\cdot P(Z_i,\omega_i|\theta)\\
=&\Big[\sum_{Z_i=0}^{1}\prod_{0\le t\le\tau}\lambda(Z_i)^{\Delta N_i(t)}\exp\Big(-\int_0^\tau \lambda(Z_i)dt\Big)\cdot p^{Z_i}(1-p)^{1-Z_i}q_m^{Z_i}q_f^{1-Z_i}\Big]^{1-\omega_i}\\
=&\big\{pq_m\cdot\exp(\beta_0+\beta_1)^{N_i(\tau)}\exp\big[-\tau\exp(\beta_0+\beta_1)\big]\\
&+(1-p)q_f\cdot\exp(\beta_0)^{N_i(\tau)}\exp\big[-\tau\exp(\beta_0)\big]\big\}^{1-\omega_i},
\end{aligned}
$$

and for an unobserved individual i, its contribution is simply $L_i = P(\delta_i = 0)$. Therefore the full likelihood function is given by

$$L(\theta|\text{data}) = \prod_{i=1}^{n} L_i^{\delta_i}\prod_{i=n+1}^{\nu} L_i^{1-\delta_i}.$$

This likelihood function contains too many unobserved individuals, so we adopt the conditional likelihood approach[5], which is based on observed individuals only, and is given by

$$L_C(\theta|\text{data}) = \prod_{i=1}^{n}\Big(\frac{L_i}{\pi_0}\Big)^{\delta_i}, \tag{1}$$

where $\pi_0 = p\big\{1-\exp[-\tau\exp(\beta_0+\beta_1)]\big\}+(1-p)\big\{1-\exp[-\tau\exp(\beta_0)]\big\}$ is the average probability of a generic individual being caught at least once.

2.2 Estimation of Model Parameters

CMLE can be obtained by maximizing (1), however, it is too complicated, so EM algorithm will be used to solve the problem. For doing this, the augmented likelihood function containing both observed data and missing covariates is given by

$$
\begin{aligned}
L_A =& \prod_{i=1}^{n}\frac{\exp(\beta_0+\beta_1 Z_i)^{N_i(\tau)}\exp[-\tau\exp(\beta_0+\beta_1 Z_i)]}{\pi_0}\\
&\cdot p^{Z_i}(1-p)^{1-Z_i}q_m^{Z_i(1-\omega_i)}q_f^{(1-Z_i)(1-\omega_i)}(1-q_m)^{Z_i\omega_i}(1-q_f)^{(1-Z_i)\omega_i}.
\end{aligned}
$$

Starting from $\theta^{(0)}$, EM algorithm will be conducted until $\theta^{(k)}$ converges.

In the E-step, $Q(\theta|\theta^{(k)}) = E(\log L_A|\theta^{(k)})$ is calculated as follows.

$$
\begin{aligned}
\log L_A =& \sum_{i=1}^{n}[N_i(\tau)\beta_0 + N_i(\tau)\beta_1 Z_i - \tau\exp(\beta_0+\beta_1 Z_i) + Z_i\log p\\
&+(1-Z_i)\log(1-p) + Z_i(1-\omega_i)\log q_m + (1-Z_i)(1-\omega_i)\log q_f\\
&+\omega_i Z_i\log(1-q_m) + \omega_i(1-Z_i)\log(1-q_f) - \log\pi_0].
\end{aligned}
$$

Define $B_{i0}^{(k)} = \exp[\beta_0^{(k)}N_i(\tau)-\tau\exp\beta_0^{(k)}]$ and $B_{i1}^{(k)} = \exp\big[N_i(\tau)(\beta_0^{(k)}+\beta_1^{(k)})-\tau\exp(\beta_0^{(k)}+\beta_1^{(k)})\big]$,

$$
\begin{aligned}
W_{i1}(\theta^{(k)}) =& P(Z_i=1|\omega_i=0,N_i(\tau),\theta^{(k)}) = \frac{P(Z_i=1,\omega_i=0,N_i(\tau),\theta^{(k)})}{P(\omega_i=0,N_i(\tau),\theta^{(k)})}\\
=&\frac{p^{(k)}{q_m}^{(k)}B_{i1}^{(k)}}{p^{(k)}{q_m}^{(k)}B_{i1}^{(k)}+(1-p^{(k)}){q_f}^{(k)}B_{i0}^{(k)}}
\end{aligned}
$$

