A new integrated likelihood for estimating population size in dependent dual-record system

Kiranmoy CHATTERJEE and Diganta MUKHERJEE

1Department of Statistics, Bidhannagar College, Kolkata 700064, India
2Sampling and Official Statistics Unit, Indian Statistical Institute, Kolkata 700108, India

Key words and phrases: Capture–recapture; direction of behavioural dependence; human population; nuisance parameter; time-behavioural response variation model.

MSC 2010: Primary 62F10, 97K80; secondary 62H99, 62P10

Abstract: Efficient estimation of the population size from dependent dual-record system (DRS) remains a statistical challenge in the capture-recapture type experiment. Owing to the non-identifiability of the suitable time-behavioural response variation model (denoted as $M_{tb}$) under DRS, few methods are developed in the Bayesian paradigm based on informative priors. Our contribution in this article is to develop a new integrated likelihood function from model $M_{tb}$ motivated by a novel approach developed by Severini (2007). A suitable weight function on the nuisance parameter is derived with the knowledge of the direction of behavioural dependency. A pseudo-likelihood function is constructed so that the resulting estimator possess some desirable properties including negligible prior (or weight) sensitiveness. Extensive simulations show the superior performance of our proposed method to that of the existing Bayesian methods. Moreover, the proposed estimator is easy to implement from the computational perspective. Applications to two real data sets are presented.

1. INTRODUCTION

The dual-record system (DRS) is a special type of capture–recapture experiment, which is particularly designed for estimating the size of a specified population, say $N$, based on two...
sampling occasions. Application of this capture–recapture technique is common in various disciplines, such as epidemiology, economics, demography, and studies of episodic events. Wolter (1986) sketched different capture–recapture models for estimating $N$ in DRS based on the pioneering work of Otis et al. (1978) who introduced several likelihood models for different situations. Model $M_t$, popular in practice, accounts for time variation effects and assumes causal independence between the sources of information. But this assumption may seriously mislead in most of the situations for the human population, especially when capture probabilities also vary with behavioural responses (Chandrasekar & Deming, 1949). When both the time variation effect and behaviour response effect act together, model $M_{tb}$ is appropriate. Gosky & Ghosh (2011) found that the model $M_{tb}$ was the most robust model in estimating $N$ using simulation studies in the Bayesian paradigm by comparing all the models proposed by Otis et al. (1978).

The underlying behavioural response effect (say, $\phi$) classifies a given population as recapture prone (when $\phi > 1$) or recapture averse (when $\phi < 1$). Usually, demographic studies exhibit recapture prone type dependence. On the contrary, studies on the drug abused population commonly reveal that the underlying list-dependence is negative, that is, the drug abused population is recapture averse. Otis et al. (1978) addressed the estimability problem related to this model. Lloyd’s (1994) martingale approach, Chao, Chu, & Chiu’s (2000) quasi-likelihood approach and Yang & Chao’s (2005) univariate Markovian method successfully solved the problem when the number of capture attempts ($T$) is strictly more than two. Lee, Hwang, & Huang (2003) and Wang, He, & Sun (2015) developed and implemented Bayesian techniques in the context of multiple capture attempts which is seldom used for the human population. Recently, Chatterjee & Mukherjee (2016a) discussed some issues related to the full Bayes method with the non-informative prior specifically for model $M_{tb}$. They also developed some empirical Bayes strategies in DRS considering the present problem in a missing data framework. Generally, in the Bayesian paradigm, a difficulty arises as the resulting estimator for $N$ may be very sensitive to the choice of prior(s). Nour (1982) proposed an estimator in DRS with an assumption equivalent to recapture proneness, but avoided using Bayesian techniques. Nour (1982) and Chatterjee & Mukherjee (2016a) reveal that if directional knowledge on $\phi$ (i.e. recapture proneness or aversion) is available, inference on $N$ could be improved. In this context, an efficient strategy for identification of the direction of lists-dependence in DRS was proposed by Chatterjee & Mukherjee (2016b). Appropriateness and the estimability issues of the present model under DRS motivate us to consider the problem of estimation of $N$. Here, we propose a novel integrated likelihood method as a suitable non-Bayesian strategy, particularly when the underlying population is correctly known as recapture prone or averse.

All the parameters in model $M_{tb}$, except the parameter of interest $N$, are regarded as nuisance parameters (say, $\psi \in \Psi$). Severini (2000) discussed some useful likelihood-based inference procedures through the construction of pseudo-likelihood functions by maximization (profile likelihood) or conditioning (conditional likelihood) or integrating the likelihood function over $\Psi$ with respect to some weight functions (integrated likelihood). Integrated likelihood has an advantage that it always exists unlike other pseudo-likelihood methods. Salasar, Leite, & Louzada (2014) considered an integrated likelihood approach with uniform and Jeffrey’s weight function for elimination of nuisance parameters in model $M_t$. Salasar, Leite, & Louzada (2016) compared four pseudo-likelihood functions (profile, conditional, uniform and Jeffrey’s integrated likelihood) derived from model $M_t$ and discussed procedures for point and interval estimation.

In the integrated likelihood method the main challenge is to choose a suitable weight function. Severini (2007) presented a novel approach for selecting the weight function so that the resulting integrated likelihood is useful for non-Bayesian inference with some nice statistical properties. Recently, Chatterjee & Mukherjee (2016c) developed an integrated likelihood for $M_t$, using the strategy proposed by Severini (2007). This article extends the work of Chatterjee & Mukherjee (2016c) for the present complex $M_{tb}$-DRS model. This article is framed to provide an alternative
to the few existing methods in the literature of traditional homogeneous and behaviourally dependent two-sample capture-recapture data.

In the next section, we discuss data structures for DRS and model $M_{tb}$. In Section 3, we first discuss the integrated likelihood method using the weight function with uniform and Jeffrey’s densities, and then we propose a new integrated likelihood method. In Section 4, evaluation of our proposed method and comparison with few competitors are carried out by an extensive simulation study. We illustrate our method by applying them to two real data sets. Finally in Section 5, we summarize our findings and comment on the usefulness of the proposed integrated likelihood.

2. ANALYSES ON DUAL-RECORD SYSTEM: PRELIMINARIES

2.1. Dual-Record Data Structure

Let us consider a human population $U$ of size $N$ to be estimated. We consider that enlisting all the individuals in $U$ in two attempts will be incomplete. We make two basic assumptions that (i) the population is closed within the time of two capture attempts and (ii) individuals are homogeneous with respect to capture probabilities. Information collected by two sources are matched and the classification is done according to a multinomial fashion (see Table 1, where the notation are introduced); this is known as DRS. Clearly, the number of missed individuals by both systems ($x_{00}$) is unknown, hence making the total population size $N = x_{..}$ unknown. Expected proportions or probabilities associated to the $(j, k)$th cell, denoted as $p_{jk}$, are also given in Table 1, where $j, k \in \{1, 0, \cdot \}$.

