Estimating the Size of Homogeneous Population via Bayesian Method under a complex Dual-record System

Kiranmoy Chatterjee
Diganta Mukherjee

Abstract

In Dual-record system, the popular Chandrasekar-Deming (1949, JASA, 44, 101-115) model incorporates only the time variation effect on capture probabilities. Some population may have behavioral change after being captured first time. In this paper we focus on the Dual-record system model with both the time as well as behavioral response variation. Two approaches are proposed from which Bayes estimates can be obtained using very simple Gibbs sampling operation from explicit conditional posterior densities. We explore the features of our two proposed methods and their usages depending on the availability or non-availability of information on the nature of behavioral response effect. Extensive simulation studies are carried out to evaluate their performances and compare with available classical approach. Finally, the model and the methods are applied on a real dataset from a Population Change Survey.

Key words: Behavioral response; Dual-record system; Generalised beta type-1 distribution; Gibbs sampling; Human population, Time variation.

1 Introduction

The problem of human population size estimation is a very important statistical concern which includes a vast area of application in the fields of epidemiology, demography and Government statistics. There are several situations where the actual size of a population or a population with certain characteristics or any vital event occurring in a specified area and time span is a matter of prime interest. For that they rely on census operation. But when population is very large and/or hardly countable accurately then census operation also fails to extract the true figure. As a remedy, one or more independent information is collected on that population near to the census operation time and estimate the population size, \(N\), by matching the available lists of information. In the context of human population, this kind of system for merging information from different lists is generally known as Multiple-record system. Indeed, this system is very much more popular in biological or epidemiological studies with the name Capture-recapture system (see Seber (1982[22], 1986[23], 1992[24]), Otis et al. (1978)[20] and Chao et al. (1996[4], 2001[6])). In dual-record (or, shortly, dual) system estimation (DSE), two samples and in triple system estimation (TSE), three samples of information are used to estimate the \(N\). Undercount estimation in census, extent of registration for vital events, estimation of the size of drug abused population, etc. are the primal potential application of this DSE model. In connection with undercount rate estimation in US

\*Bayesian and Interdisciplinary Research Unit (BIRU), Indian Statistical Institute, 203, B. T. Road, Kolkata-700108, India; E-mail: kiranmoy07@gmail.com
\†Sampling and Official Statistics Unit (SOSU), Indian Statistical Institute, 203, B. T. Road, Kolkata-700108, India; E-mail: dgantam@hotmail.com
Census, different models based on Dual-record system approach have been reviewed by Wolter (1986, [27]). The most simple and relevant model is $M_t$, which accounts for only time(t) variation effect, was first invoked by Chandrasekar and Deming (1949, [3]) on human population. Later it also became well-known as C-D estimator in population studies (El-Khorazaty (1975 [13]) and Marks et al. (1974 [18])). There are various frequentist and likelihood approaches for all the basic DSE models in the literature of animal population size estimation (see, Bishop et al.(1975, [1]), Huggins(1989, [12])) and epidemiological study from capture-recapture experiment (see, Hook (1995, [11]) and Chao et al. (2001, [6])). Bayesian approach was pioneered by Robert (1967, [21]), Castledine (1981, [2]) and Smith (1988 [25], 1991 [26]), primarily for $M_t$. George and Robert (1992, [9]) first gave an extensive account on the population size estimation through hierarchical Bayesian analysis via Gibbs sampling on model $M_t$.

But this common DSE model doesn’t work well because the underlying assumption of independence between the capture probabilities in different samples is violated in most of the situations. The capture probabilities may become correlated either due to list dependence or heterogeneity among individuals or both the factors. However, in this article, we always talk about a population which is homogeneous quite well; so list dependence remains the only causal factor for the non-appropriateness of the model $M_t$. List dependence occurs when capture probability for every individual at the time of second survey depends whether he/she is captured in previous time. This change is meant by the behavioral change. A model $M_b$ describe only this change nicely by a parameter $\phi \in \mathbb{R}^+$, called Behavioral Response Effect. When both the time(t) variation effect and behavior response(b) effect acts together on the probability of capture in second list, then we will have a more complicated model $M_{tb}$. This model has a strong relevance for a group of homogeneous individuals when the sample lists are thought to be dependent. In demographic application, this association is usually positive ([3], Greenfield (1975 [10])). But for a population with sensitive characteristics, such as drug users, population with a disease like Common Congenital Anomaly, this association might become negative.

Otis et al. (1978, [20]) addressed the identifiability problem related to the model $M_b$ when number of samples is only two. Hence, as per our knowledge, only one estimator is available in the dual system for a model equivalent to $M_b$. However, this approach proposed by Nour (1982, [19]) for recapture prone (i.e. $\phi > 1$) population is not derived through the likelihood method. Hint to find the estimator for other type of population is also given in their article. Chao et al.(2000, [5]) derived quasi-likelihood estimators following Lloyd’s (1994, [17]) assumption and compared this with the conditional and unconditional MLEs for number of samples greater than two. In Bayesian paradigm, Lee and Chen (1998, [15]) applied the Gibbs sampling idea to the model $M_b$ and $M_{tb}$ but they did not use recapture data and estimates were unstable and prior sensitive. Lee et al. (2003) considered an $M_b$ model setup with an informative uniform prior on behavioral effect $\phi$ and came up with a Bayesian solution which works well for number of samples ($t$) greater than two. However, the conditional densities for capture probabilities and behavior response effect cannot be identified separately according to their model. So, they used adaptive rejection sampling to generate from approximate densities. Their estimator performs well when a reasonably good uniform prior is chosen for $\phi$. Then the appropriate prior leads to greater precision than MLE for large amount of recapture information. Hence their approach is subjective specially when no specific information on $\phi$ can be available from experience as well as when number of recaptured individuals is not large enough.
Estimation of \( N \), from the model \( M_{tb} \), is the main interest of this paper. A Bayesian solution has been given here which has certain advantages over Lee et al. (2003) when only two independent samples are available. We have reparameterized the model \( M_{tb} \) for our convenience and suggest two different priors on \( \phi \). One is informative uniform prior as we need to specify its range using the available directional knowledge on \( \phi \), other one is a conjugate beta prior with data based values for the hyperparameters. Prior specification preserves some characteristics of the system so that nice solutions could be obtained simply, even when no information on \( \phi \) is available. If there is no behavioral response effect, our model automatically reduces to the model \( M_t \) by George and Robert (1992, [9]). The goal of the article is also to investigate the potential of the Bayesian methodology for human population size estimation as an alternative or supplement to the few existing approaches in the Dual-record system framework when behavioral effect plays a key role along with time variation.

In the next section, first we present the data structure in Dual-record system following the most common model \( M_t \) and another basic model \( M_b \) with behavioral response. Estimation of \( N \) from likelihood methods are discussed briefly for the two basic models. Then the combined model \( M_{tb} \), our main interest, is discussed. In section 3, we propose a very simply understandable Bayesian approach with the informative constant prior associated to the model \( M_{tb} \) and discuss its advantages about computations. Extensive numerical illustrations of this approach are presented over several hypothetical populations to evaluate its performance. Understanding a need to develop another approach when directional knowledge is not available, we formulated a second simple Bayesian analysis in section 4 with the hyperparameters chosen empirically from data using a nice inter-relationship among them. A comparative simulation study of the Bayes estimates under this framework is given for different loss functions over the same hypothetical populations. We also apply our models on a real data on death counts from a Population Change Survey in section 5 and compare it with the estimator proposed by Nour ([19]). Finally in section 6, we summarize our findings and provide the best possible estimation rule depending upon either availability or unavailability of the directional knowledge on the behavioral change of the population.

2 Dual-record System: Preliminaries

The idea of dual collection is equivalent to the very popular capture-recapture sampling in wildlife management to estimate \( N \). Laplace (1786, [14]) was the first man who used this capture-recapture approach to estimate the population size of France (Seber 1982, p. 104) from vital events births, marriages and deaths. Consider a given human population \( U \) and \( N \) be the size of the population \( U \). When any census or complete enumeration is done and it is believed that the census failed to capture all events, then it is considered as a sample (though a very large one, generally, for human population). To know the extent of coverage or equivalently the true \( N \), one needs two or more such samples covering that population. Then combining all the sources of information the estimate of \( N \) could be obtained assuming different conditions on the sample dependence and behavior of the individuals in the population leading to different models. In this paper we will concentrate on those models which have two common assumptions - (1) population is closed within the time of different samples taken, (2) individual are homogeneous with respect to capture probabilities means, each individual in the population \( U \) has the same probability of capture. Let, \( p_{i01} \) and \( p_{i10} \) be the capture probability of \( i \)th individual in first sample (List 1) and second sample (List 2) respectively. Hence, capture probabilities in List 1 satisfy \( p_{i01} = p_{01} \) and in List 2 satisfy \( p_{i10} = p_{10} \) for \( i=1, 2, \ldots, N \). When information is collected independently from one extra source other than census, then it
is known as Dual-record System or Dual System. Some countries use their regular survey as the second source, e.g. US Census Bureau uses the Current Population Survey (CPS) to estimate the true population. The individuals captured in first list (made from census) are matched one-by-one with the list of individuals made from second sample. We classify all the captured individuals in $U$ according to a multinomial fashion as in Table 1. The number of missed individual by both systems is denoted as $x_{22}$ and this quantity is unknown which make the total population size $N(=x_{00})$ unknown. Expected Proportions or probabilities for each cell are also given and these notations will be followed throughout in this paper.

