

A Bayesian Justification for Random Sampling in Sample Survey

Glen Meeden
School of Statistics
University of Minnesota
Minneapolis, MN 55455
glen@stat.umn.edu

Summary

In the usual Bayesian approach to survey sampling the sampling design, plays a minimal role, at best. Although a close relationship between exchangeable prior distributions and simple random sampling has been noted; how to formally integrate simple random sampling into the Bayesian paradigm is not clear. Recently it has been argued that the sampling design can be thought of as part of a Bayesian's prior distribution. We will show here that under this scenario simple random sample can be given a Bayesian justification in survey sampling.

AMS 1991 subject classifications Primary 62D05; secondary 62A15.

Keywords and phrases: Sample survey, Bayesian inference, Random sampling

1. Introduction

In the Bayesian approach to sample survey one must specify a prior distribution over the parameter space for the entire population of possible values for the characteristic of interest. Once a sample is observed, the posterior distribution is just the conditional distribution of the unobserved units given the values of the observed units computed under the prior distribution for the population. This posterior does not depend on the sampling design used to select the sample. The Bayesian approach to finite population sampling was elegantly described in the writings of D. Basu. For further discussion see his collection of essays in Ghosh (1988). In theory one can use the prior distribution to select an optimal, purposeful sample (Zacks, 1969) but this is almost never done in practice. A problem with the Bayesian approach is that it can be difficult to find prior distributions which make use of available prior information about the population.

The sampling design plays a fundamental role in the standard frequentist theory for survey sampling. The design is the only source of randomness in the model since units in the sample are assumed to be observed without error and it is upon the selection probabilities that the frequentist properties of estimators are based.

It was noted in Godambe (1955) that many sampling designs can be thought of as being defined conditionally on the order that the units appear in the sample. He then suggested that from a theoretical perspective it is convenient to ignore this fact and just consider the unordered sample (where the order is ignored). A good reason for doing this was pointed out in Murthy (1957) where it was demonstrated that for any estimator which depends on

the ordered values there exists another estimator which only uses the unordered values which has the same expectation but smaller variance except when the two estimators are the same. This application of the Rao-Blackwell theorem was also discussed by Pathak (1961). Because of this most sampling theory has concentrated on unordered designs although Raj (1956) is one example where the order in which the sample was drawn was considered.

Meeden and Noorbaloochi (1910) noted that given a design, but before any data has been collected, the actual units that will appear in the sample are unknown. They argued that this suggests the design could be considered as part of the Bayesian's prior distribution. There they considered prior distributions which were defined in two steps. First, using a design, they randomly assigned an order to the units in the population and then conditional on a given order they specified a distribution for the possible values of the units. They showed that this approach gives a flexible method to incorporate prior information into survey sampling problems.

Ericson (1969) presented subjective Bayesian models for survey sampling when the labels contain little prior information about the units in the population. His prior distributions were exchangeable and in section 2.2 he discussed the "intimate similarities" between a subjective exchangeable prior distribution and an objective distribution introduced by the design using simple random sampling. In this note we will show how in the Meeden and Noorbaloochi (1910) framework these "intimate similarities" can be formally expressed which it turn yields a Bayesian justification for simple random sampling.

2. Some notation

Consider a population of size N and let $\Lambda = \{\alpha, \beta, \dots, \tau\}$ be a set of N labels which identify the units in the population. We let λ denoting a typical label. Let $y = \{y_\lambda : \lambda \in \Lambda\}$ denote a typical unknown set of population values. Here we assume that each y_λ can only take on the values 0 and 1. Λ is an unordered set since the population labels have no order. But since order will be important for us we let $u = (\alpha, \beta, \dots, \tau)$ denote the labels in some fixed order. Then y_u denotes y arranged in this standard default order. Hence the set of possible values for y_u is given by

$$Y(0,1) = \{y_u : \text{such that for each } \lambda \in \Lambda \ y_\lambda = 0 \text{ or } 1\} \quad (1)$$

When we write y its order does not matter while it does matter in y_u . If π is a permutation of $1, 2, \dots, N$ we let $\pi(u)$ be the permutation π applied to u to give a new order for the labels. Then $y_{\pi(u)}$ will denote the values of y arranged in the order determined by $\pi(u)$. Let Π be the set of all possible permutations of $1, 2, \dots, N$. