Research Report

Department of Statistics

No. 2003:3

Estimating the Size of Hidden Populations: A Bayesian Approach

Christian Tallberg

Department of Statistics, Stockholm University, SE-106 91 Stockholm, Sweden

Estimating the Size of Hidden Populations: A Bayesian Approach

Christian Tallberg

Abstract

The problem of estimating the size of hidden populations is considered. A practical design to obtain efficient estimators is snowball sampling which allows units to provide information not only about themselves but also about other units. In classical approaches inferences about the model are based on asymptotic theory, and accuracy of confidence statements is questionable for small sample sizes. We employ Bayesian methods enabling exact finite sample inference in terms of whole distributions of the unknown parameters given the observed data. Often, prior information on the model parameters is available. The Bayesian analysis enjoys the advantage of the possibility to implement this information into the analysis which, if properly used, should improve the estimators. Simulation results are provided where the Bayesian estimator is compared to frequentist competitors. Applications of our proposed model are illustrated with analysis to three studies of hard drug use.

Keywords: Bayesian analysis; Network sampling; Random graphs.

1 Introduction

In studies where the purpose is to estimate small, hidden and hard-to-access human populations such as heroin users, standard probability sampling designs are often inefficient. This is due to the fact that in order to yield sufficiently accurate estimates very large samples have to be drawn. Often however, contact patterns between members of the population exists, which facilitates for more effective procedures to collect data. Such procedures are link-tracing sampling designs, which means that social links are followed from

one respondent to another to obtain a sample. An example of link-tracing designs is adaptive cluster sampling which has been treated in the graph setting as well as the spatial setting by Thompson (1997) and Thompson and Seber (1996). We shall consider another example of a link-tracing procedure in which individuals in an initial sample are asked to identify acquaintances, who in turn were asked to identify acquaintances, and so on for a fixed number of stages or waves. The procedure was termed snowball sampling by Goodman (1961). Various statistical methods for snowball samples are investigated by Frank (1971, 1977, 1979), Snijders (1992), and Thompson and Frank (2000) have developed survey theory for snowball sampling. Recent publications include Spreen and Coumans (2001), where an initial simple random sample is drawn from an existing sampling frame for one part of the population, and the objective is to estimate the size of the population excluded from the frame by snowball sampling technique. Chow and Thompson (1999) consider Bayesian methods for the estimation problems under the link-tracing design used by Thompson and Frank (2000). An overview of link-tracing designs and further references to the literature on the subject is provided in Spreen (1992).

Working from the frequentist point of view, Frank and Snijders (1994) use both design-based and model-based approaches to estimate the size of a hidden population. In this paper we consider their graph model, and address the estimation of the size of a hidden population from a Bayesian viewpoint. Often, prior information is available on the parameters that one wants to estimate. Using this information effectively via a Bayesian approach should yield improved estimators. The initial sample is selected under a design posed by the applications considered in this paper.

The organization of the paper is as follows. In Section 2, the notation is outlined and necessary terminology introduced. Graph models and sampling designs are described in Section 3. Assignment of prior distributions and derivation of posterior distributions are given in Section 4. The developed methods are illustrated in Section 5 on several real data sets. Section 6 provides a simulation study where characteristics of frequentist estimators and Bayesian estimators are compared under three graph models. Some concluding remarks are given in the final section.

2 Concepts of snowball sampling

Consider a directed graph on the vertex set $V = \{1, 2, ..., v\}$. By following the notation outlined by Frank and Snijders (1994), we let V^2 denote the set of all ordered pairs (i, j) from V. By denoting the adjacency matrix y, each entry $y_{ij}, (i, j) \in V^2$ takes the value 1 if an arc is present from i to j and 0 otherwise. The diagonal entries of y are equal to 1. A sample S from the graph is a subset of vertices and a subset of vertex pairs. In the snowball sampling approach S consists of an initial sample S_0, which is a subset of V, selected by some adequate design and K waves after the initial sample.

The subset of vertices that are adjacent from vertex j is denoted by $A_j = \{i \in V : y_{ji} = 1\}$, and given by row j of y. The size of A_j is called the out-degree and is denoted by a_j. It is obtained as the sum of the elements in the jth row of y

$$a_j = |A_j| = \sum_{i=1}^{v} y_{ji}.$$

If we by

$$A(S) = \bigcup_{j \in S} A_j$$

denote the subset of vertices adjacent from vertices in S, the first wave of the snowball sample is given by $S_1 = A(S_0) \cap \bar{S}_0$. The second wave is given by $S_2 = A(S_1) \cap \bar{S}_0 \cap \bar{S}_1$, and so forth. The snowball initiated by S_0 is given by $S_0 \cup S_1 \cup ... \cup S_K$ where K is the number of waves of the snowball and S_{K+1} is the first empty set in the sequence $S_1, S_2....$