### Table 1: $2 \times 2$ table for Dual-record-System Model

| List 1 | In | Out | Total |
|--------|----|-----|-------|
|        |    |     |       |
| I. Observed sample numbers |   |     |       |
| In     | $x_{11}$ | $x_{12}$ | $x_{10}$ |
| Out    | $x_{21}$ | $x_{22}$ | $x_{20}$ |
| Total  | $x_{01}$ | $x_{02}$ | $x_{00} = N$ |

|        |    |     |       |
| II. Expected Proportions |   |     |       |
| In     | $p_{11}$ | $p_{12}$ | $p_{10}$ |
| Out    | $p_{21}$ | $p_{22}$ | $p_{20}$ |
| Total  | $p_{01}$ | $p_{02}$ | 1       |

#### 2.1 Time Variation Model: $M_t$

This model is very simple and widely used under the capture-recapture framework. Two additional assumptions are required for this model. One is that the two lists are causally independent. The event of being included in List 1 is independent of the event of being included in List 2. Hence, the cross-product ratio satisfies $\theta = p_{11}p_{22}/(p_{12}p_{21}) = 1$. Another assumption is time variation in the capture probabilities, i.e., $p_{01} \neq p_{10}$. Then the associated likelihood is

$$L_t(N, p_{01}, p_{10}|D) \propto \frac{N!}{x_{11}!x_{12}!x_{21}!(N-x_0)!}p_{01}^{x_{01}}p_{10}^{x_{10}}(1-p_{01})^{N-x_{01}}(1-p_{10})^{N-x_{10}}. \quad (1)$$

The corresponding maximum likelihood estimates are

$$\hat{N}_{M_t} = x_{11} + x_{12} + x_{21} + \frac{x_{12}x_{21}}{x_{11}} = \frac{x_{10}x_{01}}{x_{11}},$$

$$\hat{p}_{01,M_t} = \frac{x_{11}}{x_{10}} \quad \text{and} \quad \hat{p}_{10,M_t} = \frac{x_{11}}{x_{01}}.$$
2.2 Behavioral Response Model: $M_b$

Causal independence assumption is criticised in surveys and censuses of human populations. The concern is that an individual’s capture probability in second list may change in response to being captured in the first list. An individual who is captured in first attempt may have more (or less) chance to be included in the second list than the individual who has not been captured in first attempt. The change may occur due to different causes (see Wolter [27]). This change is grossly known as behavioral response variation. When this chance is more then the corresponding individuals are treated as recapture prone, otherwise when it is less, the individual are recapture averse. When there is no time effect, i.e. capture probabilities in first and second samples are indifferent but causal independence is violated, then the appropriate model is denoted by $M_b$. Here the odds of capture in second survey given capture in first attempt are either greater or less than the odds of capture in second survey given non-capture in first list. The probability of first capture is the same for each individual in the population. Let the

\[
\text{Prob(An individual is captured in List 1)} = p_{01} = p \text{ (say).}
\]

Hence, \(\text{Prob(An individual is captured in List 2 | he/she is not captured in List 1)} = p_{01} = p\), also The probability of second capture is also the same for each individual and let \(\text{Prob(An individual is captured in List 2 | he/she is captured in List 1)} = p_{11}/p_{01} = c\). If \(c > p\), the cross-product ratio associated with this model is greater than 1 (positive association between lists) and when \(c < p\), the cross-product ratio is less than 1 (negative association). Hence the corresponding likelihood function for this model is

\[
L_b(N, p, c|D) \propto \frac{N!}{x_{11}!x_{12}!x_{21}!(N-x_0)!}p^{x_0}c^{x_{11}}(1 - p)^{2N-x_0-x_{01}}(1 - c)^{x_{21}}
\]

and the maximum likelihood estimates are

\[
\hat{N}_{M_b} = x_0 \left\{1 - \left(\frac{x_{12}}{x_{01}}\right)^2\right\}^{-1},
\hat{p}_{M_b} = \frac{x_{01} - x_{12}}{x_{01}} \text{ and } \hat{c}_{M_b} = \frac{x_{11}}{x_{01}}.
\]

2.3 Model $M_{tb}$

When behavioral features of the model $M_b$ is relevant along with those of the model $M_t$, one will get the relatively complex model $M_{tb}$. Here, the corresponding population is said to be affected by both the time and behavior response effects. To model this situation one has to impose the assumption following Wolter ([27]) that the probability of first capture is the same for each individual in the population and that is,

\[
\text{Prob(ith individual is captured in List 1)} = p_{01} \text{ and}
\]

\[
\text{Prob(ith individual is captured in List 2 | he/she is not captured in List 1)} = p_{12}/p_{02} = p_{10}^t
\]
Since we are restricted ourselves to a homogeneous population, all these conditional and unconditional probabilities are same over all individuals in the population $U$. But this model suffers from identifiability problem due to the number of sufficient statistics $(x_{11}, x_{12}, x_{21})$ being less than the number of relevant parameters $(N, p_{01}, p_{10}^*, c)$ as coined by Otis et al. (1978, [20]). We assume here that recapture probability at second sample, $c$, is equal to a constant multiple of the probability of first time capture in second attempt, $p_{10}^*$, hence, $c = \phi p_{10}^*$ as Chao et al. (2000, [5]) adopted this from Lloyd (1994, [17]) to get rid of from the problem. Here $\phi$ represents the behavior response effect. Hence our likelihood becomes

$$L_{th}(N, p_{01}, p_{10}^*, \phi|D) \propto \frac{N!}{(N-x_0)!} \phi^{x_{11}} p_{01}^{x_{01}} p_{10}^{x_{21}} (1 - p_{01})^{N-x_{01}} (1 - p_{10}^*)^{N-x_{01}} (1 - \phi p_{10}^*)^{x_{21}}. \quad (3)$$

Since, the recapture probability of an individual in second sample is $\phi p_{10}^*$, it can be easily shown that for any population, individuals are recapture prone only if $\phi > 1$ and they are recapture averse only if $\phi < 1$. Because in those situations, cross product ratio $(\theta)$ will be greater than 1 and less than 1 respectively. When $\phi = 1$, then (3) will be equivalent to (1) as conditional $p_{10}^*$ will be equal to unconditional $p_{10}$.

Chao et al. [5] presented maximum quasi-likelihood estimates of population size $N$ and compared this estimate with unconditional and conditional MLE in case number of capturing is more than 2. However, in this article we confine ourselves in the same problem but from dual-record system, because three sources of data is hardly ever found for human population. Very few country practices Triple System method which is equivalent to capture-recapture model with number of trapping samples, $t = 3$. Notice that for capture with only one recapture scheme, i.e. $t = 2$, the dimension of parameter space can not be reduced by assuming $c = \phi p_{10}^*$ and identifiability problem will still exist. In the remainder of this section, we give a unified Bayesian technique for estimation of the size of a closed population$(N)$ based on the model $M_{tb}$. It is also to be noted that, when $\phi = 1$, then there will be no change in behavior response and therefore $M_{tb} \Rightarrow M_t$. Again, when $\phi \neq 1$ but $p_{01} = p_{10}$, then $M_{tb} \Rightarrow M_b$.