and $W_{i0}(\theta^{(k)}) = P(Z_i = 0|\omega_i = 0,\ N_i(\tau), \theta^{(k)}) = 1 - W_{i1}(\theta^{(k)})$, then

$$\begin{aligned}Q(\theta|\theta^{(k)}) = &\sum_{i=1}^{m}[N_i(\tau)\beta_0 + N_i(\tau)\beta_1 Z_i - \tau\exp(\beta_0+\beta_1 Z_i) + Z_i\log p\\&+ (1-Z_i)\log(1-p) + Z_i\log(1-q_m) + (1-Z_i)\log(1-q_f) - \log\pi_0]\\&+\sum_{i=m+1}^{n}[N_i(\tau)\beta_0 + N_i(\tau)\beta_1 W_{i1} - \tau\exp(\beta_0+\beta_1)W_{i1} - \tau\exp(\beta_0)W_{i0}\\&+ W_{i1}\log p + W_{i0}\log(1-p) + W_{i1}\log q_m + W_{i0}\log q_f - \log\pi_0].\end{aligned}$$

In the M-step, we need to maximize $Q(\theta|\theta^{(k)})$ with respect θ to obtain $\theta^{(k+1)}$. Direct derivation tells us that

$$\begin{aligned}\frac{\partial Q(\theta|\theta^{(k)})}{\partial q_m} &= \sum_{i=1}^{m}\Big(-\frac{Z_i}{1-q_m}\Big) + \sum_{i=m+1}^{n}\frac{W_{i1}}{q_m},\\\frac{\partial Q(\theta|\theta^{(k)})}{\partial q_f} &= \sum_{i=1}^{m}\Big(-\frac{1-Z_i}{1-q_f}\Big) + \sum_{i=m+1}^{n}\frac{W_{i0}}{q_f}.\end{aligned}$$

Notice that the two equations above only contain q_m, q_f, equating them to 0, we get

$$q_m^{(k+1)} = \frac{\sum\limits_{i=m+1}^{n} W_{i1}(\theta^{(k)})}{\sum\limits_{i=1}^{m} Z_i + \sum\limits_{i=m+1}^{n} W_{i1}(\theta^{(k)})}$$

and

$$q_f^{(k+1)} = \frac{\sum\limits_{i=m+1}^{n} W_{i0}(\theta^{(k)})}{n - \sum\limits_{i=1}^{m} Z_i - \sum\limits_{i=m+1}^{n} W_{i1}(\theta^{(k)})}.$$

To update the other parameters β_0, β_1, and p, Newton-Raphson method is used to solve the equations $\dfrac{\partial Q}{\partial\beta_0} = 0$, $\dfrac{\partial Q}{\partial\beta_1} = 0$, and $\dfrac{\partial Q}{\partial p} = 0$. For calculating the first and second order derivatives of Q, we first calculate

$$\begin{aligned}\frac{\partial\pi_0}{\partial\beta_0} &= p\tau\exp(\beta_0+\beta_1)\exp[-\tau\exp(\beta_0+\beta_1)] + (1-p)\tau\exp(\beta_0)\exp[-\tau\exp(\beta_0)],\\\frac{\partial\pi_0}{\partial\beta_1} &= p\tau\exp(\beta_0+\beta_1)\exp[-\tau\exp(\beta_0+\beta_1)],\\\frac{\partial\pi_0}{\partial p} &= \exp[-\tau\exp(\beta_0)] - \exp[-\tau\exp(\beta_0+\beta_1)],\\\frac{\partial^2\pi_0}{\partial\beta_0^2} &= p\tau[\exp(\beta_0+\beta_1) - \tau\exp(2\beta_0+2\beta_1)]\exp[-\tau\exp(\beta_0+\beta_1)]\\&\quad + (1-p)\tau[\exp(\beta_0) - \tau\exp(2\beta_0)]\exp[-\tau\exp(\beta_0)],\\\frac{\partial^2\pi_0}{\partial\beta_0\partial\beta_1} &= p\tau[\exp(\beta_0+\beta_1) - \tau\exp(2\beta_0+2\beta_1)]\exp[-\tau\exp(\beta_0+\beta_1)],\\\frac{\partial^2\pi_0}{\partial\beta_0\partial p} &= \tau\exp(\beta_0+\beta_1)\exp[-\tau\exp(\beta_0+\beta_1)] - \tau\exp(\beta_0)\exp[-\tau\exp(\beta_0)],\end{aligned}$$