3 Graph models and sampling designs

The vertex set V is partitioned into two disjoint vertex subsets $V_1 = \{1, 2, ..., v_1\}$ and $V_2 = \{v_1 + 1, v_1 + 2, ..., v_1 + v_2\}$, where $v = v_1 + v_2$. The first vertex set V_1 consists of vertices that are known through some register, whereas V_2 is constituted by the remaining vertices which are unknown or hidden. In the context of sampling heroin users, V_1 is self-generated by the members of the population that are in need of medical treatment or social support. The unknown selection method of the registered members is modeled by Bernoulli

sampling. This implies that each member is drawn independently from the population with equal but unknown probability ρ. Thus, the size of the registered part of the population v_1, is binomial (v, ρ). Since a frame of V_1 exists, random sampling procedures can be designed exclusively on this set at the initial sampling stage. In this paper, we let the initial sample S_0 be a simple random sample without replacement subset of V_1 of size $n = |S_0|$. Our goal is then to estimate the size of the total population v. The estimation of the size of the hidden population is performed after the initial sample and the first wave of the snowball sample has been drawn.

Assume that the arc indicators are independent identically distributed Bernoulli variables with probability 1 on the diagonal (y_{ii}) and with probability β elsewhere. Let r be the number of nonloop arcs in the initial sample, s the number of arcs from the initial sample to non-registered vertices in the first wave of the snowball and q the number of arcs from the initial sample to registered vertices in the first wave. Conditionally on S_0, r is binomial$\{n(n-1), \beta\}$ distributed, s is binomial(nv_2, β) and q is binomial$\{n(v_1 - n), \beta\}$ distributed. For $k = 0, ..., n$ let m_k be the number of individuals in V_2 that are mentioned by exactly k members of S_0. Define

$$m = m_1 + m_2 + ... + m_n.$$

Then

$$\begin{aligned} m_0 &= v_2 - m \\ s &= m_1 + 2m_2 + ... + nm_n \end{aligned}$$

and $\mathbf{m} = (m_1, ...m_n)$ is multinomial$(v - v_1, p_1, ..., p_n)$ where

$$p_k = \binom{n}{k} \beta^k (1-\beta)^{n-k}.$$

We assume mutual independence between r, q and $\mathbf{m}$ conditional on v_1, β and v. Furthermore, since ρ provides no additional information about r and $\mathbf{m}$ conditional on v_1 and v, the likelihood is given by

$$\begin{aligned} & p(r, q, , \mathbf{m}, v_1 | \rho, \beta, v) \\ = \; & p(r | q, \mathbf{m}, v_1, \rho, \beta, v) \, p(q | \mathbf{m}, v_1, \rho, \beta, v) \, p(\mathbf{m} | v_1, \rho, \beta, v) \, p(v_1 | \rho, \beta, v) \\ = \; & p(r | v_1, \beta, v) \, p(q | v_1, \beta, v) \, p(\mathbf{m} | v_1, \beta, v) \, p(v_1 | \rho, v), \end{aligned}$$

assuming the distribution of v_1 does not depend on β. By inserting the proper conditional distributions given above, the likelihood is computed as

$$\begin{aligned}
& p\left(r,q,\mathbf{m},v_1\left|\rho,\beta,v\right.\right) \\
= & \binom{n\left(n-1\right)}{r}\beta^r\left(1-\beta\right)^{n(n-1)-r}\binom{n\left(v_1-n\right)}{q}\beta^q\left(1-\beta\right)^{n(v_1-n)-q} \\
& \times\left(v-v_1\right)!\left[\prod_{k=0}^{n}\frac{p_k^{m_k}}{m_k!}\right]\frac{v!}{v_1!\left(v-v_1\right)!}\rho^{v_1}\left(1-\rho\right)^{v-v_1} \\
= & \binom{n\left(n-1\right)}{r}\binom{n\left(v_1-n\right)}{q}\beta^{r+q}\left(1-\beta\right)^{n(v_1-1)-r-q}\beta^s\left(1-\beta\right)^{nm+nm_0-s} \\
& \times\left[\prod_{k=0}^{n}\frac{\binom{n}{k}^{m_k}}{m_k!}\right]\frac{v!}{v_1!}\rho^{v_1}\left(1-\rho\right)^{v-v_1} \\
= & \frac{v!}{\left(v-v_1-m\right)!v_1!}\beta^t\left(1-\beta\right)^{n(v-1)-t}\rho^{v_1}\left(1-\rho\right)^{v-v_1} \\
& \times\binom{n\left(n-1\right)}{r}\binom{n\left(v_1-n\right)}{q}\prod_{k=1}^{n}\frac{\binom{n}{k}^{m_k}}{m_k!},
\end{aligned}$$

where $v=v_1+m+m_0$ and $t=r+q+s$. By discarding a multiplicative constant which does not depend on model parameters, the essential part of the likelihood is given by

$$f\left(r,q,\mathbf{m},v_1\left|\rho,\beta,v\right.\right)=\frac{v!}{\left(v-v_1-m\right)!}\beta^t\left(1-\beta\right)^{n(v-1)-t}\rho^{v_1}\left(1-\rho\right)^{v-v_1}. \quad (1)$$