3 Solution with Informative Uniform Prior on $\phi$

3.1 Methodology

We can think $p_{10}^*$ as a function of $\phi$ and $c$ from conditional relation $p_{10}^* = c/\phi$ and $c$ can be well-estimated by $\hat{c} = x_{11}/x_{01}$. To obtain Bayesian solution through simple computation the parameter $p_{10}^*$ is dropped from the prior specification task and independent priors are assigned on $\Theta = (N, \phi, p_{01})$ as $\pi(\Theta) = \pi(p_{01})\pi(\phi)\pi(N)$. Hence conditional posterior distributions are as follows

$$\pi(p_{01}|N, \phi, D) \propto p_{01}^{x_{01}} (1 - p_{01})^{N-x_{01}} \pi(p_{01}), \quad (4)$$

$$\pi(\phi|N, p_{01}, D) \propto \phi^{x_{11}} (1 - \phi p_{10}^*)^{x_{21}} \pi(\phi), \quad (5)$$

$$\pi(N-x_0|p_{01}, \phi, D) \propto \frac{N!}{(N-x_0)!} \frac{((1 - p_{01})(1 - p_{10}^*))^{N}}{\pi(N)}, \quad (6)$$

where $p_{10}^* = \hat{c}/\phi$. The conditional posterior distributions in (4), (5) and (6) are same as that considered by Lee et al. (2003) for $M_{tb}$. Moreover, Lee et al. (2003) also takes into account a prior on $p_{10}^*$; so they have identified conditional posterior densities only for $N$ and $p_{01}$. Thus, they employed adaptive rejection sampling to generate $p_{10}^*$ and $\phi$. 
We consider the noninformative prior for \( p_{01} \) as \( \pi(p_{01}) = \text{Unif}(0, 1) \). A constant prior distribution \( \pi(\phi) = \text{Unif}(\alpha, \beta) \) is chosen for \( \phi \) which is no longer be noninformative since we need to specify its range. It follows that (4) and (5) reduce to

\[
\pi(p_{01}|N, \phi, D) \propto \text{Beta}(x_{01} + 1, N - x_{01} + 1),
\]

\[
\pi(\phi|N, p_{01}, D) \propto \text{Generalized Beta Type-1}(x_{11} + 1, x_{21} + 1, 1, \text{rate} = p_{10}^{*}) \times I_{[\alpha, \beta]}(\phi),
\]

where \( I_{[\alpha, \beta]}(\phi) \) is an indicator function for \( \phi \in [\alpha, \beta] \). Now, the hyperparameters \( \alpha \) and \( \beta \) are to be chosen. If no other information on \( \phi \) is available then choosing prior distribution is not at all easy. Lee et al. (2003) described a trial-and-error procedure for this. They opt for such \( \alpha \) and \( \beta \) for which the range of the posterior credible interval for \( \phi \) is not too close to either side of the prior limits. They argued that this procedure seems to work well when there is a high capture probability or wealth of recapture information. Usually for human population high capture probabilities can be attained. However, their approach remain very subjective.

One can always think of \( c \) as a good choice for the lower limit of \( \alpha \), since \( c = \phi p_{10}^{*} < \phi \). If it is correctly known that the population is recapture prone, then we recommend to set \( \alpha \) to 1. Our approach is not subjective as we propose to always fix the lower limit by the mle \( \hat{c} \) or by 1 depending upon the availability of knowledge about the direction of \( \phi \) value. Only upper limit \( \beta \) is left to be specified. When we have information that the population is recapture averse, then \( \beta = 1 \) will be the good choice. Otherwise we evaluate two alternatives \( \beta = 2 \) or 3. Two different priors on \( N \) are considered here:

I. Poisson prior: \( \pi(N) = \text{Poi}(\lambda) \), then conditional posterior (5) becomes

\[
\pi(N - x_{0}|p_{01}, \phi, D) \propto \text{Poi}(\lambda(1 - p_{01})(1 - p_{10}^{*})),
\]

II. Jeffrey’s prior: \( \pi(N) \propto 1/N \), then conditional posterior (5) becomes

\[
\pi(N - x_{0}|p_{01}, \phi, D) \propto \text{NegBinom}(x_{0}, \mu),
\]

where, \( \mu = 1 - (1 - p_{01})(1 - p_{10}^{*}) \) and \( p_{10}^{*} = c/\phi \approx \hat{c}/\phi \). Jeffrey’s prior for \( N \) is equivalent to the Jeffrey’s prior for hyperparameter \( \lambda \) in the poisson case, \( \pi(\lambda) \propto 1/\lambda \). Alternatively we can adopt Empirical Bayes approach for \( \lambda \), as stated in the article by George and Robert(1992). We replace \( \lambda \) by \( \hat{N}_{\text{MLE}} \). At first we generate initial values \( p_{01}^{(0)} \) and \( \phi^{(0)} \) from their corresponding prior densities and initial value \( p_{10}^{* (0)} \) is obtained using the relation \( p_{10}^{*} = \hat{c}/\phi \). Then generate \( N^{(0)} \) from the conditional posterior distribution (6) with chosen prior \( \pi(N) \) and replacing \( p_{01} \) and \( p_{10}^{*} \) by \( p_{01}^{(0)} \) and \( p_{10}^{* (0)} \). After having the initial values for all the associated parameters we perform a simple Gibbs sampling operation to obtain Bayes estimates as follows.

Step 1: Simulate \( p_{01}^{(1)} \) and \( \phi^{(1)} \) from \( \pi(p_{01}|N^{(0)}, \phi^{(0)}, D) \) and \( \pi(\phi|N^{(0)}, p_{10}^{* (0)}, D) \) respectively.

Step 2: Obtain \( p_{10}^{* (1)} \) using the relation \( p_{10}^{*} = \hat{c}/\phi \).

Step 3: Generate \( N \) from \( \pi(N - x_{0}|p_{01}^{(1)}, \phi^{(1)}, D) \) corresponding to the chosen prior on \( N \).

Step 4: Repeat the above three steps until the convergence is reached.

In these way we produce a Gibbs sequence \( \{p_{01}^{(h)}, \phi^{(h)}, N^{(h)}; h = 1, 2, 3,...\} \). The above Bayesian
approach is presented here as an alternative solution to the population size estimation problem under the model $M_b$ when only two trapping samples are available. One advantage of our approach over Lee et al. (2003) is that we have got all closed form conditional posterior densities for $\theta = \{p_0, \phi, N\}$. It is clear that to avoid the computational burden due to the complicated form of $p^*_1$ in (3), we have used the relation $p^*_1 = \hat{c}/\phi$. Another relation, $p^*_0 = x_{12}/(N-x_0)$, suggested by Llyod(1994), also can be used in this context. The initial value of parameter vector $\theta$ is coming from a wide range of choices, so it is generally unstable at the beginning of the process. To avoid the influence of the starting value, we discard the first $k$ iterative values as under the burn-in period and consider the remaining consecutive $k$ values to construct the posterior distributions for the parameters. There are many methods available for diagnostic checking of convergence (see Cowles and Bradley (1996)). In this study, we use the iterative simulation technique using multiple sequence method developed by Gelman and Rubin (1992). This is well-described in Gelman (1996,[7]). Lee et al. (2003, [16], p.p. 483) mention it briefly to workout their approach.

3.2 Numerical Illustrations

In this section we examine the performance of our approach and understand its efficiency in order to apply the method to know about the actual population size in human population when only two data sources are available. We emphasis on the dependence case between the two sources and this dependence is assumed to originate from the change in the population prior to second time survey. Let us consider a hypothetical population with $N = 500$ and two alternative possibilities $\phi = \{0.80, 1.25\}$, $0.80$ represents recapture averse and $1.25$ is for recapture prone population. Tables 4-7 and Tables 8-11 present the result from simulation study for $\phi = 0.80$ and $\phi = 1.25$ respectively. Under each value of $\phi$, we consider four pairs of population with capture probabilities $(p_{01}, p_{10})$. Those are $(0.50, 0.65)$, $(0.60, 0.70)$, $(0.80, 0.70)$ and $(0.70, 0.55)$. First two cases, where $p_{01} < p_{10}$, represent the usual situation in Dual-record system framework when a specialised survey in conducted after census operation, e.g. Post Enumeration Survey (PES). The hyper-parameters $\alpha$ and $\beta$ for $\pi(\phi) \propto U(\alpha, \beta)$ are taken as $\pi(\phi) \propto U(\alpha, \beta)$

$$
(\alpha, \beta) = \begin{cases} 
(\hat{c}, 1), (\hat{c}, 2), (\hat{c}, 3) & \text{for } \phi = 0.80 \\
(\hat{c}, 2), (\hat{c}, 3), (1, 2), (1, 3) & \text{for } \phi = 1.25
\end{cases}
$$

$U(\hat{c}, 2)$ or $U(\hat{c}, 3)$ are taken as common uniform prior specification for both the $\phi$ values, as a recommendation when no information is available on $\phi$. We will examine their performances as common prescription for any population irrespective of capture aversion or proneness. We again compare our estimate with Nour (1982) for $\phi = 1.25$ case. Because, Nour ([19]) deduced his approach for $\phi > 1$ case only and gave a hint to find the estimate of $N$ for the alternative situation on $\phi$. But following his approach and hint, it is not possible to find solution for $\phi < 1$. However, when $\phi > 1$, Nour’s estimator works considerably well. 200 data sets on $(x_{11}, x_{12}, x_{21})$ were generated. Our Bayes estimates have been obtained through simple Gibbs sampling (as in our model we found explicit form of all conditional posteriors) from 5 independent parallel chains. Burn-in period is fixed at $k=2000$ for simplicity after observing the performance of $\hat{R}^{1/2}$. For Bayesian approach, estimate of population size $N$ is obtained by averaging 200 posterior means. Based on those 200 estimates after averaging over the results obtained from different chains, the sample s.e. and the sample RMSE (Root Mean Square Error) were calculated. We also calculated the 95% credible interval(C.I.) based on sample quantile of the posterior distribution of $N$. For Nour’s approach (for $\phi > 1$ only), estimate as well as its S.E., RMSE, 95% confidence interval are computed over 200 generated datasets and present them as average estimate, sample SE, sample RMSE, 95%CI
respectively in Tables 6–9. The expected number of distinct captured individual($E(x_0)$) and values of $c$ are also listed in all the tables.