$$\frac{\partial^2\pi_0}{\partial\beta_1^2}=\frac{\partial^2\pi_0}{\partial\beta_0\partial\beta_1},\qquad \frac{\partial^2\pi_0}{\partial\beta_1\partial p}=\tau\exp(\beta_0+\beta_1)\exp[-\tau\exp(\beta_0+\beta_1)],\qquad \frac{\partial^2\pi_0}{\partial p^2}=0.$$

Therefore

$$\begin{aligned}
\frac{\partial Q(\theta|\theta^{(k)})}{\partial\beta_0} =& \sum_{i=1}^{m}[N_i(\tau)-\tau\exp(\beta_0+\beta_1 Z_i)] \\
&+\sum_{i=m+1}^{n}[N_i(\tau)-\tau W_{i1}\exp(\beta_0+\beta_1)-\tau W_{i0}\exp(\beta_0)]-\frac{n}{\pi_0}\frac{\partial\pi_0}{\partial\beta_0}, \\
\frac{\partial Q(\theta|\theta^{(k)})}{\partial\beta_1} =& \sum_{i=1}^{m}[N_i(\tau)Z_i-\tau Z_i\exp(\beta_0+\beta_1 Z_i)] \\
&+\sum_{i=m+1}^{n}[N_i(\tau)W_{i1}-\tau W_{i1}\exp(\beta_0+\beta_1)]-\frac{n}{\pi_0}\frac{\partial\pi_0}{\partial\beta_1}, \\
\frac{\partial Q(\theta|\theta^{(k)})}{\partial p} =& \sum_{i=1}^{m}\Big(\frac{Z_i}{p}-\frac{1-Z_i}{1-p}\Big)+\sum_{i=m+1}^{n}\Big(\frac{W_{i1}}{p}-\frac{W_{i0}}{1-p}\Big)-\frac{n}{\pi_0}\frac{\partial\pi_0}{\partial p}, \\
\frac{\partial^2 Q(\theta|\theta^{(k)})}{\partial\beta_0^2} =& \sum_{i=1}^{m}[-\tau\exp(\beta_0+\beta_1 Z_i)]+\sum_{i=m+1}^{n}[-\tau W_{i1}\exp(\beta_0+\beta_1)-\tau W_{i0}\exp(\beta_0)] \\
&+\frac{n}{\pi_0^2}\Big(\frac{\partial\pi_0}{\partial\beta_0}\Big)^2-\frac{n}{\pi_0}\frac{\partial^2\pi_0}{\partial\beta_0^2}, \\
\frac{\partial^2 Q(\theta|\theta^{(k)})}{\partial\beta_0\partial\beta_1} =& \sum_{i=1}^{m}[-\tau Z_i\exp(\beta_0+\beta_1 Z_i)]+\sum_{i=m+1}^{n}[-\tau W_{i1}\exp(\beta_0+\beta_1)] \\
&+\frac{n}{\pi_0^2}\frac{\partial\pi_0}{\partial\beta_0}\frac{\partial\pi_0}{\partial\beta_1}-\frac{n}{\pi_0}\frac{\partial^2\pi_0}{\partial\beta_0\partial\beta_1}, \\
\frac{\partial^2 Q(\theta|\theta^{(k)})}{\partial\beta_0\partial p} =& \frac{n}{\pi_0^2}\frac{\partial\pi_0}{\partial\beta_0}\frac{\partial\pi_0}{\partial p}-\frac{n}{\pi_0}\frac{\partial^2\pi_0}{\partial\beta_0\partial p}, \\
\frac{\partial^2 Q(\theta|\theta^{(k)})}{\partial\beta_1^2} =& \sum_{i=1}^{m}[-\tau Z_i^2\exp(\beta_0+\beta_1 Z_i)]+\sum_{i=m+1}^{n}[-\tau W_{i1}\exp(\beta_0+\beta_1)] \\
&+\frac{n}{\pi_0^2}\Big(\frac{\partial\pi_0}{\partial\beta_1}\Big)^2-\frac{n}{\pi_0}\frac{\partial^2\pi_0}{\partial\beta_1^2}, \\
\frac{\partial^2 Q(\theta|\theta^{(k)})}{\partial\beta_1\partial p} =& \frac{n}{\pi_0^2}\frac{\partial\pi_0}{\partial\beta_1}\frac{\partial\pi_0}{\partial p}-\frac{n}{\pi_0}\frac{\partial^2\pi_0}{\partial\beta_1\partial p}, \\
\frac{\partial^2 Q(\theta|\theta^{(k)})}{\partial p^2} =& \sum_{i=1}^{m}\Big[-\frac{Z_i}{p^2}-\frac{1-Z_i}{(1-p)^2}\Big]+\sum_{i=m+1}^{n}\Big[-\frac{W_{i1}}{p^2}-\frac{W_{i0}}{(1-p)^2}\Big]+\frac{n}{\pi_0^2}\Big(\frac{\partial\pi_0}{\partial p}\Big)^2-\frac{n}{\pi_0}\frac{\partial^2\pi_0}{\partial p^2}.
\end{aligned}$$