Estimates of $N$ can be obtained with different conditions assumed on the individual’s capture probabilities (thus leading to different models). A common practice is to assume independence between two capture attempts. Therefore, the conditional likelihood estimate from model $M_t$ is

$$\hat{N}_{ind} = \left[ x_{11}, x_{10}, x_{01} \right],$$

which is equivalent to the Petersen estimator or the dual system estimator (Wolter, 1986). However, this model is highly criticized due to the unrealistic independence assumption for real life applications. In demographic studies, violation of the independence assumption is commonly observed due to the positive correlation between two capture attempts (Chandrasekar & Deming, 1949; Nour, 1982). Assuming a positive dependency, Nour (1982) proposed an estimator of $N$ as

$$\hat{N}_{Nour} = x_0 + \frac{2x_{11}x_{10}x_{01}}{x_{11}^2 + x_{10}x_{01}},$$

where $x_0 = x_{11} + x_{10} + x_{01}$ is the total number of captured individuals in DRS. On the contrary, negative dependence is observed in epidemiological surveillance of rare or critical diseases.

2.2. Model $M_{tb}$

The independence assumption is often criticized in surveys and censuses of human populations. An individual who is captured in the first attempt may have a different chance to be included in List 2 than an individual who has not been captured in the first attempt. This change is grossly known as the behavioural response variation. When this chance is more, the corresponding individuals are treated as recapture prone; when the chance is less, individuals are treated to be recapture averse. When the behavioural variation effect is combined with the time variation effect, one would get a complex model denoted as $M_{tb}$ (Otis et al., 1978; Chatterjee & Mukherjee, 2016a). Using Table 1, we derive that \(\text{Prob(An individual present in List 2|not in List 1)} = \frac{p_{01}}{1-p_1}\).
Table 1: The 2 × 2 data structure from the dual-record system (DRS) with corresponding cell probabilities displayed in brackets.

| List 1 | List 2 | In | out | Total |
|--------|--------|----|-----|-------|
| In     | x_{11}[p_{11}] | x_{10}[p_{10}] | x_{1}[p_{1}] |
| Out    | x_{01}[p_{01}] | x_{00}[p_{00}] | x_{0}[p_{0}] |
| Total  | x_{1}[p_{1}]     | x_{0}[p_{0}]     | x = N[p_{.}] |

denoted as \( p \), and \( \text{Prob}(\text{An individual present in List 2 in List 1}) = \frac{p_{11}}{p_{1}} \), denoted as \( c \). Therefore, the likelihood function for model \( M_{ib} \) in DRS is

\[
L(N, p_{1}, p, c) \propto \frac{N!}{(N-x_0)!} c^{x_{11}} p_{1}^{x_{11}} p_{0}^{x_{01}} (1-p_{1})^{N-x_{1}} (1-p)^{N-x_{0}} (1-c)^{x_{10}},
\]

where \( N > x_0, 0 < p_{1}, p, c < 1 \), consisting of lesser number of sufficient statistics \( (x_{11}, x_{01}, x_{10}) \) than the parameters \( (N, p_{1}, p, c) \). Note that the parameters of model \( M_{ib} \) are non-identifiable in the sense that for every \( (N, p_{1}, p, c) \), there exist many other parameter choices which give the same likelihood value.

We rewrite \( c \) as \( c = \phi p \) for some \( \phi (>0) \), which is termed as the behavioural response effect characterizing the behavioural dependency of an individual at the time of second capture attempt (Chao, Chu, & Chiu, 2000). Note that the dimension of the parameters cannot be reduced by any reparameterization of likelihood (3), and therefore, the estimability problem persists. However, the reparameterized likelihood in terms of \( \phi \) and \( p \) may be of interest in lieu of likelihood (3) (Chao, Chu, & Chiu, 2000); \( \phi \) and \( p \) are not estimable separately but their product \( c \) is. We will use the \( \phi \) in Section 4.1 to construct various simulated populations as \( \phi \) has a clear interpretation about the direction of behavioural dependence in DRS (i.e. recapture proneness or aversion).

3. INTEGRATED LIKELIHOOD METHOD

For the statistical model with likelihood function \( L(\theta, \psi|\mathbf{x}) \) for the given data \( \mathbf{x} \), where \( \theta (\in \Theta) \) is the parameter of interest and \( \psi (\in \Psi) \) represents the nuisance parameter, our aim is to construct an integrated likelihood so that the nuisance parameter is eliminated through integration. Suppose \( \mathcal{L}^* = \{ L(\theta, \psi|\mathbf{x}) : \psi \in \Psi \} \), the set of likelihoods, is summarized over \( \Psi \) by a weighted average with respect to a chosen function on \( \Psi \), say \( \pi(\psi|\theta) \) defined on \( \Psi \). Hence, the integrated likelihood function with respect to the weight function \( \pi(\psi|\theta) \) is

\[
L^I(\theta) = \int_{\Psi} L(\theta, \psi|\mathbf{x}) \pi(\psi|\theta) d\psi;
\]

see Severini (2000) for a detailed discussion. One advantage of the integrated likelihood over other pseudo-likelihoods (conditional, marginal) is that it is always possible to construct, but this is not true for conditional or marginal likelihood. Moreover, it is not necessary to choose \( \pi(\psi|\theta) \) as a proper density function. One drawback of the integrated likelihood is its subjectiveness due to \( \pi(\psi|\theta) \). The basic aim always remains to choose a suitable \( \pi(\psi|\theta) \) such that \( L^I(\theta) \) could be efficiently useful for non-Bayesian likelihood inference.

In the capture–recapture context, for a fixed \( \theta \), Jeffrey’s and uniform weight functions are the two most popular non-informative weight functions on \( \psi \) for model \( M_t \) (Salasar, Leite, &
Louzada, 2014, 2016). In the following Result 1, we compute the integrated likelihood using (4) with \( \theta = N \) and \( \psi = (p_1, c, p) \).

**Result 1.** Integrated likelihood functions of \( N(> x_0) \), derived from the likelihood of model \( M_{th} \) with uniform and Jeffrey’s weight functions \( \pi(\psi|\theta) \), are

\[
L_U'(N) = \int_L L(N, \psi|x)d\psi \propto \frac{1}{(N + 1)(N - x_1 + 1)}
\]

and \( L_J'(N) = \int_L L(\theta, \psi|x)\pi_J(\psi|\theta)d\psi = \frac{(N - x_1)}{(N + 1)(N - x_0)} \), respectively, where Jeffrey’s weight function, \( \pi_J(\psi|\theta) \), is given in the appendix. Clearly, both \( L_U'(N) \) and \( L_J'(N) \) are strictly decreasing over its domain \( N > x_0 \). Thus, these weight functions cannot add any informative weight on \( \psi \) in the likelihood (3) and hence, they fail to draw inference on \( N \).