Table 2 and 3 demonstrates a case of recapture averse population with $p_{01} < p_{10}$. In Table 2, Bayes estimate corresponding to Jeffrey’s prior performs moderately for $U(\hat{c}, 1)$ and if it is not known that $\phi$ is less than 1, then the other priors $U(\hat{c}, 2)$ and $U(\hat{c}, 3)$ overestimate the $N$. Both of these happen because of low capture probabilities. Poisson prior with $\lambda = \hat{N}_M$ also misdirect the estimator due to same reason. In case of moderately high capture probabilities with $p_{01} < p_{10}$ in Table 3, Bayes estimate against $U(\hat{c}, 1)$ performs very well and other two estimates are also reasonably good. Irrespective of the availability of the knowledge that $\phi$ is less than 1, Jeffrey’s prior is always a better choice than poisson.

Table 4 considers high capture probabilities with $p_{01} > p_{10}$. Bayes estimate corresponding to the prior limit ($\hat{c}, 1$) is better than other two estimates. Though these other two estimates can be considered as good if we ignore their slight overestimation. For Table 5 also, prior $U(\hat{c}, 1)$ with Jeffrey’s prior on $N$ produces reasonably good estimate whereas the other two are highly overestimates as the second capture probability is very small. For the situation when $p_{01} > p_{10}$, we recommend the use of poisson prior when no directional information on $\phi$ is available. Overall results from Tables 2-5 indicate that Bayes estimate with prior $U(\hat{c}, 1)$ for $\phi$ works very well but one can use this range only when it is known that $\phi$ is less than 1. If such kind of directional information on $\phi$ is not available, then other two priors can be employed with poisson prior for $N$. The results in these four tables also tell us that Bayes estimates, from other two prior limits, are reasonably good for high capture probabilities. It is also noted that estimates from Poi($\lambda = \hat{N}_M$) prior have smaller RMSE than that of Jeffrey’s for high census capture probabilities.

### Table 2: Case (i): $p_{01} = 0.50, p_{10} = 0.65, \phi = 0.80, E(x_0) = 430, c = 0.577$

| Method  | Prior for $N$ | Average Estimate | Sample SE | Sample RMSE | 95% CI $^a$ |
|---------|---------------|-----------------|-----------|-------------|------------|
| Bayes   |               |                 |           |             |            |
| $U(\hat{c}, 1)$ | Jeffrey       | 482             | 12.39     | 21.62       | (461, 508) |
|         | Poi($\hat{N}_M$) | 542             | 69.93     | 81.45       | (488, 676) |
| $U(\hat{c}, 2)$ | Jeffrey       | 534             | 21.25     | 40.33       | (497, 581) |
|         | Poi($\hat{N}_M$) | 609             | 108.33    | 153.77      | (523, 823) |
| $U(\hat{c}, 3)$ | Jeffrey       | 555             | 29.03     | 61.86       | (505, 612) |
|         | Poi($\hat{N}_M$) | 619             | 115.79    | 166.14      | (527, 849) |

$^a$the 95% equal-tailed credible interval(C.I.) based on sample posterior distribution for Bayesian approach

Now we turn to the recapture prone cases in Tables 6–9. Table 6 and 7 demonstrate two population with $p_{01} < p_{10}$ which is the usual situation for human population size or total vital events estimation through post enumeration survey. Here we evaluate the performance of our proposed Bayes approach with Jeffrey’s prior (in first row), poisson priors with $\lambda = \hat{N}_M$ (in second row) corresponding to each of the four priors on $\phi$ out of which two are commonly prescribed priors. All priors considered are compared with the available Nour’s (1982) estimator. When one
Table 3: Case (ii): $p_{01} = 0.60, p_{10} = 0.70, \phi = 0.80, E(x_0) = 459, c = 0.636$

| Method | Prior for $N$ | Average Estimate | Sample SE | Sample RMSE | 95% CI      |
|--------|---------------|------------------|-----------|-------------|-------------|
| Bayes  |               |                  |           |             |             |
| $U(\hat{c},1)$ | Jeffrey | 495 | 8.38 | 9.98 | (480, 510) |
|         | Poi($\hat{N}_{Mb}$) | 504 | 11.38 | 12.03 | (485, 526) |
| $U(\hat{c},2)$ | Jeffrey | 526 | 13.56 | 28.98 | (502, 550) |
|         | Poi($\hat{N}_{Mb}$) | 529 | 16.80 | 33.54 | (495, 539) |
| $U(\hat{c},3)$ | Jeffrey | 529 | 15.44 | 36.91 | (505, 560) |
|         | Poi($\hat{N}_{Mb}$) | 529 | 17.29 | 34.19 | (496, 541) |

Table 4: Case (iii): $p_{01} = 0.80, p_{10} = 0.70, \phi = 0.80, E(x_0) = 483, c = 0.667$

| Method | Prior for $N$ | Average Estimate | Sample SE | Sample RMSE | 95% CI      |
|--------|---------------|------------------|-----------|-------------|-------------|
| Bayes  |               |                  |           |             |             |
| $U(\hat{c},1)$ | Jeffrey | 500 | 5.00 | 5.01 | (490, 508) |
|         | Poi($\hat{N}_{Mb}$) | 499 | 4.97 | 5.07 | (489, 508) |
| $U(\hat{c},2)$ | Jeffrey | 513 | 6.50 | 14.61 | (501, 525) |
|         | Poi($\hat{N}_{Mb}$) | 507 | 5.92 | 9.29 | (497, 518) |
| $U(\hat{c},3)$ | Jeffrey | 516 | 7.00 | 17.10 | (502, 528) |
|         | Poi($\hat{N}_{Mb}$) | 507 | 5.92 | 9.29 | (497, 518) |

has the information that the population of interest is recapture prone, i.e. $\phi > 1$, then the uniform priors $U(1, 2)$ or $U(1, 3)$ for $\phi$ is recommended for use. Then another Bayes estimator is produced with prior Poisson($\lambda = \hat{N}_{Nour}$) (in third row) for $N$. Tables show that the Bayes results from these two recommended priors are significantly better than the Nour’s estimate based on RMSE and closeness of confidence intervals around the true value $N$. On the contrary it is also observed that if the underlying true information is not known to us, then use of the priors $U(\hat{c},2)$ and $U(\hat{c},3)$ may be seriously misleading for relatively low capture probabilities. Results from Jeffrey’s or Poi($\lambda = \hat{N}_{Nour}$) prior on $N$ is better than that from Poi($\lambda = \hat{N}_{Mb}$). Overall, our proposed Bayes estimate performs very well and shows improvement over the Nour’s approach (available for for $\phi > 1$) when the prior on $\phi$ is properly chosen from the available knowledge on the possible range of $\phi$. When this knowledge is not available, the two commonly prescribed priors $U(\hat{c},2)$ and $U(\hat{c},3)$ perform well with $\pi(N) \equiv \text{Poi}(\lambda = \hat{N}_{Mb})$ only if chance to be captured in census is very high. This is because $U(\hat{c},2)$ or $U(\hat{c},3)$ gives a constant prior over such a wide range, so that estimates became negatively biased.

4 Alternative Solution with Conjugate Prior on $\phi$

4.1 Methodology

We support the argument made by Lee et al. (2003) that, in practice, it is necessary to restrict the range of $\phi$ to be between some $\alpha$ and $\beta$. The suggested values of $\alpha$ and $\beta$ are discussed in earlier section when some information on the possible range of $\phi$ is available. Otherwise a reasonable
Table 5: Case (iv): $p_{01} = 0.70, p_{10} = 0.55, \phi = 0.80, E(x_0) = 446, c = 0.511$

| Method   | Prior for N | Average Estimate | Sample SE | Sample RMSE | 95% CI         |
|----------|-------------|------------------|-----------|-------------|----------------|
| Bayes    |             |                  |           |             |                |
| U(\hat{c},1) | Jeffrey   | 483              | 8.72      | 19.31       | (468, 498)     |
| Poi(\hat{N}_{M_b}) |          | 481              | 8.90      | 21.08       | (466, 497)     |
| U(\hat{c},2) | Jeffrey   | 522              | 20.17     | 29.77       | (494, 562)     |
| Poi(\hat{N}_{M_b}) |          | 504              | 16.51     | 17.07       | (481, 538)     |
| U(\hat{c},3) | Jeffrey   | 545              | 35.43     | 57.98       | (503, 614)     |
| Poi(\hat{N}_{M_b}) |          | 513              | 21.89     | 25.33       | (484, 555)     |