4 Prior and posterior distributions

The Bayesian setting allows us to incorporate our subjective information of the model parameters, prior to looking at data, into the inference stage in form of a probability distribution. We shall assume mutual independence between β,ρ and v a priori, and that $\beta\sim$beta(a,b) and $\rho\sim$beta(a_ρ,b_ρ). The prior distribution of v is discussed below.

We consider a statistical model where the unknown parameters β,ρ and v are included, although β and ρ are not of our immediate concern. Despite describing relevant aspects of the reality they are modeling, they are regarded as nuisance parameters and it is our desire to eliminate them from the analysis

in order to concentrate on v. From a Bayesian point of view, this is achieved by computing the marginal posterior distribution of v in the following way

$$\begin{aligned} & p\left(v\left|r,q,\mathbf{m},v_1\right.\right) \\ \propto\ & p\left(r,q,\mathbf{m},v_1\left|v\right.\right)p\left(v\right) \\ =\ & \int_0^1\int_0^1 p\left(r,q,\mathbf{m},v_1\left|v,\beta,\rho\right.\right)p\left(v\right)p\left(\rho\right)p\left(\beta\right)d\rho d\beta. \end{aligned}$$

If we insert the likelihood derived in (1) and the assumed priors, the posterior distribution of v is given by

$$\begin{aligned} p\left(v\left|r,q,\mathbf{m},v_1\right.\right) \propto & \int_0^1\int_0^1 \frac{v!}{(v-v_1-m)!}\beta^{t+a-1}\left(1-\beta\right)^{n(v-1)+b-t-1} \\ & \times\rho^{v_1+a_\rho-1}\left(1-\rho\right)^{v-v_1+b_\rho-1}p\left(v\right)d\beta d\rho \\ = & \frac{v!}{(v-v_1-m)!}\frac{\Gamma\left(t+a\right)\Gamma\left(n\left(v-1\right)+b-t\right)}{\Gamma\left(n\left(v-1\right)+b+a\right)} \\ & \times\frac{\Gamma\left(v_1+a_\rho\right)\Gamma\left(v-v_1+b_\rho\right)}{\Gamma\left(v+a_\rho+b_\rho\right)}p\left(v\right) \\ \propto & \frac{v!}{(v-v_1-m)!}\frac{\Gamma\left(n\left(v-1\right)+b-t\right)}{\Gamma\left(n\left(v-1\right)+b+a\right)}\frac{\Gamma\left(v-v_1+b_\rho\right)}{\Gamma\left(v+a_\rho+b_\rho\right)}p\left(v\right). \end{aligned}$$

So far nothing has been said about the priors on v. A non-informative prior for v is given by the uniform distribution over the set of integers $1,...,N$, for some upper limit N. An alternative non-informative prior advocated by Jeffreys (1961) which is also the leading term of Rissanen's (1983) prior based on information theory argument, is given by $p\left(v\right)=1/v$, $v=1,2,....$ In this paper we will consider both priors. A convenient method to obtain a discrete informative prior is to assign a proportional functional value in a continuous distribution as a probability to the associated discrete outcome; see Bernardo and Smith (1994). An appropriate choice of such a prior for v, which only takes positive values, is a zero truncated "discretized" gamma distribution.

To determine a point estimator, within the Bayesian framework, that summarizes the entire information of the distribution into a single value one has to consider the associated loss functions. Unless one has strong evidence against a reasonably symmetric posterior distribution, a natural candidate is a symmetric loss function. We will consider one of the most commonly used loss functions, the quadratic loss function. The optimal decision rule (Bayes estimator) with respect to the quadratic loss function is the posterior

mean. An alternative estimator, associated with the 0-1 loss function, is the posterior mode.

5 Empirical examples

We now illustrate the methodology presented in previous sections on three real data sets. In all three examples, the adjacency matrices involved are non-symmetric.