### 3.1. Proposed Integrated Likelihood Approach

In a conventional statistical setting, Severini (2007) suggested using

\[
\widetilde{L}'(\theta) = \int L(\theta, \gamma|x)\pi(\gamma)d\gamma,
\]

as an integrated likelihood for \( \theta \), where the parameter \( \gamma \) is chosen so that the maximum likelihood estimate of \( \gamma \), for a given \( \theta \), is approximately constant as a function of \( \theta \). Severini (2007) showed that the solution, \( \gamma \), of

\[
E \{ \ell_\psi (\theta \gamma); \gamma \} \equiv E \{ \ell_\psi (\theta \gamma); \gamma \} |_{\theta_0 = \hat{\theta}, \psi_0 = \gamma} = 0,
\]

is strongly unrelated to the parameter \( \theta \) in the sense that \( \hat{\gamma}_0 = \hat{\gamma} + O(n^{-1/2}) \), for fixed values of \( \theta \) so that \( \theta - \hat{\theta} = O(1) \). The setting here is different in that asymptotics are in \( N \) rather than \( n \), where \( n \) represents sample size in conventional statistical settings (Sen & Sen, 1981). Nevertheless, we will use (6) to determine a choice for the nuisance parameter \( \gamma \), in terms of \( \theta \) and \( \psi \), in such a way so that \( \gamma \) and \( \theta \) become independent under \( \pi(\psi|\theta) \); \( \pi(\psi|\theta) = \pi(h(\gamma, \theta)|\theta) = \pi(\gamma|\theta) = \pi(\gamma) \).

Thereafter, we have to choose a weight function \( \pi(\gamma) \) for \( \gamma \) only, which does not depend on \( \theta \). Hence, we will use (5), with \( \theta = N \), as the integrated likelihood.

The following theorem implements the proposed idea of integrated likelihood (5) for the model \( M_{th} \) in (3). The proof is given in the appendix.

**Theorem 1.** For likelihood (3) with \( \theta = N \), \( \psi = (p_1, c, p) \) and \( \hat{\theta} = \hat{N}_0 \), the parameterization satisfying (6) is \( \gamma = (\gamma_1, \gamma_2, \gamma_3) \), where \( \gamma_1 = (N/\hat{N}_0)p_1, \gamma_2 = c \) and \( \gamma_3 = \frac{\bar{p}(1-p_1)}{(N_0/N) - p_1} \). Here \( \hat{N}_0 \) is any working estimate of \( N \).

Results from Theorem 1 imply that the maximum likelihood estimate of \( \gamma \), for given \( N \), is

\[
\hat{\gamma}_N = (\hat{\gamma}_1,N, \hat{\gamma}_2,N, \hat{\gamma}_3,N),
\]

where

\[
\hat{\gamma}_1,N = (N/\hat{N}_0)\bar{p}_1,N = (N/\hat{N}_0)\frac{x_1}{N} = \frac{x_1}{\hat{N}_0},
\]

\[
\hat{\gamma}_2,N = \hat{c}_N = \hat{c} = \frac{x_{11}}{x_1},
\]

and

DOI: 10.1002/cjs  The Canadian Journal of Statistics / La revue canadienne de statistique
As we find here that \( \hat{\gamma}_N \) does not depend on \( N \), in the non-Bayesian inference about \( N \) based on the integrated likelihood function, the weight function \( \pi(\psi|N) \) reduces to \( \pi(\gamma) \) after replacing \( \psi \) in terms of \( N \) and \( \gamma \) obtained from (6). Now, one can choose any suitable weight function \( \pi(\gamma) \) for \( \gamma \) as the resulting integrated likelihood \( \hat{L}_I(\theta) \) in (5) does not heavily depend on the chosen weight function. Since the current model suffers from the non-identifiability problem, as discussed in Section 2.2, we consider some informative weight function for \( \gamma \) subject to the condition that hyper-parameters satisfy some relations that lead to a suitable integrated likelihood.

**Result 2.** In connection with Theorem 1, suppose the weight function \( \pi(\gamma) \) is of the form \( \pi(\gamma) = \pi(\gamma_1)\pi(\gamma_2|\gamma_3)\pi(\gamma_3|\gamma_4) \), where \( \pi(\gamma_1) = GB1 \left( b_1 = \frac{N}{\hat{N}_0}, r_1, s_1 \right) \), \( \pi(\gamma_2) = \text{Unif}(0, 1) \), and \( \pi(\gamma_3|\gamma_4) = GB1 \left( b_2 = \frac{N/\hat{N}_0 - \gamma_1}{1-\gamma_1}, r_2, s_2 \right) \) and \( GB1(\cdot) \) stands for Generalized beta distribution of first kind. Then for any positive real numbers \( r_1, s_1, r_2, s_2 \) satisfying \( r_2 + s_2 = s_1 \), the integrated likelihood (5) becomes

\[
\hat{L}_I(N) = \frac{\Gamma(N - x_0 + s_2)\Gamma(N + 1)}{\Gamma(N + r_1 + s_1)\Gamma(N - x_0 + 1)}
\]

for \( N \geq x_0 \), where \( \Gamma(a) \) denotes the Gamma function of \( a \), given by \( \int_0^\infty e^{-x}x^{a-1}dx \). Thus, for the given values of \( r_1, s_1 \) and \( r_2, s_2 \), \( \hat{L}_I(N) \) is non-decreasing in \( N \) for

\[
N \leq \frac{x_0(r_1 + s_1 - 1)}{r_1 + r_2} - 1
\]

and \( \hat{L}_I(N) \) converges to 0 as \( N \to \infty \).

The proof is given in the appendix. The rationale behind the assumption \( r_2 + s_2 = s_1 \) in Result 2 is primarily to make the resulting integrated likelihood well-behaved, and secondarily, to reduce the dimension of the hyper-parameters in \( \pi(\gamma) \). By the relationship between \( \gamma \) and \( \psi \), established in Theorem 1, we suggest the hyper-parameters \( r_2 = x_{01}, s_1 = (\hat{N}_0 - x_1) \) and \( r_1 = x_1.s_1/(N - x_1) \), adopting posterior unbiasedness condition for \( \gamma_1 \) and \( \gamma_3 \), where \( \hat{N}_0 \) is a working estimate of \( N \), such as \( \hat{N}_{\text{ind}} \) or \( \hat{N}_{\text{Nour}} \) defined in Section 2.1. See the appendix for detailed computations. As \( r_1 = x_1.s_1/(N - x_1) \) (denoted as \( r_1(N, s_1) \)) depends on \( N \) and \( s_1 \), the right side of (7) also becomes dependent on \( N \). So, \( N \leq \frac{s_0[r_1(N,s_1)+s_1-1]}{[r_2+r_1(N,s_1)]} - 1 \) is equivalent to \( \frac{s_0[r_1(N,s_1)+s_1-1]}{[r_2+r_1(N,s_1)]} - N \geq 1 \) and \( \frac{s_0[r_1(N,s_1)+s_1-1]}{[r_2+r_1(N,s_1)]} - N \) is non-increasing in \( N \) for fixed \( s_1 \) and \( r_2 \).