Now $\beta_0^{(k+1)},\beta_1^{(k+1)},p^{(k+1)}$ is obtained by simply repeating the procedure

$$\begin{pmatrix}\beta_0^{(new)}\\ \beta_1^{(new)}\\ p^{(new)}\end{pmatrix}=\begin{pmatrix}\beta_0^{(old)}\\ \beta_1^{(old)}\\ p^{(old)}\end{pmatrix}-\begin{pmatrix}\frac{\partial^2 Q}{\partial\beta_0^2} & \frac{\partial^2 Q}{\partial\beta_0\partial\beta_1} & \frac{\partial^2 Q}{\partial\beta_0\partial p}\\ \frac{\partial^2 Q}{\partial\beta_0\partial\beta_1} & \frac{\partial^2 Q}{\partial\beta_1^2} & \frac{\partial^2 Q}{\partial\beta_1\partial p}\\ \frac{\partial^2 Q}{\partial\beta_0\partial p} & \frac{\partial^2 Q}{\partial\beta_1\partial p} & \frac{\partial^2 Q}{\partial p^2}\end{pmatrix}_{\theta^{(k)}}^{-1} * \begin{pmatrix}\frac{\partial Q}{\partial\beta_0}\\ \frac{\partial Q}{\partial\beta_1}\\ \frac{\partial Q}{\partial p}\end{pmatrix}_{\theta^{(k)}}.$$

until $\beta_0^{(new)}, \beta_1^{(new)}, p^{(new)}$ converges.

Repeat E-step and M-step until $||\theta^{(k+1)} - \theta^{(k)}||$ is sufficiently small, then $\theta^{(k)}$ converges to the CMLE $\widehat{\theta}$.

2.3 Estimation of Population Size

The famous Horvitz-Thompson method[4] is used to estimate the population size ν. If θ is known, then it is given by

$$\widehat{\nu}(\theta) = \sum_{i=1}^{\nu} \frac{\delta_i}{\pi_0(\theta)} = \sum_{i=1}^{n} \frac{1}{\pi_0(\theta)} = \frac{n}{\pi_0(\theta)},$$

which is unbiased. Based on the result of previous subsection, the estimator actually used is given by plug-in the parameter estimate $\widehat{\theta}$, i.e., $\widehat{\nu} = \widehat{\nu}(\widehat{\theta})$.

It is not difficult to show that $E\frac{\partial \log(L_C)}{\partial \theta} = 0$, and based on the central limit theorem, $\frac{1}{\sqrt{\nu}}\frac{\partial \log(L_C)}{\partial \theta}$ converges in distribution to a normal distribution $N(0, \Sigma)$, where Σ is a 5×5 positive definite matrix. Its elements are calculated and are available from the authors.