To proceed with the Bayesian approach, we need to specify prior distributions of the model parameters. In each of the three data sets we will use two priors on v. The first one is the "discretized" gamma distribution with various combinations of the hyperparameters, γ and θ. A way to determine γ and θ is to equate the mean $E(v) = \gamma\theta$ to a value which represents our initial belief about v, and the variance $Var(v) = \gamma\theta^2$ (as a result of the discretization, these relations are only approximate) to a value which represents our uncertainty about v. Subjective specification of the mean and variance yield the following system of equations

$$\begin{cases} \gamma = \frac{[E(v)]^2}{Var(v)} \\ \theta = \frac{Var(v)}{E(v)} \end{cases} . \tag{2}$$

The second prior is the uniform distribution which can be seen as a reference prior with a minimum of subjective prior information on the population size. This approach allows us to derive results more in line with those used in non-Bayesian settings. We avoid making choices about ρ and β by setting the hyperparameters in the beta distributions to one. That is, ρ and β are uniformly distributed over their range zero to unity.

5.1 Data set 1

The first considered data set is a study of heroin use in the town of Groningen described and analyzed by classical methods in Frank and Snijders (1994). A snowball sample of heroin users was taken, which consisted of an initial sample of size $n = 34$. In their analysis all the registered heroin users were included in the initial sample, i.e. the inclusion probability α equals 1. Interviews were carried out in which the respondents were asked to mention other heroin users in the town of Groningen. The number of heroin users in

the first wave was $m = 237$ and the number of nominations was $t = 311$, of which $r = 15$ were within the initial sample. From independent information, the police estimate the number of heroin users to $v = 800$ in Groningen. We use the police estimates as a guideline for assignment of our priors which are centered on their estimates with various precisions expressing various uncertainty about v. The hyperparameters are then determined by (2).

Figure 1 gives marginal posterior distributions of v with the corresponding prior distributions. The solid vertical lines depict the lower and upper boundaries in 95% high posterior density (HPD)-intervals. The police estimate $v = 800$ is covered by the three intervals under informative priors, and it just falls outside the interval under the non-informative prior. This result stresses that the police estimate is reasonable under the condition that the underlying model is correct.

Besides providing HPD-intervals for the Bayes estimate, without any additional difficulty the posterior distribution also allows us to answer various number of questions such us

$$\begin{aligned} \Pr\left(v \geqslant 700 \,|\text{data}\right) &= 0.262 \\ \Pr\left(v \leqslant 500 \,|\text{data}\right) &= 0.00079 \end{aligned}$$

under our non-informative prior.

5.2 Data set 2 and 3

The second data set we analyze, is network data obtained from Heerlen Drug Monitoring System (DMS). It is a study of daily users of opiates and/or other drugs like cocaine in Parkstad Limburg in 1999, analyzed previously by classical methods in Spreen and Coumans (2001) where a more detailed description of the data set is given. The number of registered hard drug users as clients at aid agencies were $v_1 = 435$. A simple random sample without replacement of size $n = 55$ was drawn from V_1. They mentioned 204 drug users of which $m = 97$ were non-registered. The reported number of nominations was $t = 302$.

The third data set, collected by DMS, is a study of the same population in 2002. The number of registered users was 326 from which a random sample without replacement of size $n = 71$ was drawn. The number of nominees was 223 constituting the first wave of which $m = 64$ was non-registered, and the number of nominations was $t = 349$.

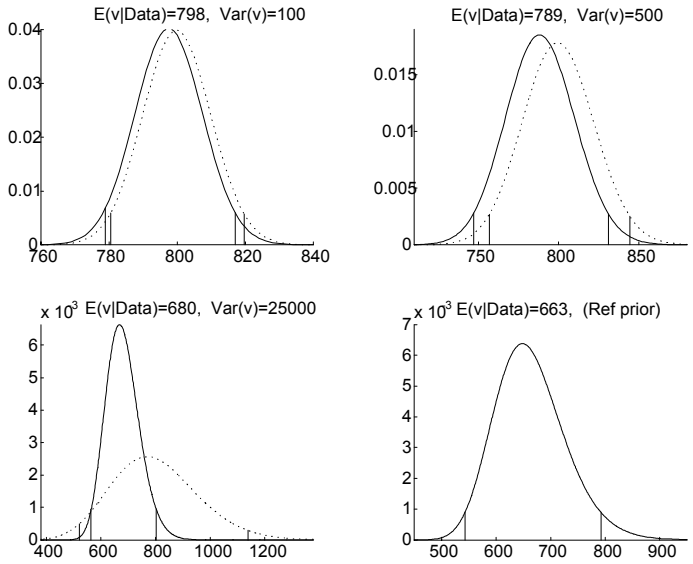

Figure 1: Priors (dotted curves) and marginal posteriors (solid curves) of v under various prior variances for data set 1.