The result stated in the following theorem discusses the possibility of existence of the corresponding maximum likelihood estimate, whose proof is sketched in the appendix.

**Theorem 2.** Suppose \( r_1 = x_1.s_1/(N - x_1) \). Consider the quadratic equation

\[
N^2r_2 - N(x_1.r_2 + x_{01} .s_1 - r_2 - x_0) + x_1 .(s_1 - x_0 + r_2) = 0,
\]

derived from (7). If (8) has roots \( N^I < N^F \) with \( [N^I] + 1 > x_0 \) and \( L_{10}^I([N^I] + 1) \geq L_{10}^I(x_0) \), the maximizer of \( L_{10}^I(N) \) is \( [N^I] + 1 \). Otherwise it is \( x_0 \).
We explore several properties of the resulting estimator, as pointed out in Severini (2007), through simulation in Section 4.1. In order to implement the above specification on the weight function, when it is known that $\phi > 1$, we suggest $\hat{N}_0 = \hat{N}_{\text{Nour}}$, where $\hat{N}_{\text{Nour}}$ is defined in (2). On the other hand, if $\phi < 1$ then we suggest $\hat{N}_0 = \hat{N}_{\text{ind}}$, where $\hat{N}_{\text{ind}}$ is defined in (1).

Remark 3.1. If we consider $\theta = (N, c, p)$ for likelihood (3), then $\gamma$ can be obtained from the relevant log-likelihood functions as before. Now, for $\psi = p_1$, if the weight function is taken on the associated $\gamma$ as $\pi(\gamma) = GB1(b = N/\hat{N}_{\text{ind}}, r, s)$ for any positive real numbers $r$ and $s$, then the integrated likelihood for $N(\geq x_0)$ would be

$$
\tilde{L}_M^{(1)}(N, c, p) = \frac{\Gamma(N - x_1 - s)\Gamma(N + 1)}{\Gamma(N + r + s)\Gamma(N - x_0 + 1)} e^{x_1 p x_0 (1 - c)}(1 - p)^{N - x_0},
$$

and it will fail to produce the maximum likelihood estimate of $\theta = (N, c, p)$. The failure is perhaps due to insufficient information to draw inference about $N$ and $p$.

4. NUMERICAL ILLUSTRATION

4.1. Simulation Study

We conduct a simulation study to evaluate the performance of our proposed approach and compare it with a few existing competing methods. We simulate hypothetical populations corresponding to six pairs of capture probabilities $(p_1, p_2) = (0.50, 0.65), (0.60, 0.70), (0.80, 0.70), (0.70, 0.55), (0.55, 0.75), (0.70, 0.50)$ for each of the recapture prone (represented by $\phi = 1.25, 1.50$) and recapture averse (represented by $\phi = 0.60, 0.80$) situations. We denote six populations corresponding to the six pairs of $(p_1, p_2)$ as P1 to P6, respectively, representing recapture prone situations and associated results are presented in Tables 2 and 3 with true $N = 200$ and 500, respectively. Results of the other six populations comprising the same six pairs of $(p_1, p_2)$ reflecting recapture averse situations, namely A1–A6, are shown in Tables 4 and 5 for the same two $N$ values respectively. From each of the 12 populations (i.e. P1–P6 and A1–A6), 1,000 data sets on $(x_{11}, x_{10}, x_{01})$ are generated and the proposed integrated likelihood estimate $\hat{N}_\text{lb}$ of $N$ is obtained for each data set. Finally, the average over those 1,000 replicated estimates, $\hat{N}_\text{lb}$, is reported in tables. Next, the root mean square error (RMSE) is computed in which expectation of squared error is empirically obtained over 1,000 replicates. We also compute 95% confidence interval (C.I.) based on the log-transformation method used by Chao (1987) for which the estimate of s.e.($\hat{N}$) is obtained based on parametric bootstrapping with 1,000 bootstrap samples. Finally, the average of those C.I. over 1,000 replications is reported as C.I. and the associated coverage rate (in %) are obtained over 1,000 replications. Note that similar statistics are computed for two existing Bayesian methods, Lee and SEMWiG, proposed respectively by Lee, Hwang, & Huang (2003) and Chatterjee & Mukherjee (2016a). To conduct Bayesian inference on $N$ using Lee and SEMWiG methods, we use the same priors as used in Chatterjee & Mukherjee (2016a) when information is available on the direction of $\phi$. The results are summarized in Tables 2–5.

Tables 2–5 show that the estimates from our proposed method are more efficient than that of by Lee, in terms of accuracy, RMSE and the length of C.I. When $\phi$ is far below 1 (i.e. for $\phi = 0.60$), SEMWiG produces slightly better results for small populations. This discrepancy increases when $N$ is larger. In other situations, the performance of the proposed $\hat{N}_\text{lb}$ tended to give better RMSE than SEMWiG. However, SEMWiG is better than $\hat{N}_\text{lb}$ which is better than Lee in terms of the length of the associated interval estimates of $N$. When $\phi \in [0.80, 1.25]$, our method is better than or comparable with existing competitors. When the value of $\phi$ is far away from 1 (i.e. 0.6 or 1.5 in our illustration), coverage rate by the C.I. associated with the proposed
Table 2: Estimation results for the simulated populations with \( \phi > 1 \) and \( N = 200 \).

| Method | P1         | P2         | P3         | P4         | P5         | P6         |
|--------|------------|------------|------------|------------|------------|------------|
|        | \( \hat{N} \) (RMSE) | \( \hat{N} \) (RMSE) | \( \hat{N} \) (RMSE) | \( \hat{N} \) (RMSE) | \( \hat{N} \) (RMSE) | \( \hat{N} \) (RMSE) |
|        | \( \phi = 1.25 \) | \( \phi = 1.50 \) | \( \phi = 1.25 \) | \( \phi = 1.50 \) | \( \phi = 1.25 \) | \( \phi = 1.50 \) |
| Lee\(^a\) | 193 (15.38) | 173 (28.33) | 191 (12.33) | 178 (23.11) | 195 (6.90) | 179 (11.48) |
| C.I. | (172, 229) | (158, 195) | (167, 193) | (167, 194) | (183, 200) | (183, 201) |
| Coverage | 86 | 29 | 19 | 19 | 41 | 41 |
| SEMWiG\(^b\) | 197 (12.79) | 176 (25.80) | 194 (10.59) | 179 (21.94) | 198 (5.89) | 191 (10.53) |
| C.I. | (194, 201) | (174, 178) | (192, 197) | (178, 181) | (197, 200) | (190, 193) |
| Coverage | 21 | 17 | 9 | 19 | 18 | 18 |
| \( \hat{N}^I_{tb} \) | 191 (13.16) | 175 (26.99) | 195 (9.22) | 180 (21.66) | 199 (5.07) | 191 (10.56) |
| C.I. | (177, 211) | (164, 194) | (183, 213) | (171, 197) | (191, 213) | (185, 213) |
| Coverage | 87 | 25 | 92 | 34 | 94 | 73 |

\(^a\)Prior on \( \phi \) is chosen as \( U(1, 2) \) since \( \phi > 1 \) is known (see Chatterjee & Mukherjee, 2016a).