Using the Law of Large Numbers, $\lim\limits_{\nu\to\infty} \frac{1}{\nu}\frac{\partial^2 \log(L_C)}{\partial \theta^2} = \widetilde{\Sigma}$ almost surely. It can be shown that $\widetilde{\Sigma} = -\Sigma$. By applying Taylor's series expansion, we have

$$\frac{\partial \log(L_C)}{\partial \theta}\Big|_{\theta} - \frac{\partial \log(L_C)}{\partial \theta}\Big|_{\widehat{\theta}} \approx -\frac{\partial^2 \log(L_C)}{\partial \theta^2}\Big|_{\theta}(\widehat{\theta} - \theta),$$

so

$$\sqrt{\nu}(\widehat{\theta} - \theta) \approx -\nu\Big(\frac{\partial^2 \log(L_C)}{\partial \theta^2}\Big)^{-1}\frac{1}{\sqrt{\nu}}\frac{\partial \log(L_C)}{\partial \theta} \xrightarrow{d} N(0, \Sigma^{-1}).$$

Because

$$\begin{aligned}
\widehat{\nu}(\widehat{\theta}) - \widehat{\nu}(\theta) \approx& \Big[\frac{\partial \widehat{\nu}(\theta)}{\partial \theta}\Big]^T (\widehat{\theta} - \theta) = -\Big[\frac{n\partial \pi_0(\theta)/\partial \theta}{\pi_0^2(\theta)}\Big]^T (\widehat{\theta} - \theta),\\
\frac{\widehat{\nu}(\widehat{\theta}) - \nu}{\sqrt{\nu}} =& \frac{\widehat{\nu}(\widehat{\theta}) - \widehat{\nu}(\theta)}{\sqrt{\nu}} + \frac{\widehat{\nu}(\theta) - \nu}{\sqrt{\nu}}\\
\approx& -\frac{n}{\nu}\Big[\frac{\partial \pi_0(\theta)/\partial \theta}{\pi_0^2(\theta)}\Big]^T \sqrt{\nu}(\widehat{\theta} - \theta) + \frac{1}{\sqrt{\nu}}\sum_{i=1}^{\nu}\Big[\frac{\delta_i}{\pi_0(\theta)} - 1\Big]\\
\approx& -\Big[\frac{\partial \pi_0(\theta)/\partial \theta}{\pi_0(\theta)}\Big]^T \sqrt{\nu}(\widehat{\theta} - \theta) + \frac{1}{\sqrt{\nu}}\sum_{i=1}^{\nu}\Big[\frac{\delta_i}{\pi_0(\theta)} - 1\Big]\\
\approx& \Big[\frac{\partial \pi_0(\theta)/\partial \theta}{\pi_0(\theta)}\Big]^T \nu\Big(\frac{\partial^2 \log(L_C)}{\partial \theta^2}\Big)^{-1}\frac{1}{\sqrt{\nu}}\frac{\partial \log(L_C)}{\partial \theta} + \frac{1}{\sqrt{\nu}}\sum_{i=1}^{\nu}\Big[\frac{\delta_i}{\pi_0(\theta)} - 1\Big].
\end{aligned}$$

Now

$$\begin{aligned}
&\mathrm{Cov}\Big(\frac{\partial \log(L_C)}{\partial \beta_0}, \frac{\delta_i}{\pi_0(\theta)} - 1\Big)\\
=& \frac{1}{\pi_0(\theta)} E\delta_i \frac{\partial \log(L_C)}{\partial \beta_0} - E\frac{\partial \log(L_C)}{\partial \beta_0}\\
=& \frac{1}{\pi_0(\theta)}\Big[\sum_{j\neq i}\frac{1}{\nu}E\frac{\partial \log(L_C)}{\partial \beta_0}\cdot E\delta_i + \frac{1}{\nu}E\frac{\partial \log(L_C)}{\partial \beta_0}\Big] = 0,
\end{aligned}$$

so

$$\mathrm{Cov}\Big(\frac{1}{\sqrt{\nu}}\frac{\partial\log(L_C)}{\partial\beta_0}, \frac{1}{\sqrt{\nu}}\sum_{i=1}^{\nu}\Big[\frac{\delta_i}{\pi_0(\theta)}-1\Big]\Big) = \frac{1}{\nu}\sum_{i=1}^{\nu}\mathrm{Cov}\Big(\frac{\partial\log(L_C)}{\partial\beta_0}, \frac{\delta_i}{\pi_0(\theta)}-1\Big) = 0.$$