For these two data sets, the police estimate the number of heroin users to $v = 1500$ at both time periods. As in the previous example, we use the police estimates as a guideline for assignment of our priors, and center the priors on their estimates with various precisions expressing various uncertainty about v.

Figures 2 and 3 show that the 95% HPD-intervals only cover the police estimate with high precision on the priors, i.e. when they are quite certain of their prior beliefs. Hence, according to the collected data the police rather seriously overestimates the number of drug users. In the analysis given by Frank and Snijders (1994) for the first data set, and Spreen and Coumans (2001) for the second data set, their obtained estimates are lower than the police estimates as well. We note that a Bayesian approach, where empirical observations and prior information are coupled in a natural way, offers a compromise between the police and data estimates which may be used for policy decisions.

From prior knowledge, the proportion of registered drug users, ρ, in average varies between 0.58 and 0.66 in Dutch cities. By using this information, a convenient way to determine the hyperparameters is by solving the following

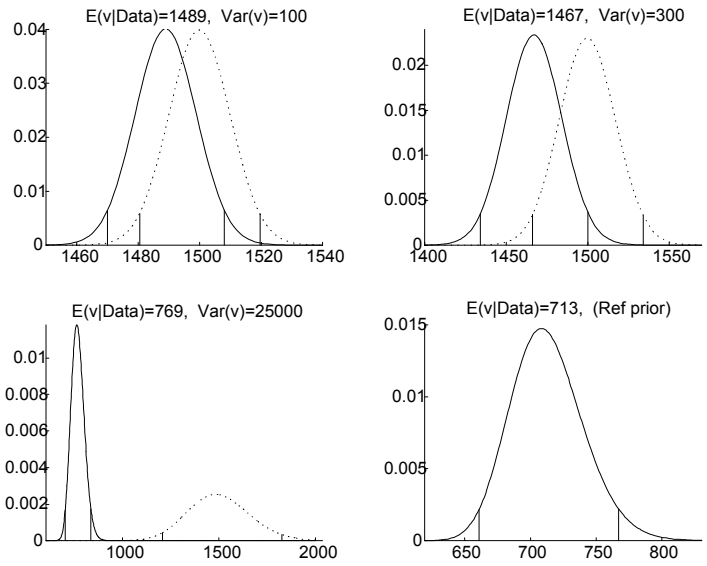

Figure 2: Priors (dotted curves) and marginal posteriors (solid curves) of v under various prior variances for data set 2.

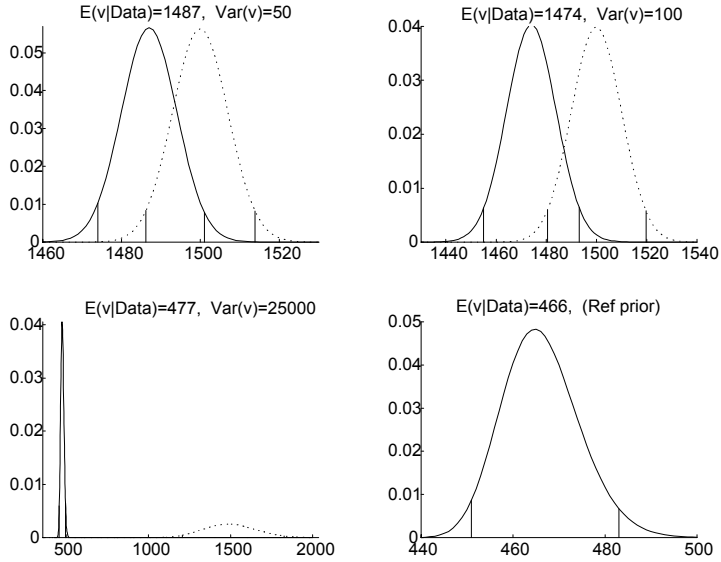

Figure 3: Priors (dotted curves) and marginal posteriors (solid curves) of v under various prior variances for data set 3.

equations

$$\Pr\left(k_l \leqslant \rho \leqslant k_u \,|a_\rho, b_\rho\right) = \lambda \tag{3}$$

$$E\left(\rho \,|a_\rho, b_\rho\right) = \frac{a_\rho}{a_\rho + b_\rho}, \tag{4}$$

where k_l and k_u are percentiles of the beta distribution. The numbers k_l and k_u, and the probability λ are set to values which represent our initial belief of ρ. By specifying the first moment in the beta distribution to $(k_l + k_u)/2$, the solution to (3) and (4) is $c = 871$ and $d = 534$ for the values $k_1 = 0.58$, $k_2 = 0.66$ and $\lambda \approx 0.998$, which corresponds to a beta distribution with mean 0.62 and variance 0.00017. In Figures 4 and 5, the posteriors of v are compared for the two considered priors on ρ on data sets 2 and 3, respectively. Conditional on a prior on v, it seems that the posteriors of v are insensitive to the choices of priors on ρ since the posteriors of v are quite similar, although the priors of ρ are not. The location of the posterior of v is slightly moved in the direction towards $\{1 - E\left(\rho \,|a_\rho, b_\rho\right)\}\rho^{-1}v_1$ if an informative prior is assigned on ρ. Simulation results, not presented here, demonstrate that the robustness of the posterior of v holds if extremely informative priors is assigned to the other nuisance parameter β.