Table 6: Case (v): $p_{01} = 0.50, p_{10} = 0.65, \phi = 1.25, E(x_0) = 394, c = 0.722$

| Method   | Prior for N | Average Estimate | Sample SE | Sample RMSE | 95% CI         |
|----------|-------------|------------------|-----------|-------------|----------------|
| NOUR     |             |                  |           |             |                |
| Bayes    |             |                  |           |             |                |
| U(\hat{c},2) | Jeffrey   | 433              | 12.57     | 67.88       | (410, 459)     |
| Poi(\hat{N}_{M_b}) |          | 447              | 16.33     | 55.68       | (416, 481)     |
| U(\hat{c},3) | Jeffrey   | 436              | 13.54     | 65.16       | (412, 465)     |
| Poi(\hat{N}_{M_b}) |          | 449              | 16.86     | 53.98       | (417, 484)     |
| U(1,2)   | Jeffrey     | 497              | 20.51     | 20.89       | (459, 539)     |
| Poi(\hat{N}_{M_b}) |          | 520              | 29.73     | 35.92       | (469, 582)     |
| Poi(\hat{N}_{Nour}) |         | 489              | 17.99     | 21.04       | (456, 526)     |
| U(1,3)   | Jeffrey   | 505              | 23.34     | 23.88       | (463, 554)     |
| Poi(\hat{N}_{M_b}) |          | 525              | 31.18     | 40.10       | (471, 592)     |
| Poi(\hat{N}_{Nour}) |         | 493              | 18.83     | 20.17       | (458, 532)     |

common choice for $[\alpha, \beta]$ is $[\hat{c}, 2]$ or $[\hat{c}, 3]$. Through simulation study in last section we notice that these common choices may yield confidence intervals too wide to give efficient result or it may lead to misleading results in some situations, if we consider uniform or constant type prior on that wide range. Hence, it is quite clear that if any one wants to use constant or uniform prior on \phi, then it is almost necessary to shrink the domain of the prior applying the correct knowledge available on \phi. In practice, for human population, demographic or any other beneficiary type survey reflects a recapture prone nature among the individuals in the population. But, if no such information is available, can we build up a Bayesian method which will be able to produce a reasonably good estimate? In this section we try to set an informative prior for \phi, of course within a finite range $[\alpha, \beta]$. Then we investigate how the Bayes estimate performs if a conjugate prior is elicited with empirically estimated hyperparameters. Specifically, we like to answer the following questions after formation of the approach. (1) Does this Bayes estimate perform better than the constant prior case when one has the directional information on \phi? and (2) Is it possible to suggest an unique Bayes estimator, from a comparative study of suitable loss functions, even when we do not have any information on \phi?
### Table 7: Case (vi): $p_{01} = 0.60, p_{10} = 0.70, \phi = 1.25, E(x_0) = 422, c = 0.761$

| Method  | Prior for $N$ | Average $N$ | Sample SE | Sample RMSE | 95% CI |
|---------|---------------|-------------|-----------|-------------|-------|
| NOUR    |               | 487         | 13.18     | 18.47       | (461, 512) |
| Bayes   |               |             |           |             |       |
| U(\hat{c},2) | Jeffrey | 448         | 10.24     | 53.38       | (428, 466) |
|         | Poi($\hat{N}_{Mb}$) | 450         | 10.68     | 51.28       | (428, 470) |
| U(\hat{c},3) | Jeffrey | 448         | 10.42     | 52.86       | (428, 467) |
|         | Poi($\hat{N}_{Mb}$) | 450         | 10.80     | 50.92       | (428, 470) |
| U(1,2)  | Jeffrey       | 491         | 15.36     | 17.97       | (460, 521) |
|         | Poi($\hat{N}_{Mb}$) | 492         | 15.84     | 17.77       | (460, 524) |
|         | Poi($\hat{N}_{Nour}$) | 488         | 14.57     | 18.78       | (459, 517) |
| U(1,3)  | Jeffrey       | 494         | 16.70     | 17.68       | (461, 527) |
|         | Poi($\hat{N}_{Mb}$) | 494         | 16.48     | 17.57       | (461, 527) |
|         | Poi($\hat{N}_{Nour}$) | 490         | 15.19     | 18.20       | (460, 520) |

Unlike the previous case, $\phi$ is considered here as dependent on $N$ and therefore, a joint prior distribution $\pi(\Theta) = \pi(p_{01})\pi(\phi|N)\pi(N)$ is considered. Let us restrict $\phi \in [\alpha, \beta]$ and consider a conjugate prior on $\phi$, $\pi(\phi) =$ Generalized Beta Type-1($p, q, 1, rate = 1/\beta$), for given $p$, $q$ and $\beta$. Hence, $\pi(\phi | \alpha, \beta, p, q) \propto \phi^{p-1}(1-\phi)^{q-1} \times I_{[\alpha, \beta]}(\phi)$. Since, $c = \phi p_{10}^*$ and $c < 1$, then $p_{10}^*$ might be thought as a good choice for the upper limit of $\phi$, hence $\beta = p_{10}^*-1$. In practice, $c$ is a common choice for $\alpha$ and $c$ will be replaced by its mle $\hat{c}$ and $p_{10}^*$ could be obtained using the relation $p_{10}^* = x_{12}/(N - x_{01})$, suggested by Llyod(1994). This suggestion leads to the conditional posterior distribution of $\phi$ as a well-known probability density function from which one can easily generate Gibbs sample. Remaining parameters in $\Theta$ have same prior setup discussed in last section 3. The conditional posterior density will be

$$
\pi(\phi|N, \beta, D) \propto \text{Generalized Beta Type-1}(x_{11} + p, x_{21} + q, 1, rate = 1/\beta) \times I_{[\hat{c}, \beta]}(\phi), \quad (9)
$$

where $\beta = (N - x_{01})/x_{12}$ and $p$ and $q$ are chosen by equating $E_\pi(\phi | \alpha, \beta, p, q)$ with $c/p_{10}^* = c\beta$. Now since, $E_\pi(\phi | \alpha, \beta, p, q) = \beta(p + q)^{-1}$, we choose $p = b(x_{11}/x_0)$ and $q = b(x_{21}/x_0)$ where $b > 0$ is a tuning parameter that regulates the variance of the prior density such that $V_\pi(\phi) = O(b^{-1})$. Here we have used $\hat{c} = x_{11}/x_{01}$ in place of $c$. Hence we can perform a simple Gibbs sampling operation with $\pi(\phi|N, p_{01}, D)$ and other two conditional posterior densities $\pi(N|\phi, p_{01}, data)$ and $\pi(p_{01}|N, \phi, data)$ exactly same as in section 3. Because, no other classical estimate of $N$ is available for $\phi < 1$ as per our knowledge, the hyper-parameter $\lambda$ is replaced by the estimate of $N$ under $M_b$ model, denoted as $\hat{N}_{Mb}$, for poisson prior. But for the case $\phi > 1$, we judge the performances under poisson prior using both $\lambda = \hat{N}_{Mb}$ and $\lambda = \hat{N}_{Nour}$. Initial value $p_{01}^{(0)}$ is generated as in similar way stated in section 3. $\phi^{(0)}$ is simulated from the informative prior Generalized Beta Type-1 with initial $\beta^{(0)} = (\hat{N}_{Mb} - x_{01})/x_{12}$. $p_{10}^{(0)}$ will be obtained by the relation $p_{10}^{(0)} = (x_{11}/x_{01})/\phi$. Then generate $N^{(0)}$ from (6) with chosen prior $\pi(N)$ and replacing $p_{01}$ and $p_{10}^{(0)}$ by $p_{01}^{(0)}$ and $p_{10}^{(0)}$. Then we perform a simple Gibbs sampling operation as follows.

**Step 1:** Simulate $p_{01}^{(1)}$ and $\phi^{(1)}$ from $\pi(p_{01}^{(1)}|N^{(0)}, \phi^{(0)}, D)$ and $\pi(\phi^{(1)}|N^{(0)}, \beta, D)$ (in (9)) respectively,
Table 8: Case (vii): \( p_{01} = 0.80, p_{10} = 0.70, \phi = 1.25, E(x_0) = 458, c = 0.729 \)

| Method for \( N \) | Prior for \( N \) | Average Sample Estimate | SE | RMSE | 95\% CI |
|---------------------|------------------|-------------------------|----|------|---------|
| NOUR               |                  | 499                     | 8.74 | 8.76 | (481, 516) |
| Bayes              | U(\( \hat{c},2 \)) Jeffrey | 474                     | 6.63 | 26.67 | (460, 486) |
|                    | Poi(\( \hat{N}_{Mb} \))               | 473                     | 6.52 | 27.64 | (459, 485) |
|                    | U(\( \hat{c},3 \)) Jeffrey | 475                     | 6.77 | 26.03 | (460, 488) |
|                    | Poi(\( \hat{N}_{Mb} \))               | 474                     | 6.57 | 27.34 | (460, 485) |
|                    | U(1,2) Jeffrey | 499                     | 9.06 | 9.08  | (481, 517) |
|                    | Poi(\( \hat{N}_{Mb} \))               | 495                     | 8.15 | 9.61  | (478, 511) |
|                    | Poi(\( \hat{N}_{Nour} \))              | 499                     | 8.86 | 9.98  | (480, 516) |
|                    | U(1,3) Jeffrey | 502                     | 9.74 | 9.99  | (483, 521) |
|                    | Poi(\( \hat{N}_{Mb} \))               | 496                     | 8.44 | 9.17  | (479, 512) |
|                    | Poi(\( \hat{N}_{Nour} \))              | 500                     | 9.23 | 9.24  | (481, 519) |

where \( \beta = \frac{1}{p_{10}^{*0}} \).