It is straight forward that $\frac{1}{\sqrt{\nu}}\sum\limits_{i=1}^{\nu}\big[\frac{\delta_i}{\pi_0(\theta)}-1\big]$ converges in distribution to $N\big(0, \frac{1-\pi_0(\theta)}{\pi_0(\theta)}\big)$, therefore $\frac{\widehat{\nu}(\widehat{\theta})-\nu}{\sqrt{\nu}}$ converges in distribution to a normal distribution $N(0, V)$, where

$$V = \Big[\frac{\partial\pi_0(\theta)/\partial\theta}{\pi_0(\theta)}\Big]^T\Sigma^{-1}\Big[\frac{\partial\pi_0(\theta)/\partial\theta}{\pi_0(\theta)}\Big] + \frac{1-\pi_0(\theta)}{\pi_0(\theta)}.$$

The estimate of V is given by

$$\widehat{V} = \Big[\frac{\partial\pi_0(\widehat{\theta})/\partial\theta}{\pi_0(\widehat{\theta})}\Big]^T\widehat{\Sigma}^{-1}\Big[\frac{\partial\pi_0(\widehat{\theta})/\partial\theta}{\pi_0(\widehat{\theta})}\Big] + \frac{1-\pi_0(\widehat{\theta})}{\pi_0(\widehat{\theta})},$$

where $\widehat{\Sigma}^{-1} = \Sigma^{-1}(\widehat{\theta})$ is the estimate for the asymptotic variance of $\sqrt{\nu}(\widehat{\theta}-\theta)$. So the 95% confidence interval of ν is $\big[\widehat{\nu}(\widehat{\theta}) - 1.96\sqrt{\widehat{V}\cdot\widehat{\nu}(\widehat{\theta})}, \widehat{\nu}(\widehat{\theta}) + 1.96\sqrt{\widehat{V}\cdot\widehat{\nu}(\widehat{\theta})}\big]$.

3 Simulation

To test the performance of the proposed method, a series of simulation experiments are conducted. Each experiment is repeated 5000 times, with parameter values being $\beta_0 = -0.4$, $\beta_1 = 0.6$, $p = 0.5$, $q_m = 0.4$ and $q_f = 0.2$. When $\tau = 1$, the average capture probabilities for males and females are 0.7052 and 0.4885, respectively, and the overall capture probability is 0.5968; when $\tau = 2$, capture probabilities for males and females are 0.9131 and 0.7383, respectively, and the overall capture probability is 0.8257. The population size ν is taken to be 200, 400 and 800, respectively.

The averaged biases of the estimates (with standard errors in parenthesis), the means of the estimated standard deviations and the coverage rates of the 95% confidence intervals are summarized in Table 1 and Table 2. From the simulation results, it can be seen that, the averaged biases are small, which means that the proposed method provides asymptotically unbiased estimates for all model parameters. As the population size getting larger, and/or the time length of the experiment getting longer, the estimates $\widehat{\theta}$ getting closer to the true value θ, and the standard error getting smaller. The averaged standard deviation estimates are generally close to the standard errors of the 5000 replications, and the coverage rates of the 95% confidence intervals are close to their nominal values. When the model parameters take other values, similar results are obtained and are not presented for brevity.

Table 1. Simulation Results with $\tau = 1$

ν		$\widehat{\beta_0}$	$\widehat{\beta_1}$	$\widehat{p}$	$\widehat{q_m}$	$\widehat{q_f}$
200	bias	–0.092 (0.307)	0.093 (0.329)	-0.013 (0.107)	-0.020 (0.133)	-0.038 (0.179)
	mean of std	0.302	0.331	0.153	0.215	0.251
	CR of 95% CI	97.02	97.46	95.2	97.06	97.6
400	bias	–0.041 (0.200)	0.039 (0.219)	-0.004 (0.087)	-0.013 (0.112)	-0.027 (0.145)
	mean of std	0.206	0.227	0.103	0.142	0.169
	CR of 95% CI	96.46	96.58	94.94	96.46	96.96
800	bias	–0.015 (0.144)	0.012 (0.159)	-0.001 (0.066)	-0.009 (0.088)	-0.014 (0.107)
	mean of std	0.144	0.160	0.071	0.097	0.111
	CR of 95% CI	95.34	95.08	95.6	96	96.22