6 Some simulation results

In this section we apply the Bayesian approach to six simulated datasets. Data was simulated as follows. For each of the population sizes $v = 100$ and $v = 1000$, one directed graph was generated from the stochastic models used by Frank and Snijders (1994). In each of the models the expected in-degree (excluding the self-loop) was fixed at 5, whereas the expected variance of the in-degrees, σ^2, differ. The three models described in their paper are the following

1. Constant in-degree: for each vertex i, five other vertices are chosen at random (without replacement) to have an arc going to i; here $\sigma^2 = 0$.

2. Bernoulli: all arcs are determined independently, and each ordered pair (i, j) with $i \neq j$ has a probability $5/(v-1)$ for an arc; here $\sigma^2 \approx 5$.

3. Two-block model: vertices are distinguished in two equal size groups and all arcs are determined independently; within the first group arcs

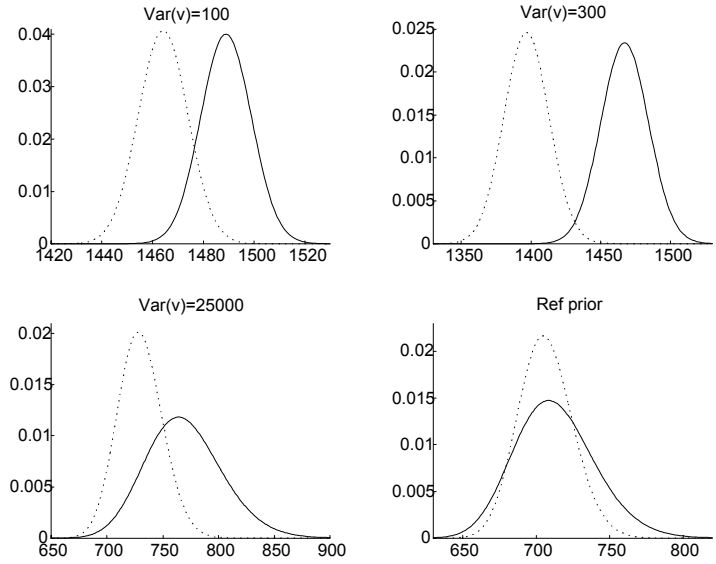

Figure 4: Marginal posteriors of v under a non-informative (dotted line) and an informative (solid line) prior on ρ for data set 2.

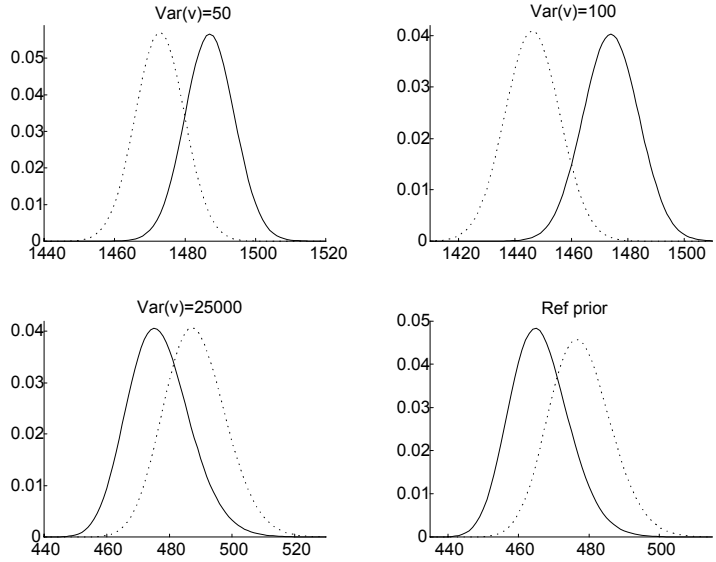

Figure 5: Marginal posteriors of v under a non-informative (dotted line) and an informative (solid line) prior on ρ for data set 3.

have probability $12/(v-2)$, within the second group $6/(v-2)$, and between the two groups $1/v$; here $\sigma^2 \approx 7.25$.