Step 2: Obtain \( p_{10}^{*1} \) using the relation \( p_{10}^{*} = \frac{\hat{c}}{\phi} \).

Step 3: Generate \( N \) from \( \pi(N - x_0|p_{10}^{*1}, \phi^{(1)}, D) \) corresponding to chosen prior on \( N \).

Step 4: Repeat the above three steps until the convergence is reached.

Hence for a general \( h \)-th step, \( \phi^{(h)} \) is generated from its conditional density for given \( p_{10}^{* (h-1)} \) and then it used to generate \( N^{(h)} \), for \( h = 1, 2, 3,... \). Hence, the values \( \{N^{(h)}: k < h \leq 2k\} \), where \( k \) is the chosen burn-in period, are believed to be a very large sample from the resultant posterior distribution \( \pi(N|D) \). \( k \) is chosen based on the performance of \( \hat{R}^{1/2} \) as earlier. To obtain an estimate of true population size (\( N \)), we have chosen the common prescription of posterior median (\( \hat{N}_{MED} \)) or the posterior mode, which is usually referred to as MAP (maximum a posteriori) estimator.

Casella (1986) and Raftery (1988) suggested another estimate obtained by minimizing the squared relative error loss function

\[
L(N, \hat{N}) = \left( \frac{N - \hat{N}}{N} \right)^2 .
\]

The corresponding estimate is \( \hat{N}_{SRE} = E_{\pi(N|D)}(N^{-1})/E_{\pi(N|D)}(N^{-2}) \). Estimate of \( N \) with its s.e. are obtained over 200 posterior replications. One feature of this setup is that though it uses informative prior but the parameters \( (p, q) \) are taken as functions of data for given tuning value \( b \). Hence, one can say this is an empirical Bayes procedure. Now we try to answer the questions already raised in the first paragraph of this section.

4.2 Numerical Illustrations

The numerical illustrations of this second approach are presented according to both the directional restriction on \( \phi \). Here prior belief on \( \phi \) is considered with a reasonable value of \( b = 20 \). Table 10 shows the performance of our Bayes estimator with informative prior when \( \phi < 1 \). We have observed the performance of \( \hat{R}^{1/2} \) and fixed \( k \) at 7000. It is clear that MAP-based estimator is far better than other Bayes estimates for \( N \) from both the Jeffrey’s and Poisson prior. When \( \phi > 1 \), \( \hat{N}_{MED} \) and \( \hat{N}_{SRE} \) performs better than \( \hat{N}_{MAP} \).
Table 9: Case (viii): \( p_{01} = 0.70, p_{10} = 0.55, \phi = 1.25, E(x_0) = 420, c = 0.585 \)

| Method         | Prior for \( N \) | Average Estimate | Sample SE | Sample RMSE | 95% CI       |
|----------------|-------------------|------------------|-----------|-------------|--------------|
| NOUR          | -                 | 499              | 13.53     | 13.55       | (473, 523)   |
| Bayes         |                   |                  |           |             |              |
| U(\( \hat{c},2 \)) Jeffrey | 458              | 10.56            | 43.66     | (436, 477)  |
| Poi(\( \hat{N}_{M_b} \)) | 452              | 9.75             | 49.07     | (432, 470)  |
| U(\( \hat{c},3 \)) Jeffrey | 462              | 11.50            | 39.49     | (440, 483)  |
| Poi(\( \hat{N}_{M_b} \)) | 454              | 10.11            | 47.07     | (434, 472)  |
| U(1,2) Jeffrey | 511              | 17.20            | 20.50     | (481, 543)  |
| Poi(\( \hat{N}_{M_b} \)) | 489              | 12.81            | 16.84     | (463, 511)  |
| Poi(\( \hat{N}_{Nour} \)) | 505              | 15.47            | 16.32     | (478, 533)  |
| U(1,3) Jeffrey | 528              | 20.72            | 34.71     | (487, 568)  |
| Poi(\( \hat{N}_{M_b} \)) | 495              | 13.47            | 14.33     | (469, 518)  |
| Poi(\( \hat{N}_{Nour} \)) | 513              | 16.47            | 20.93     | (482, 542)  |

Table 11 tells us that the estimator \( \hat{N}_{SRE} \) is best among these three estimators and one can use it with prior \( \text{Poi}(\lambda = N_{Nour}) \) when it is known that underlying \( \lambda \) is greater than one. Jeffrey’s prior also works well for both the \( \hat{N}_{SRE} \) and \( \hat{N}_{MED} \). Hence, to the question no. 1 we have raised previously in this section, we answer that this Bayes estimate performs uniformly better than constant prior case when \( \phi > 1 \). Indeed, \( \phi > 1 \) corresponds to the most likely case for human population. For high capture probabilities and \( p_{01} < p_{10} \), Bayes estimate \( \hat{N}_{MAP} \) from conjugate generalized beta prior performs relatively better than the constant prior case when \( \phi < 1 \) (Compare the results corresponding to the prior \( U(\hat{c},1) \) for Table 3 and 4 with Table 10). \( p_{01} < p_{10} \) is almost certain for Dual-record-Record System when a specialised survey is conducted following the census count, e.g. Post Enumeration Survey (PES). It is also observed from Table 11 that \( \hat{N}_{MAP} \) with \( \text{Poi}(\lambda = N_{M_b}) \) performs reasonably well for \( \phi > 1 \) when \( p_{01} < p_{10} \), though not as good as the other two estimators. Hence, in order to find an unique estimator corresponding to the situation when no information on \( \phi \) is available, the \( \hat{N}_{MAP} \) estimator using \( \text{Poi}(\lambda = N_{M_b}) \) prior is suggested as a reasonably good choice under usual scenario of Dual-record System for human population. Figure 2 shows the posterior distributions for \( N \) based on generalised beta prior on \( \phi \) and \( \pi(N) \propto \text{Poi}(\lambda = N_{M_b}) \) for both the cases \( \phi = 0.80 \) and \( \phi = 1.25 \).

5 Real Data Application

Greenfield (1975, [10]) reports some of the results of a Population Change Survey to estimate birth, death and migration rates conducted by the National Statistical Office in Malawi between 1970 and 1972. The sample was stratified into five areas: Blantyre, Lilongwe and Zomba urban areas; other urban areas and rural areas. A Dual-record system of data collection was employed to record births, deaths and migration which occurred during that time interval. The results of the two recording systems were matched and eliminated any events which had been recorded wrongly. However, in this article we confine our attention to data on death records only. According to the estimate of \( c \), we have chosen only two strata - Lilongwe (\( \hat{c} = 0.593 \)) and Other urban areas (\( \hat{c} = 0.839 \)). Table 12 summarizes the results of the second year operation of the survey on death count only for the two strata - Lilongwe and the Other urban areas. Results on death count for
Table 10: Estimates of $N$ with estimated s.e. in () for different situations in Tables 4-7 for $\phi = 0.80$ under different loss functions. The burn-in period($k$) is fixed at 7000.

| Case Corresponding to | Prior | $\hat{N}_{MAP}$ | $\hat{N}_{MED}$ | $\hat{N}_{SRE}$ |
|-----------------------|-------|------------------|------------------|------------------|
| Table 2               | Jeffrey | 466 (9.87) | 538 (16.60) | 531 (14.36) |
|                       | Poi($\lambda = \hat{N}_{Mb}$) | 554 (10.12) | 662 (20.36) | 602 (16.09) |
| Table 3               | Jeffrey | 488 (7.13) | 540 (11.08) | 539 (10.44) |
|                       | Poi($\lambda = \hat{N}_{Mb}$) | 498 (10.12) | 560 (20.36) | 550 (16.09) |
| Table 4               | Jeffrey | 499 (5.20) | 527 (7.27) | 533 (7.75) |
|                       | Poi($\lambda = \hat{N}_{Mb}$) | 498 (6.96) | 524 (7.37) | 525 (7.38) |
| Table 5               | Jeffrey | 469 (7.26) | 514 (9.60) | 519 (10.86) |
|                       | Poi($\lambda = \hat{N}_{Mb}$) | 468 (9.50) | 511 (9.75) | 509 (9.19) |

the other three strata have been omitted at this stage since the only purpose of the example is to show the relative efficiency of the classical and Bayes solutions used for estimating the unknown population size. Nour (1982, [19]) used this dataset and estimated the death sizes as 378 and 3046 for Lilongwe and Other urban areas respectively assuming the fact that two data sources in a Dual-record collection systems are positively correlated in a human demographic study.