\(^b\)Prior on \( \phi \) is chosen as \( U(1, p^{-1}) \) since \( \phi > 1 \) is known (see Chatterjee & Mukherjee, 2016a).

Sensitivity analysis shows that the proposed estimate \( \hat{N}^I_{tb} \) is lower in some cases than the existing methods but the C.I. is tighter also. Further, the invariance property of the estimates obtained through our proposed integrated likelihood is verified as the estimates for true \( N = 500 \) are 2.5 times higher than that for true \( N = 200 \). Lastly, another important feature of \( \hat{N}^I_{tb} \) is that it does not incur serious computational effort as for existing Bayesian strategies: Lee and SEMWiG.

We also conducted a sensitivity analysis of the proposed estimate \( \hat{N}^I_{tb} \) with respect to the nuisance parameters \( \phi \) and \( p \). We consider true \( N = 500 \) and the values of \( p_1 \), set as 0.50, 0.60, 0.70, 0.80. Then, for each population with each of the four values of \( p_1 \), we plotted RMSE of \( \hat{N}^I_{tb} \) over \( \phi \in (1, 1.5) \) and \( p \in (0.3, 0.7) \) in the left panel of Figure 1. A similar graph is drawn over \( \phi \in (0.5, 1) \) and \( p \in (0.3, 0.7) \) in the right panel of Figure 1. It shows that the changes in RMSE are less with respect to \( \phi \) compared to \( p \). Also, RMSE values are more stable in the recapture prone cases than the recapture averse cases.

4.2. Real Data Example I: Children Injury Data

We consider a work by Jarvis et al. (2000), in which authors illustrate the serious drawbacks in \( \hat{N}_{ind} \) specifically for injury related data. The problem was to get the count of children under...
Table 3: Estimation results for the simulated populations with \( \phi > 1 \) and \( N = 500 \).

| Method | P1 (RMSE) | P2 (RMSE) | P3 (RMSE) | P4 (RMSE) | P5 (RMSE) | P6 (RMSE) |
|--------|-----------|-----------|-----------|-----------|-----------|-----------|
| \( \phi = 1.25 \) | \( \hat{N} \) | \( \hat{N} \) | \( \hat{N} \) | \( \hat{N} \) | \( \hat{N} \) | \( \hat{N} \) |
| Lee\(^a\) | 472 (34.11) | 478 (24.33) | 490 (12.02) | 488 (17.03) | 504 (13.71) | 527 (32.80) |
| C.I. | 438, 518 | 451, 519 | 473, 515 | 457, 531 | 445, 576 | 471, 599 |
| Coverage | 72 | 88 | 90 | 94 | 100 | 98 |
| SEMWiG\(^b\) | 484 (23.92) | 485 (21.25) | 495 (9.87) | 502 (16.32) | 480 (24.64) | 514 (22.89) |
| C.I. | 451, 522 | 456, 511 | 477, 512 | 474, 532 | 473, 493 | 505, 521 |
| Coverage | 93 | 80 | 82 | 96 | 28 | 28 |
| \( \hat{N}_t \) \(^b\) | 478 (26.75) | 486 (19.60) | 498 (9.06) | 497 (13.85) | 476 (27.19) | 496 (13.35) |
| C.I. | 455, 509 | 466, 514 | 484, 519 | 475, 526 | 458, 502 | 474, 526 |
| Coverage | 73 | 83 | 96 | 95 | 51 | 97 |
| \( \phi = 1.50 \) | \( \hat{N} \) | \( \hat{N} \) | \( \hat{N} \) | \( \hat{N} \) | \( \hat{N} \) | \( \hat{N} \) |
| Lee\(^a\) | 437 (64.17) | 455 (47.37) | 481 (21.58) | 483 (28.67) | 445 (56.40) | 492 (20.01) |
| C.I. | 405, 487 | 424, 500 | 452, 522 | 437, 538 | 419, 488 | 449, 550 |
| Coverage | 22 | 46 | 93 | 95 | 15 | 96 |
| SEMWiG\(^b\) | 442 (59.56) | 450 (51.53) | 477 (22.66) | 476 (27.53) | 436 (64.69) | 480 (25.67) |
| C.I. | 436, 452 | 447, 456 | 475, 582 | 470, 482 | 433, 439 | 475, 487 |
| Coverage | 0 | 0 | 3 | 10 | 0 | 19 |
| \( \hat{N}_t \) \(^b\) | 442 (60.52) | 455 (47.24) | 480 (21.69) | 479 (23.62) | 439 (62.40) | 479 (24.25) |
| C.I. | 421, 467 | 437, 476 | 468, 500 | 455, 503 | 426, 460 | 454, 505 |
| Coverage | 2 | 5 | 51 | 58 | 1 | 62 |

\(^a\)Prior on \( \phi \) is chosen as \( U(1, 2) \) since \( \phi > 1 \) is known (see Chatterjee & Mukherjee 2016a).

\(^b\)Prior on \( \phi \) is chosen as \( U(1, p^{-1}) \) since \( \phi > 1 \) is known (see Chatterjee & Mukherjee 2016a).

15 years of age from addresses in Northumbria who were seriously injured in local motor vehicles accidents (MVA) between April 1, 1990 and March 31, 1995. One source was Stats19 data covering all incidents reported to the police and another was the Hospital Episode data (HES) covering admissions of children. The associated DRS data are presented in Jarvis et al. (2000), Table 4, p. 48) for three different strata—cyclists, passengers, and pedestrians. Jarvis et al. argued that children injured in MVA as pedestrians or cyclists rarely enter insurance claims for which they have to inform police for case diary. It is also noted that estimates \( \hat{c} \) for these three strata are 0.25, 0.40, and 0.59, respectively, which indicate recapture aversion. The estimates under independence (\( \hat{N}_{ind} \)) are also presented along with the three competitors analyzed in Section 4.1 including our proposed \( \hat{N}_{tb} \). Standard errors associated to the estimates and C.I. of the true population size are obtained using 1,000 bootstrap samples. We summarize all the results in the top panel of Table 6.