The inclusion probabilities for the registered part of the population were determined to $\rho = 0.2$ for $v = 100$ and $\rho = 0.06$ for $v = 1000$. By allowing all the registered vertices into the initial sample, i.e. $\alpha = 1$, the sizes of the initial samples are $E(n) = 20$ and $E(n) = 60$, respectively. Conditional on n, the associated sizes of the first waves can be shown to be $m \approx 50$ for $v = 100$ and $m \approx 240$ for $v = 1000$.

In our analysis we avoid making choices about ρ and β by setting the hyperparameters in the beta distributions to one. That is, ρ and β are uniformly distributed over their range zero to unity. We will consider two priors for v. Our first prior is an improper reference prior reflecting lack of information on the population size, obtained by setting $p(v)$ equal to $1/v$ for $v = 1, 2, ...$

Our second prior is an informative "discretized" gamma distribution where the hyperparameters are determined by solving the equations

$$\Pr(l_l \leqslant v \leqslant l_u \,|\gamma, \theta) = \lambda \tag{5}$$

$$E(v \,|\gamma, \theta) = \gamma\theta, \tag{6}$$

where l_l and l_u are percentiles of the gamma distribution. By specifying the first moment in the gamma distribution to $(l_l + l_u)/2$, the solutions to (5) and (6) are $\gamma = 46.44$ and $\theta = 2.15$ for $l_l = 70$, $l_u = 130$ and $\lambda \approx 0.96$ when v equals 100, which corresponds to a gamma distribution with mean 100 and variance 215. Analogously for $l_l = 800$, $l_u = 1200$, $\lambda \approx 0.96$ and $v = 1000$, equations (5) and (6) yield the solution $\gamma = 105.01$ and $\theta = 9.52$, which corresponds to a gamma distribution with mean 1000 and variance 9520. Thus, our rather informative prior beliefs are that v lies in the interval $(70, 130)$ for $v = 100$, and in the interval $(800, 1200)$ for $v = 1000$ with an approximate probability 0.96.

The simulated results are based on 20,000 iterations. In Table 1, the mean and root mean square errors of the 20,000 computed estimates of the Bayes estimator under the two given sets of priors are compared to three of the estimators, including the Horvitz-Thompson estimator, presented in Frank and Snijders (1994). The first one is the model-based maximum likelihood estimator denoted by $\hat{v}_3$. The second estimator is derived from a design-based approach where the population (digraph) is considered as fixed

and probability plays a role only via the sampling procedure. Hence, the arc indicators y_{ij} are not random but unknown. The estimator, which is a Horvitz-Thompson type estimator, is denoted by $\hat{v}_7$ and given by

$$\hat{v}_7 = \sum_{j=1}^{v} \max_i x_i y_{ij} / \hat{\eta}_j = \sum_{j \in S_0 \cup S_1} \hat{\eta}_j^{-1} \tag{7}$$

where $\hat{\eta}_j$ is an estimate of the inclusion probability given by

$$\eta_j = \Pr\{j \in S_0 \cup S_1\} = E\left(\max_i x_i y_{ij}\right) = 1 - (1-\pi)^{r_j},$$

where r_j is the in-degree of vertex j; the estimate of η_j is obtained by replacing π with $\hat{\pi} = n/\hat{v}_3$. The third "estimator", denoted by $\hat{v}_{HT}$, is the Horvitz-Thompson estimator. It is given by (7) except that calculation of the sampling inclusion probabilities requires the value of π. The choice of $\hat{v}_3$ and $\hat{v}_7$ as competitors to the Bayes estimators is because of on average better performance than the remaining estimators discussed in Frank and Snijders (1994) in terms of unbiasedness and root mean square errors. Note that $\hat{v}_7$ and $\hat{v}_{HT}$ require more information from data such as the in-degree of all sampled vertices.

In general, it seems that the Bayes estimates and their associated root mean squared errors under a non-informative prior agrees with the estimates and root mean squared errors of $\hat{v}_3$ and $\hat{v}_7$. An advantage with Bayesian methods is that we are allowed to implement conceivable information of the parameters of interest, besides the data, into the analysis in functional form of a prior distribution. It is clear that an estimator such as $\hat{v}_{2B}$, with an informative prior centered on the true value of v, enjoys a drastic decrease in the root mean square error. Furthermore, although $\hat{v}_{1B}$ is computed under a non-informative prior it is slightly better than $\hat{v}_3$ and $\hat{v}_7$ in terms of root mean square errors. Although the concept of unbiased estimators is of less importance from a Bayesian viewpoint, we see that $\hat{v}_{1B}$ seems to be less biased than its non-Bayesian competitors $\hat{v}_3$ and $\hat{v}_7$ under the Bernoulli model (Model 2), whereas $\hat{v}_7$ is less biased under the block model (Model 3).