Adopting the same assumption made by Nour (1982) on this data, we employ our Bayes approach (discussed in section 3) to estimate the population size considering $\phi > 1$. Table 13 presents the results of our Bayesian approach proposed in section 3 for both the strata. The prior distribution for $p_{01}$ is uniform over $(0, 1)$ and the common prescription on prior distribution for $\phi$ are considered at first. We always use $\hat{c}$ as a choice of lower limit $\alpha$ and 2 or 3 as a choice of upper limit $\beta$ for $\phi$ unless we know whether the population is recapture averse or recapture prone. Hence, the priors are $U(\hat{c}, 2)$ and $U(\hat{c}, 3)$. The priors for $N$ are Jeffery’s (upper case) and Poisson (lower case). For poisson prior, the hyper-parameter $\lambda$ is replaced by the classical estimate $\hat{N}_{Mb}$. Since the data is already assumed as generated from a capture prone population, so we have also tried the poisson prior with $\lambda = \hat{N}_{Nour}$. For each case, we generate 100 parallel chains from different randomly selected starting points of $p_{01}, \phi$. Then, we compute $\hat{R}^{1/2}$ (with respect to $N$) to determine the burn-in period $k$. The procedure to select burn-in period $k$ is clearly described in Gelman (1996). As we can see from Figure 1, $\hat{R}^{1/2}$ becomes smaller than 1.1 after 3500 iterations in each of the two cases for the strata Lilongwe and Other urban areas but $\hat{R}^{1/2}$ becomes stable from 5000. We therefore fix $k$ simply at 5000 for both of the Lilongwe and Other urban areas and record the remaining 5000 values in each chain respectively. This way the recorded values mimic the posterior distribution of $N$.

For Lilongwe, we first put a uniform prior on $\phi$ with range from $\hat{c}$ to 2 (a reasonable upper limit in general situations). Then we get a 95% creditable interval for $\phi$ ranging from 0.806 to 1.057. If we extend the upper limit of $\phi$ to 3 or 5 (a very wide range for $\phi$), we find that the outcomes are quite similar. Instead of Jeffrey’s, if poisson prior is used, estimates are quite similar but poisson
Table 11: Estimates of N with estimated s.e. in () for different situations in Tables 8-11 for \( \phi = 1.25 \) under different loss functions. The burn-in period(\( k \)) is fixed at 5000.

| Case Corresponding to | Prior | \( \hat{N}_{MAP} \) | \( \hat{N}_{MED} \) | \( \hat{N}_{SRE} \) |
|-----------------------|-------|---------------------|---------------------|---------------------|
| Table 6               | Jeffrey | 426 (11.00) | 499 (17.45) | 496 (14.57) |
|                       | Poi(\( \lambda = \hat{N}_{Mb} \)) | 502 (45.25) | 528 (30.31) | 510 (22.79) |
|                       | Poi(\( \lambda = \hat{N}_{Nour} \)) | 478 (28.33) | 492 (16.07) | 484 (13.83) |
| Table 7               | Jeffrey | 451 (10.19) | 514 (18.71) | 514 (14.74) |
|                       | Poi(\( \lambda = \hat{N}_{Mb} \)) | 502 (45.25) | 528 (30.31) | 510 (22.79) |
|                       | Poi(\( \lambda = \hat{N}_{Nour} \)) | 478 (28.33) | 492 (16.07) | 484 (13.83) |
| Table 8               | Jeffrey | 476 (8.96) | 496 (9.36) | 503 (9.02) |
|                       | Poi(\( \lambda = \hat{N}_{Mb} \)) | 481 (7.81) | 492 (8.05) | 494 (7.22) |
|                       | Poi(\( \lambda = \hat{N}_{Nour} \)) | 482 (11.45) | 498 (9.60) | 500 (8.46) |
| Table 9               | Jeffrey | 436 (8.63) | 495 (12.94) | 496 (11.80) |
|                       | Poi(\( \lambda = \hat{N}_{Mb} \)) | 431 (8.18) | 478 (10.59) | 473 (9.47) |
|                       | Poi(\( \lambda = \hat{N}_{Nour} \)) | 434 (8.66) | 497 (14.75) | 489 (12.41) |

Table 12: Malawi Population Change Survey 2nd Year results for death count (grossed up)

| Stratum                | \( x_{11} \) | \( x_{21} \) | \( x_{12} \) | \( x_{0} \) |
|------------------------|--------------|--------------|--------------|--------------|
| Lilongwe               | 192          | 132          | 24           | 348          |
| Other urban areas      | 1645         | 315          | 805          | 2765         |

\( x_{0} = x_{11} + x_{21} + x_{12}, \) total number of distinct deaths counted in the area by the two surveys.

Data source: Greenfield (1975)

prior gives relatively efficient results. As \( N \) is large, effect of \( \beta \) value on the estimate becomes less. Hence, for this data, we believe the number of deaths in Lilongwe is around 360 with a 95% credible interval from 357 to 365. Following same approach, we may say that estimate of the number of deaths is 2856 for Other urban areas with 95% credible interval from 2828 to 2929. Our Bayes estimate suggests that two data sources on the death records are not positively dependent which differs from the underlying assumption taken by Nour (1982, [19]).

Secondly, we use the information that the two data sources in Dual-record collection systems are positively correlated in the context of most of the demographic applications. That means \( \phi \) is greater than 1. Here we use this information and employ the two priors \( U(1, 2) \) and \( U(1, 3) \) as per our recommendation for this type of data. Under each of these two priors on \( \phi \), we add another prior on \( N \) as Poi(\( \lambda = \hat{N}_{Mb} \)). In the bottom half of both panels of the Table 13, for each of \( U(1, 2) \) and \( U(1, 3) \), first, second and third row correspond to the Jeffrey’s, Poi(\( \lambda = \hat{N}_{Mb} \)) and Poi(\( \lambda = \hat{N}_{Nour} \)) priors on \( N \) respectively. Result from \( U(1, 3) \) do not differ significantly from \( U(1, 2) \) as \( N \) is large enough in case of Jeffrey’s and Poi(\( \lambda = \hat{N}_{Nour} \)). Hence, our Bayes estimate with
properly chosen poisson prior on $N$ with $\lambda = \hat{N}_{\text{Nour}}$ gives more efficient results than that of Nour’s. Moreover, it is also found that the two data sources are positively correlated and in Lilongwe, peoples are more keen to capture the death records again than the Other urban areas in survey time. This second set of results indicate that our Bayes estimate with constant prior performs very well under the assumption that here, $\phi$ is more than 1.

6 Summary and Conclusions

We have presented two Bayesian approaches under a general framework for dual-record system where behaviour response effect might play a significant role along with time variation effect. Here we suggest an efficient Bayes estimator conditionally and unconditionally on the directional knowledge available on $\phi$ respectively. Both of our approaches have a key feature of simpler computation than the methods existing in the literature so far. The first one is formulated with uniform prior and the second one depends on conjugate prior constructed exactly based on functional relationships among the underlying parameters. Some features of the first approach with uniform prior on behaviour effect ($\phi$) are: Noninformative prior for $N$ and $p_{01}$ is used and a reasonable range for $\phi$ is always available with or without the help of the available information on $\phi$. When $\phi < 1$, specification of the lower bound of $\phi$ by $\hat{c}$ works successfully. In case of $\phi < 1$, prior $U(\hat{c}, 1)$ works better. But when $\phi > 1$, our study recommends that Bayes estimate with $\pi(\phi) \equiv U(1, 2)$ or $U(1, 3)$ and $\pi(N) \propto N^{-1}$ is expected to be superior than Nour’s estimate in terms of smaller rmse and reasonably better CI. For $\phi > 1$, $\beta$ is not at all influential if the nature of $\phi$ is correctly known. We also used poisson prior on $N$ and solved the problem with empirical Bayes approach. It is found that estimates from poisson prior with $\lambda = \hat{N}_{M_b}$ are less efficient than Jeffrey’s. Hence, we conclude that the first Bayes approach performs very well when correct information on the possible range of underlying $\phi$ is available and used. In practice, experts can usually judge whether the specified population is either recapture prone (i.e. $\phi > 1$) or recapture averse (i.e. $\phi < 1$). If it is so, our Bayes estimate with informative uniform prior on $\phi$ is a significant improvement over Nour ([19]) in terms of efficiency.