Table 2. Simulation Results with $\tau = 2$

ν		$\widehat{\beta_0}$	$\widehat{\beta_1}$	$\widehat{p}$	$\widehat{q_m}$	$\widehat{q_f}$
200	bias	–0.017 (0.149)	0.014 (0.166)	0.000 (0.076)	-0.011 (0.106)	-0.018 (0.117)
	mean of std	0.152	0.169	0.083	0.122	0.126
	CR of 95% CI	96.08	96.14	96.24	95.66	96.7
400	bias	–0.011 (0.107)	0.010 (0.119)	0.000 (0.054)	-0.006 (0.078)	-0.008 (0.082)
	mean of std	0.107	0.119	0.056	0.080	0.081
	CR of 95% CI	95.28	95.26	95.34	95	95.2
800	bias	–0.004 (0.074)	0.003 (0.083)	-0.001 (0.038)	-0.004 (0.054)	-0.004 (0.056)
	mean of std	0.075	0.084	0.038	0.055	0.055
	CR of 95% CI	95.36	95.56	94.86	95.16	95.38

We compare our proposed method with existing method[6] where only the individuals without missing data are used, and the results are presented in Table 3, together with the corresponding population size estimates. Some parameter values are still $\beta_0 = -0.4$, $\beta_1 = 0.6$ and $p = 0.5$, ν is taken to be 400, but parameters q_m and q_f take different values. Because the estimates for β_0 and β_1 are all reasonable, and the estimates for q_m and q_f are not available in existing method, only the estimates for p and ν are listed.

Table 3. Comparison with Existing Method

			Existing Method		Proposed Method	
τ	(q_m, q_f)		$\widehat{p}$	$\widehat{\nu}$	$\widehat{p}$	$\widehat{\nu}$
1	(0,0)	bias	-0.002 (0.044)	5.322 (31.029)	-0.002 (0.044)	5.322 (31.029)
		mean of std	0.045	31.056	0.045	31.056
		CR of 95% CI	95.46	95.56	95.46	95.56
1	(0.2,0.4)	bias	0.067 (0.056)	-114.324 (27.398)	-0.002 (0.079)	8.218 (34.369)
		mean of std	0.055	25.727	0.104	33.960
		CR of 95% CI	73.98	6.18	95.6	95.82
1	(0.3,0.3)	bias	-0.003 (0.054)	-114.511 (28.433)	-0.008 (0.084)	7.688 (33.925)
		mean of std	0.054	26.618	0.115	33.447
		CR of 95% CI	94.56	6.48	94.54	95.86
1	(0.4,0.2)	bias	-0.074 (0.052)	-114.619 (29.176)	-0.016 (0.081)	7.833 (33.600)
		mean of std	0.051	27.540	0.121	33.238
		CR of 95% CI	70.280	7.180	94.020	96.180
2	(0,0)	bias	0.000 (0.029)	1.080 (12.944)	0.000 (0.029)	1.080 (12.944)
		mean of std	0.030	12.841	0.030	12.841
		CR of 95% CI	95.26	94.96	95.26	94.96
2	(0.2,0.4)	bias	0.070 (0.036)	-118.802 (13.773)	0.000 (0.051)	1.735 (13.262)
		mean of std	0.036	10.284	0.055	13.443
		CR of 95% CI	49.84	0	96.14	95.56
2	(0.3,0.3)	bias	-0.001 (0.036)	-118.718 (14.100)	0.000 (0.054)	1.590 (13.248)
		mean of std	0.035	10.796	0.059	13.333
		CR of 95% CI	94.52	0	95.9	95.4
2	(0.4,0.2)	bias	-0.072 (0.035)	-118.649 (14.434)	-0.003 (0.053)	1.646 (13.166)
		mean of std	0.034	11.301	0.064	13.286
		CR of 95% CI	45	0	96.28	95.64

It is observed from Table 3 that, whenever the problem of missing covariate value arises, the population size estimate of existing method has a non-negligible negative bias, and the coverage rates of the confidence intervals are very low, meanwhile, the proposed method gives satisfactory results, with small positive biases which is common in capture-recapture studies. This is mainly because existing method uses only part of the full data. If the missing mechanism is MAR, i.e., $q_m = q_f$, then for the estimation of p, both methods gives reasonable results, although the precision of existing method seems better. When the missing mechanism is non-ignorable, i.e., $q_m \neq q_f$, then for the existing method, the bias in the estimation of p is not negligible, and the coverage rates of the confidence intervals are poor. It is not surprising that if $q_m < q_f$, the bias is positive, and the opposite is true.