7 Concluding remarks

In the present paper we have employed a Bayesian method to estimate the size of a hidden population. The method yields not only point estimators,

v	Model	$\hat{v}_3$	$\hat{v}_7$	$\hat{v}_{HT}$	$\hat{v}_{1B}$	$\hat{v}_{2B}$
100	1	111(17)	108(15)	100(12)	110(15)	106(9)
100	2	103(12)	103(13)	100(13)	102(11)	100(6)
100	3	97(12)	100(13)	100(13)	97(12)	97(8)
1000	1	1165(228)	1146(218)	998(115)	1146(210)	1043(63)
1000	2	1015(128)	1018(139)	1001(117)	1001(123)	992(47)
1000	3	968(130)	978(137)	1000(117)	956(129)	971(58)

Table 1: Means and root mean squared errors (between parentheses) for various estimators of v under various models.

but also provides us with a distribution of the unknown parameter, valid for all sample sizes, from which we easily can obtain various information, such as interval estimates, without any additional difficulty. By setting independent beta priors on the nuisance parameters, the marginal posterior of v is easily evaluated analytically.

A simulation study is carried out where the Bayes estimators are compared with frequentist candidates such as the maximum likelihood estimator and a Horwitz-Thompson type estimator. Under non-informative prior knowledge of v, the performance of the Bayes estimator is at least as good as its competitors in terms of point estimates and root mean square errors. Furthermore, the simulation results demonstrate that incorporating informative prior of v into the analysis increases the precision in our estimate. It is also shown via an example that the posteriors of v is rather insensitive to prior information of the nuisance parameters.

Our analysis is based on a model that assumes that the relations between any two vertices are independent. It may seem rather simplistic as a probability model in the social sciences. Efforts should be put to develop more elaborate models that considers dependency aspects. Further, we assume that the probability of a relation between two vertices is equal for all vertices. To assess more accurate estimators, an alternative could be to consider block models where the probability distribution of the relations between two vertices depends on their block affiliation. Allowing the selection probabilities in the initial sample to vary between the blocks should improve the estimators further. In a forthcoming paper we will consider blockmodeling in the context of snowball sampling.

References

[1] Bernardo, J.M. and Smith, A.F.M (1994). *Bayesian Theory.* New York: Wiley.

[2] Chow, M. and Thompson, S.K. (1999). Estimation with Link-Tracing Sampling Designs - A Bayesian Approach. Technical Report 99-03, Department of Statistics, The Penn State University.

[3] Frank, O. (1971). *Statistical Inference in Graphs.* Stockholm: Försvarets forskningsanstalt.

[4] Frank, O. (1977). Survey Sampling in Graphs. *Journal of Statistical Planning and Inference.* **1**, 235-264.

[5] Frank, O. (1979). Estimation of Population Totals by Use of Snowball Samples. In Holland, P.W. and Leinhardt, S. eds., *Perspectives on Social Network Research.* New York:Academic Press, 319-347.

[6] Frank, O. and Snijders, T.A.B. (1994). Estimating the Size of Hidden Populations Using Snowball Sampling. *Journal of Official Statistics***. 10**, 53-67.

[7] Goodman, L.A. (1961). Snowball Sampling. *Annals of Mathematical Statistics.* **20**, 572-579.

[8] Jeffreys, H. (1961). *Theory of probability.* Third Edition. Oxford at the Clarendon Press.

[9] Rissanen, J. (1983). A Universal Prior for Integers and Estimation by Minimum Description Length. *The Annals of Statistics.* **11,** No. 2, 416-431.

[10] Snijders, T.A.B. (1992). Estimation on the Basis of Snowball Samples: How to Weight? *Bulletin de Methodologie Sociologique.* **36**, 59-70.

[11] Spreen, M. (1992). Rare Populations, Hidden Populations, and Link-Tracing Designs: What and Why? *Bulletin de Methodologie Sociologique.* **36**, 34-58.

[12] Spreen, M. and Coumans, M. (2001). A Note on Network Sampling in Drug Abuse Research. *Connections***. 24(3)**, 44-59.

[13] Thompson, S.K. (1997). Adaptive sampling in behavioral surveys. In Harrison, L. and Hughes, A. eds., *The Validity of Self-Reported Drug Use: Improving the Accuracy of Survey Estimates*. NIDA Research Monograph **167**. Rockville, MD: National Institute of Drug Abuse, 296-319.

[14] Thompson, S.K. and Frank O. (2000). Model-Based Estimation With Link-Tracing Sampling Designs. *Survey Methodology*. **26**, 87-98.

[15] Thompson, S.K. and Seber, G.A.F. (1996). *Adaptive Sampling*. New York: Wiley.