An alternative Bayes approach with informative generalised beta prior is also proposed when there is no reliable information available on $\phi$. Some features of this empirical Bayesian approach with informative beta prior on behaviour effect($\phi$) are the following. For $\phi < 1$, MAP-based estimates are very efficient (compared to estimates minimising the other two loss functions) when the capture probabilities are high. In contrast, the other two estimates $\hat{N}_{\text{MED}}$ and $\hat{N}_{\text{SRE}}$ with $\pi(N) \equiv \text{Poi}(\lambda = \hat{N}_{\text{Nour}})$ perform relatively better than $\hat{N}_{\text{MAP}}$ for $\phi > 1$. But one can use $\hat{N}_{\text{MAP}}$ with $\pi(N) \equiv \text{Poi}(\lambda = \hat{N}_{M_b})$ in the second Bayes approach when it is not known in advance whether $\phi > 1$ or not. From the empirical evaluation of these two approaches on same populations, it is found that the second approach improves the performance for $\phi > 1$. The second approach works better than the first when we use priors $U(\hat{c}, 2)$ and $U(\hat{c}, 3)$ due to non-availability of the directional prior information on $\phi$. Hence, the proposed methods can be used to have a better and easily computable estimate of population size. If we have the directional knowledge on $\phi$, then Bayesian approach with conjugate prior works better than the first approach but obviously second approach is less robust. Beside the computational advantage, our methods are transparent and relatively easy to explain. Though our methods incorporate subjective elements through the choice of priors, this subjectiveness helps the underlying model to successfully get rid of the identifiability problem.
Acknowledgement

This work is partially supported by the research fellowship award (No. 09/093(0125)/2010–EMR–I) given to the first author from Council of Scientific and Industrial Research (CSIR), India. We are grateful to Dr. Sourabh Bhattacharya for his valuable comments and suggestions which has helped us to improve the paper considerably.

References

[1] Bishop, Y., Fienberg, S. and Holland, P. (1975). *Discrete Multivariate Analysis, Theory and Practice*. MIT Press, Cambridge, Massachusetts.

[2] Castledine, B. J. (1981). A Bayesian Analysis of Multiple Recapture. *J. Amer. Statist. Assoc.* 81, 338-346.

[3] ChandraSekar, C. and Deming, W.E. (1949). On a method of estimating birth and death rates and the extent of registration. *J. Amer. Statist. Assoc.* 44, 101-115.

[4] Chao, A., Tsay, P. K., Lin, S. H., Shau, W. Y. and Chao, D. Y. (1996). Population Size Estimation for CaptureRecapture Models With Applications to Epidemiological Data. *Proceedings of the American Statistical Association, Biometrics Section*, 108-117.

[5] Chao, A., Chu, W. and Chiu, H.H. (2000). Capture-Recapture when Time and Behavioral Response Affect Capture Probabilities. *Biometrics* 56, 427-433.

[6] Chao, A., Tsay, P. K., Lin, S. H., Shau, W. Y. and Chao, D. Y. (2001). Tutorial in Biostatistics: The Application of Capture-Recapture Models to Epidemiological Data. *Statistics in Medicine* 20, 3123-3157.

[7] 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.

[8] George, E. I. and Robert, C. P. (1990). Capture-recapture models and Bayesian sampling. Technical Report No. 435. Department of Statistics, Stanford University, Stanford, California.

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

[10] Greenfield, C. C. (1975). On the estimation of a missing cell in a 2 × 2 contingency table. *J. R. Statist. Soc. A* 138, 51-61.

[11] Hook, E. B. and Regal, R. R. (1995). Capture-recapture methods in epidemiology: Methods and limitations. *Epidemiologic Review* 17, 243-264.

[12] Huggins, R. (1989). On the statistical analysis of capture-recapture experiments. *Biometrika* 76, 133-140.

[13] El-Khorazaty, M. N. (1975). *Methodological strategies for the analysis of categorical data from multiple-record systems*. An unpublished Ph.D. thesis submitted to Department of Biostatistics, University of North Carolina at Chapel Hill. Institute of Statistics Mimeo Series No. 1019.
[14] Laplace, P. S. (1786). Sur les Naissances, les Mariages et les Morts. Histoire de L’Academie Royale des Sciences, 1783, Paris.

[15] Lee, S. M. and Chen, C. W. S. (1998). Bayesian inference of Population Size for behavioral response models. Statistica Sinica 8, 1233-1247.

[16] Lee, S. M., Hwang, W. H. and Huang, L. H. (2003). Bayes estimation of Population Size from Capture-recapture Models with Time Variation and Behavior response. Statistica Sinica 13, 477-494.

[17] Lloyd, C. J. (1994). Efficiency of Martingle Methods in Recapture Studies. Biometrika 81, 305-315.

[18] Marks, E.S., Seltzer, W. and Krotki, K. R. (1974). Population Growth Estimation. The Population Council, New York.

[19] Nour, E. S. (1982). On the Estimation of the Total Number of Vital Events with Data from Dual-record Collection Systems. J. R. Statist. Soc. A 145, 106-116.

[20] Otis, D. L., Burnham, K. P., White, G. C. and Anderson, D. R. (1978). Statistical Inference from Capture Data on Closed Animal Populations. Wildlife Monographs 62, 1-135.

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

[22] Seber, G. A. F. (1982). The Estimation of Animal Abundance and Related Parameters. 2nd edition. Charles W. Griffin, London.

[23] Seber, G. A. F. (1986). A Review of Estimating Animal Abundance. Biometrics 42, 267-292.

[24] Seber, G. A. F. (1992). A Review of Estimating Animal Abundance II. International Statistical Review, 60, 129-166.

[25] Smith, P. J. (1988). Bayesian Analysis for multiple capture-recapture surveys. Biometrics 44, 1177-1189.

[26] Smith, P. J. (1991). Bayesian Analysis for a multiple capture-recapture model. Biometrika 78, 399-407.

[27] Wolter, K. M. (1986). Some Coverage Error Models for Census Data. J. Amer. Statist. Assoc. 81, 338-346.
Figure 1: Posterior distributions for $N$ obtained from $\text{Poi}(\lambda = \hat{N}_{M_b})$ prior for $N$ and generalised beta prior for $\phi$. Figures (i) to (iv) correspond with the cases (i) to (iv) respectively under $\phi = 0.80$ and Figures (v) to (viii) for the different cases (v) to (viii) of $\phi = 1.25$
Figure 2: Plot of $\hat{R}^{1/2}$ against burn-in period $k$. Horizontal line is drawn corresponding to suggested threshold value 1.1 for $\hat{R}^{1/2}$. First row is for Lilongwe and second row is for Other urban areas. Three columns represent three prior specifications on $\phi$ respectively. X-axis is in the scale of 500.
Table 13: Bayes estimates of total number of deaths using uniform prior on $\phi$

| Method | $N$ | s.e.($N$) | 95% CI of $N$ | $\phi$ | 95% CI of $\phi$ |
|--------|-----|-----------|---------------|-------|-----------------|
|        |     |           |               |       |                 |
|        |     |           |               |       |                 |
| Lilongwe |     |           |               |       |                 |
| (c, 2) | 362 | 2.65      | (356, 368)    | 0.926 | (0.806, 1.057)  |
|        | 360 | 2.09      | (357, 365)    | 0.928 | -               |
| (c, 3) | 365 | 4.98      | (357, 376)    | 1.006 | (0.811, 1.253)  |
|        | 362 | 3.37      | (357, 370)    | 1.005 | -               |
| (1, 2) | 379 | 1.27      | (377, 382)    | 1.35  | (1.29, 1.40)    |
|        | 374 | 1.03      | (372, 376)    | 1.35  | (1.30, 1.40)    |
|        | 378 | 1.00      | (377, 380)    | 1.35  | (1.30, 1.40)    |
| (1, 3) | 386 | 3.61      | (380, 394)    | 1.52  | (1.36, 1.70)    |
|        | 378 | 2.08      | (374, 382)    | 1.52  | (1.36, 1.71)    |
|        | 383 | 2.41      | (378, 388)    | 1.52  | (1.36, 1.70)    |
|        |     |           |               |       |                 |
| Other urban areas |     |           |               |       |                 |
| (c, 2) | 2858 | 34.16    | (2822, 2932)  | 0.936 | (0.898, 1.014)  |
|        | 2856 | 29.92    | (2828, 2929)  | 0.936 | -               |
| (c, 3) | 2858 | 34.17    | (2822, 2933)  | 0.936 | (0.899, 1.015)  |
|        | 2856 | 29.93    | (2828, 2929)  | 0.936 | -               |
| (1, 2) | 3030 | 41.33    | (2980, 3128)  | 1.12  | (1.06, 1.22)    |
|        | 3056 | 35.86    | (3008, 3140)  | 1.12  | (1.06, 1.22)    |
|        | 3027 | 32.19    | (2985, 3100)  | 1.12  | (1.06, 1.22)    |
| (1, 3) | 3030 | 41.91    | (2982, 3134)  | 1.16  | (1.07, 1.22)    |
|        | 3056 | 36.41    | (3010, 3144)  | 1.12  | (1.06, 1.22)    |
|        | 3027 | 32.54    | (2985, 3103)  | 1.12  | (1.06, 1.22)    |

*a.s.e. is computed based on sample posterior distribution

b.the 95% posterior credible interval is determined based on percentile method