It is observed that \( \hat{N}_{tb} \), Lee and SEMWiG agree with the negative departure from independence, that is recapture aversion and the resulting dual system estimates \( \hat{N}_{ind} \) produce large estimates which seriously overestimates \( N \). Estimates from Lee’s method possess larger variation than all other estimates and hence its credible intervals are too wide. Estimate \( \hat{N}_{tb} \) has better efficiency.
TABLE 4: Estimation results for the simulated populations with $\phi < 1$ and $N = 200$.

| Method          | A1     | A2     | A3     | A4     | A5     | A6     |
|-----------------|--------|--------|--------|--------|--------|--------|
| $\phi = 0.60$   |        |        |        |        |        |        |
| $\hat{N}$ (RMSE) | 236 (43.96) | 222 (28.77) | 214 (15.51) | 222 (26.56) | 234 (38.12) | 224 (29.91) |
| C.I.            | (186, 316) | (195, 271) | (200, 244) | (188, 291) | (198, 295) | (186, 297) |
| Coverage        | 92     | 88     | 17     | 96     | 79     | 90     |
| $\hat{N}$ (RMSE) | 213 (13.44) | 211 (11.28) | 208 (7.90) | 200 (0.76) | 216 (16.71) | 199 (2.05) |
| C.I.            | (200, 230) | (203, 219) | (203, 211) | (193, 209) | (210, 222) | (193, 204) |
| Coverage        | 2      | 0      | 0      | 98     | 0      | 99     |
| $\hat{N}$ (RMSE) | 214 (15.72) | 215 (16.23) | 210 (11.26) | 203 (5.42) | 220 (21.52) | 200 (4.35) |
| C.I.            | (199, 237) | (204, 233) | (206, 218) | (196, 219) | (209, 239) | (191, 216) |
| Coverage        | 34     | 2      | 9      | 71     | 0      | 91     |
| $\phi = 0.80$   |        |        |        |        |        |        |
| $\hat{N}$ (RMSE) | 196 (7.78) | 194 (8.64) | 199 (3.89) | 191 (10.92) | 199 (9.26) | 203 (11.26) |
| C.I.            | (169, 242) | (184, 220) | (193, 213) | (178, 247) | (186, 229) | (175, 258) |
| Coverage        | 94     | 97     | 95     | 99     | 99     | 98     |
| $\hat{N}$ (RMSE) | 192 (8.43) | 195 (4.70) | 195 (4.48) | 192 (9.34) | 199 (3.45) | 185 (14.62) |
| C.I.            | (188, 198) | (189, 200) | (193, 197) | (186, 196) | (194, 204) | (182, 190) |
| Coverage        | 0      | 0      | 0      | 50     | 100    | 0      |
| $\hat{N}$ (RMSE) | 196 (8.25) | 201 (5.32) | 201 (3.25) | 193 (9.32) | 201 (4.39) | 190 (11.80) |
| C.I.            | (185, 214) | (193, 215) | (197, 210) | (186, 206) | (195, 219) | (183, 203) |
| Coverage        | 95     | 94     | 86     | 80     | 84     | 68     |

$^a$Prior on $\phi$ is chosen as $U(0.2, 1.4)$ since $\phi < 1$ is known (see Chatterjee & Mukherjee 2016a).

$^b$Prior on $\phi$ is chosen as $U(\hat{\phi}, 1)$ since $\phi < 1$ is known (see Chatterjee & Mukherjee 2016a).

than all other estimates. The proposed $\hat{N}_{ib}$ with the suggested directional knowledge (i.e. $\phi < 1$) produces an estimate between those of $Lee$ and $SEMWiG$; in most cases, it has a smaller variance and tighter confidence bounds, as expected.

4.3. Real Data Example II: Handloom Data

We consider another dataset from a survey aimed to estimate the undercount in the census of handloom workers residing at Gangarampur in South Dinajpur district of West Bengal in India. The survey was post enumeration type and conducted in November 2013, 3 months after the handloom census operation (Mukherjee, 2014). The classification strategy proposed by Chatterjee & Mukherjee (2016b) infers that ward 2 and ward 16 have the nature of recapture proneness and aversion. For details on the data sets, we refer to Chatterjee & Mukherjee (2016b).

Being quite assured about the homogeneity within wards from the experts of Textile Directorate, we apply the model $M_{tb}$ to these data and compute the estimates using our proposed integrated likelihood method and other rival methods. Similar statistics to those in Section 4.2 are computed and the results are presented in the bottom panel of Table 6.
TABLE 5: Estimation results for the simulated populations with $\phi < 1$ and $N = 500$.

| Method | A1   | A2   | A3   | A4   | A5   | A6   |
|--------|------|------|------|------|------|------|
| $\phi = 0.60$ |      |      |      |      |      |      |
| Lee$^a$ | $\hat{N}$ (RMSE) | 625 (136.80) | 608 (115.19) | 553 (56.30) | 530 (34.31) | 619 (124.43) | 529 (40.26) |
| C.I.     | (485, 827) | (505, 744) | (511, 619) | (472, 671) | (510, 775) | (464, 669) |
| Coverage | 80 | 49 | 3 | 99 | 24 | 97 |
| SEMWiG$^b$ | $\hat{N}$ (RMSE) | 521 (25.48) | 517 (18.23) | 517 (17.16) | 517 (19.69) | 527 (27.73) | 513 (15.87) |
| C.I.     | (490, 549) | (505, 529) | (514, 521) | (499, 539) | (512, 542) | (491, 538) |
| Coverage | 88 | 18 | 0 | 50 | 1 | 86 |
| $\hat{\eta}_{tb}$ | $\hat{N}$ (RMSE) | 541 (43.00) | 534 (35.24) | 528 (28.34) | 514 (16.47) | 548 (49.80) | 505 (10.61) |
| C.I.     | (516, 577) | (517, 560) | (519, 540) | (499, 535) | (530, 577) | (489, 527) |
| Coverage | 7 | 0 | 2 | 51 | 0 | 91 |
| $\phi = 0.80$ |      |      |      |      |      |      |
| Lee$^a$ | $\hat{N}$ (RMSE) | 472 (35.42) | 490 (13.65) | 509 (10.34) | 481 (20.08) | 531 (39.79) | 529 (38.53) |
| C.I.     | (431, 565) | (460, 560) | (485, 545) | (448, 545) | (475, 633) | (459, 608) |
| Coverage | 99 | 99 | 100 | 100 | 95 | 98 |
| SEMWiG$^b$ | $\hat{N}$ (RMSE) | 478 (25.27) | 492 (11.74) | 498 (5.28) | 484 (19.34) | 490 (12.15) | 480 (22.96) |
| C.I.     | (455, 507) | (477, 510) | (489, 508) | (465, 506) | (485, 496) | (462, 493) |
| Coverage | 60 | 96 | 75 | 71 | 26 | 34 |
| $\hat{\eta}_{tb}$ | $\hat{N}$ (RMSE) | 487 (17.81) | 499 (8.30) | 502 (5.29) | 484 (17.55) | 506 (10.19) | 488 (13.58) |
| C.I.     | (469, 513) | (486, 518) | (494, 515) | (473, 501) | (493, 527) | (476, 505) |
| Coverage | 79 | 96 | 88 | 51 | 86 | 68 |

$^a$Prior on $\phi$ is chosen as $U(0.2, 1.4)$ since $\phi < 1$ is known (see Chatterjee & Mukherjee, 2016a).