4 Mai Po Bird Example

We apply the proposed method to a real example which is also used in other related works, e.g.,[3]. Detailed information of the data can be found in [7,14]. Of all the 131 captured birds, 49 birds' sex were recorded correctly, of which 32 are males and 17 are females. The estimation results using our proposed method are displayed in Table 4, with standard deviation in parenthesis.

Table 4. Mai Po Data Results

β_0	β_1	p	q_m	q_f	ν
–1.196 (0.377)	1.129 (0.435)	0.159 (0.062)	0.199 (0.271)	0.813 (0.049)	414.758 (102.108)

It is verified again that male birds are more catchable than females, which is the conclusion of [3,7,14]. The population size estimates coincides approximately with the second estimate of [3], which uses weight as the only covariate, and with previous results. Our analysis shows strongly that the missing mechanism for the covariate sex is non-ignorable, therefore the conjecture of [3] is confirmed. Furthermore, the estimate for p is updated to be even smaller, which should be more reasonable, because in [3], it is assumed that the covariate is missing at random.

5 Discussion

In this paper we have successfully solved the problem of missing covariate values, possibly non-ignorable, by establishing a sub-model, which can deal with various missing mechanisms. Combined with the conditional likelihood approach, it is assured that information in the observed data is completely used in the statistical inference procedure.

We only consider one bivariate covariate case here, so the binomial distribution assumption is the only natural choice. Theoretically the proposed method can be extended to multi-covariate situation, where continuous covariate may also exist. However, more parameters should be included, and even non-parametric structure might be constructed, so strong and disputable assumptions on covariate distribution and its link to the missing mechanism will be introduced to the sub-model, which is not the main theme of the present work. The assumption that the baseline hazard rate $\lambda_0(t)$ takes the form $\exp(\beta_0)$ can be also weakened to it is time-dependent, but be known, up to a constant, and if it is completely unknown, then the estimation of model parameters is usually via a martingale estimating equation, which uses only recaptured data, so is less efficient. Finally, it is apparent that a discrete time version of this study is straight forward.

References

[1] Chao, A. An overview of closed capture-recapture models. *Journal of Agricultural, Biological, and Environmental Statistics*, 6: 158–175 (2001)

[2] Cox, D.R. Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society:*

Series B (Methodological), 34: 187–220 (1972)

[3] Duan, X., Liu, L., Zhao, P. Estimation of covariate distribution with capture-recapture data. *Communications in Statistics-Theory and Methods*, 38: 3705–3712 (2009)

[4] Horvitz, D.G., Thompson, D.J. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47: 663–685 (1952)

[5] Huggins, R.M. On the statistical analysis of capture experiments. *Biometrika*, 76: 133–140 (1989)

[6] Hwang, W.H., Chao, A. Continuous-time capture-recapture models with covariates. *Statistica Sinica*, 12: 1115–1131 (2002)

[7] Lin, D.Y., Yip, P.S.F. Parametric regression models for continuous time removal and recapture studies. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61: 401–411 (1999)

[8] Little, R.J.A., Rubin, D.B. Statistical Analysis with Missing Data, 2nd edition. Wiley, New York, 2002

[9] Pollock, K.H. Modeling capture, recapture, and removal statistics for estimation of demographic parameters for fish and wildlife populations: past, present, and future. *Journal of the American Statistical Association*, 86: 225–238 (1991)

[10] Pollock, K.H. Capture-recapture models. *Journal of the American Statistical Association*, 95: 293–296 (2000)

[11] Wang, Y. A semiparametric regression model with missing covariates in continuous time capture-recapture studies. *Australian and New Zealand Journal of Statistics*, 47: 287–297 (2005)

[12] Xi, L., Watson, R., Wang, J., Yip, P.S.F. Estimation in capture-recapture models when covariates are subject to measurement errors and missing data. *The Canadian Journal of Statistics*, 37: 645–658 (2009)

[13] Xu, Y., Liu, L., You, N., Pan, H., Yip, P.S.F. Estimating population size for a continuous time frailty model with covariates in a capture-recapture study. *Biometrics*, 63: 917–921 (2007)

[14] Yip, P.S.F., Wang, Y. A unified parametric regression model for recapture studies with random removals in continuous time. *Biometrics*, 58: 192–199 (2002)