$^b$Prior on $\phi$ is chosen as $U(\hat{\phi}, 1)$ since $\phi < 1$ is known (see Chatterjee & Mukherjee, 2016a).

$\hat{\eta}_{tb}$ estimates 164 handloom workers under the consideration of recapture proneness of ward 2. The efficiency of our proposed estimator is better than Lee but little worse than SEMWiG. However, estimated C.I. based on SEMWiG is very much tighter. For the other sampled ward 16, results from our method assuming recapture aversion is very close to the results found from SEMWiG method.

5. DISCUSSION

The independence assumption between the lists does not hold in many instances. Various data on epidemiology and undercount in demography indicate that model $M_{tb}$ is most relevant. Here, we consider integrated likelihood method as a non-Bayesian strategy which has potential to overcome the non-identifiability in $M_{tb}$-DRS model. We have shown that general integrated likelihoods using common non-informative weight functions fail to produce a reasonable estimate of $N$. Our method produces efficient estimates with several properties including invariance, less weight function sensitiveness, and so on. To the best of our knowledge, this article presents the first efficient non-Bayesian strategy for the complex $M_{tb}$-DRS model. Our simulation study...
Figure 1: Sensitivity analysis of the performance of proposed $\hat{N}_t^p$ in terms of RMSE over $\phi$ and $p$; (a) to (d) in both the left (recapture prone) and right (recapture averse) panels have $p_1 = 0.50, 0.60, 0.70, \text{ and } 0.80$ respectively.

supports the fact that this newly developed integrated likelihood method is either more efficient in some regular situations than the existing methods or it is comparable to them in other situations. Our method also incurs much lower computational burden than other existing methods for estimating the true population size.
Proof of Result 1. If uniform prior \( \pi_U(\psi|N) \propto 1 \) on \( \psi = (p_1, p, c) \) is used, the resulting integrated likelihood of \( N \) using (4) based on the model \( M_{lb} \) in (3) becomes

\[
L_U^l(N) = \frac{N!}{(N-x_0)!} \int_{p_1} \int_{p} \int_{c} c^{x_1} p_1^{x_1} p_1^{x_0} (1-p_1)^{N-x_1} (1-p)^{N-x_0} (1-c)^{x_0} dc dp dp_1.
\]

Similarly, we obtain the integrated likelihood of \( N \) using Jeffrey’s weight function for \( \psi \), that is,

\[
\pi_J(\psi|N) \propto \sqrt{L_N(\psi)} = \left[ \det \left( \frac{N}{p_1(1-p_1)} \right) \right]^{1/2} = (c(1-c)p(1-p))^{-1},
\]

where \( I_N(\psi) \) is 3 \times 3 Fisher’s information matrix, as

\[
L_J^l(N) = \frac{N!}{(N-x_0)!} B(x_{11}, x_{10}) B(x_{01}, N-x_0) B(x_1. + 1, N-x_1. + 1)
\]

\[
\propto \frac{(N-x_0)}{(N+1)(N-x_0)}. \tag{5}
\]

Note that here \( B(a, b) \) denotes the ‘beta function’ which equals \( \frac{(a-1)! (b-1)!}{(a+b-1)!} \).

Proof of Theorem 1. The partial derivatives of the log-likelihood of model \( M_{lb} \) in (3) are

\[
\ell_{p_1}(N, p_1, c, p) = (x_{11}/p_1) - (N-x_1)/1-p_1, \quad \ell_c(N, \psi_1, \psi_2) = x_{11}/c - x_{10}/(1-c), \quad \ell_p(N, p_1, c, p) = (x_{01}/p) - (N-x_0)/(1-p).
\]

Now, following (6), if we take expectations on the partial
derivatives of the log-likelihood function over the distribution fixing \((\theta, \psi)\) at \((\theta_0, \psi_0)\), we have

\[
E(\ell_{p_1}(N_{p_1}p_1)) : N_{p_1}p_{1:0}, c_0, p_0) = \frac{N_0p_{1:0}}{p_1} - \frac{N - N_0p_{1:0}}{1 - p_1}, \tag{A.1}
\]

\[
E(\ell_{c}(N_{p_1}p_1)) : N_{p_1}p_{1:0}, c_0, p_0) = \frac{N_0p_{1:0}}{c} - \frac{N - N_0(p_{1:0} - p_{1:0})}{1 - c}, \tag{A.2}
\]

\[
E(\ell_{p}(N_{p_1}p_1)) : N_{p_1}p_{1:0}, c_0, p_0) = \frac{N_0p_{0:1}}{p} - \frac{N - N_0(p_{1:0} + p_{0:1})}{1 - p}. \tag{A.3}
\]

Now, we consider a system of equations by equating the expressions (A.1)–(A.3) to 0. By solving the system of equations evaluated at \(N_0 = \hat{N}_0, \psi_0 = \gamma\), we obtain the following new set of nuisance parameters \(\gamma = (\gamma_1, \gamma_2, \gamma_3)\) with \(\gamma_1 = (N/\hat{N}_0)p_1, \gamma_2 = c, \gamma_3 = \frac{p(p_1 - p_1)}{(N_0/N) - p_1}. \)

**Proof of Result 2.** Nuisance parameters \(p_1, c, p\) of the model \(M_{tb}\) are expressed in terms of \(\gamma\), stated in Theorem 1, as

\[
p_1 = (\hat{N}_0/N)\gamma_1, c = \gamma_2, p = \frac{\gamma_3(1 - \gamma_1)}{(N/\hat{N}_0) - \gamma_1}. \tag{A.4}
\]

Now, by replacing \(\psi = (p_1, c, p)\) using the relations mentioned in (A.4), we rewrite the likelihood (3) as

\[
L_{tb}(N, \gamma) \propto \frac{N!(1 - \gamma_2)^{x_{01}}(1 - \gamma_1)^{x_{11}}}{(N - x_0)!N^{x_1}} \left[ \frac{\gamma_3(1 - \gamma_1)}{(N/\hat{N}_0) - \gamma_1} \right]^{x_{01}} \left(1 - \frac{\hat{N}_0}{N}\gamma_1\right)^{N - x_1} \times \left[1 - \frac{\gamma_3(1 - \gamma_1)}{(N/\hat{N}_0) - \gamma_1}\right]^{N - x_0}, \tag{A.5}
\]

provided \(\gamma_1 < (N/\hat{N}_0)\). Now, we consider weight functions for \(\gamma_1\) and \(\gamma_3\), respectively, as

\[
\pi(\gamma_1) \equiv GB\left( b_1 = \frac{N}{\hat{N}_0}, r_1, s_1 \right) \propto \gamma_1^{r_1 - 1} \left(1 - \frac{\hat{N}_0}{N}\gamma_1\right)^{s_1 - 1},
\]

\[
\pi(\gamma_3) \equiv GB\left( b_2 = \frac{(N/\hat{N}_0) - \gamma_1}{1 - \gamma_1}, r_2, s_2 \right) \propto \gamma_3^{r_2 - 1} \left[1 - \frac{(1 - \gamma_1)}{(N/\hat{N}_0) - \gamma_3}\right]^{s_2 - 1},
\]

which are conjugate with \(L_{tb}(N, \gamma)\) in (A.5), for positive real numbers \(r_1, s_1, r_2, s_2\). We consider non-informative weight function on \(\gamma_2\) as \(\pi(\gamma_2) \equiv \text{Unif}(0, 1)\). So, \(\pi(\gamma)\) is of the form \(\pi(\gamma) = \pi(\gamma_1)\pi(\gamma_2)\pi(\gamma_3|\gamma_1)\). Integration of the form \(\int_0^b y^{q-1}(1 - y/b)^{q-1}dy\) equals \(b^qB(p, q)\). Hence, assuming \(r_2 + s_2 = s_1\), we perform an integration of the likelihood in (A.5) over \(\gamma_3, \gamma_2, \gamma_1\) and obtain

\[
\bar{L}_{tb}(N) \propto \frac{N!(N - x_0 + s_2 - 1)!(N - x_1 + s_1 - 1)!}{(N - x_0)!(N - x_1 + r_2 + s_2 - 1)!(N + r_1 + s_1 - 1)!}.
\]
Hence, we have the $\tilde{L}_I(N)$ stated in Result 2. Further note that $\tilde{L}_I(N + 1)/\tilde{L}_I(N)$ is proportional to $(N - x_0 + s_2)(N + 1)/(N + r_1 + s_1)(N - x_0 + 1)$. Therefore, since $r_2 + s_2 = s_1$, $\tilde{L}_I(N + 1)/\tilde{L}_I(N) \geq 1$ is equivalent to $(s_2 + s_2 - N - x_0) \geq (r_1 + s_1)N + (r_1 + s_1)(1 - x_0)$, which is same as $N \leq x_0(r_1 + s_1 - 1)/(r_1 + s_1 - s_2) - 1$.

**Proof of Theorem 2.** The inequality in (7) can be simplified to the quadratic equation in (A.7). If there are no roots, or if $N \leq x_0$ or $N \geq x_0$ and the maximizer is at $x_0$. If there are roots and $N > x_0$, the function is increasing only when $\max(x_0, N \gamma) \leq N \leq N$, so that the maximum is at either $x_0$ or $[N_1] + 1$. In all cases the maximizer is either $x_0$ or $[N_1] + 1$.

**Acknowledgements**

The authors gratefully acknowledge the valuable inputs of the Editor, the Associate Editor and the anonymous referees on previous versions of this article. This work was partially supported by the
research fellowship award (No. DS/JSTK-CC-0020F dated 11 July, 2013) given to the first author by the Indian Statistical Institute, India.

BIBLIOGRAPHY

Chandrasekar, C. & Deming, W. E. (1949). On a method of estimating birth and death rates and the extent of registration. *Journal of American Statistical Association*, 44, 101–115.
Chao, A. (1987). Estimating the population size for capture-recapture data with unequal catchability. *Biometrics*, 43, 783–791.
Chao, A., Chu, W., & Chiu, H. H. (2000). Capture–recapture when time and behavioral response affect capture probabilities. *Biometrics*, 56, 427–433.
Chatterjee, K. & Mukherjee, D. (2016a). On the estimation of homogeneous population size from a complex dual-record system. *Journal of Statistical Computation and Simulation*, 86, 3562–3581.
Chatterjee, K. & Mukherjee, D. (2016b). On the identification of the nature of behavioural dependence in two-sample capture-recapture study. Technical Report. https://sites.google.com/site/kiramjoychatterjee/.
Chatterjee, K. & Mukherjee, D. (2016c). An improved integrated likelihood population size estimation in dual-record system. *Statistics and Probability Letters*, 110, 146–154.
Gosky, R. & Ghosh, S. K. (2011). A comparative study of Bayes estimators of closed population size from capture–recapture data. *Journal of Statistical Theory and Practice*, 5, 241–260.
Jarvis, S. N., Lowe, P. J., Avery, A., Levene, S., & Cormack, R. M. (2000). Children are not goldfish-mark/recapture techniques and their application to injury data. *Injury Prevention*, 6, 46–50.
Lee, S. M., Hwang, W. H., & Huang, L. H. (2003). Bayes estimation of population size from capture–recapture models with time variation and behavior response. *Statistica Sinica*, 13, 477–494.
Mukherjee, D. (2014). Survey of looms and work-sheds in comprehensive handloom development programme in Dakshin Dinajpur district, West Bengal, India. A project report, Sampling and Official Statistics Unit, Indian Statistical Institute, Commissioned by: Directorate of Textiles, Government of West Bengal, India.
Nour, E. S. (1982). On the estimation of the total number of vital events with data from dual-record collection systems. *Journal of the Royal Statistical Society Series A*, 145, 106–116.
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, 1–135.
Salasar, L. E. B., Leite, J. G., & Louzada, F. (2014). On the integrated maximum likelihood estimators for a closed population capture–recapture model with unequal capture probabilities. *Statistics*, 49, 1204–1220.
Salasar, L. E. B., Leite, J. G., & Louzada, F. (2016). Likelihood-based inference for population size in a capture–recapture experiment with varying probabilities from occasion to occasion. *Brazilian Journal of Probability and Statistics*, 30, 47–69.
Sen, A. R. & Sen, P. K. (1981). Schnabel type estimators for closed populations with multiple markings. *Sankhyā Series B*, 43, 68–80.
Severini, T. A. (2000). *Likelihood Methods in Statistics*. Oxford University Press, New York.
Severini, T. A. (2007). Integrated likelihood functions for non-Bayesian inference. *Biometrika*, 94, 529–542.
Wang, X., He, C. Z., & Sun, D. (2015). Bayesian estimation of population size via capture-recapture model with time variation and behavioral response. *Journal of Ecology*, 5, 1–13.
Wolter, K. M. (1986). Some coverage error models for census data. *Journal of American Statistical Association*, 81, 338–346.
Yang, H. C. & Chao, A. (2005). Modeling animals’ behavioral response by Markov chain models for capture–recapture experiments. *Biometrics*, 61, 1010–1017.

Received 27 June 2017
Accepted 24 June 2018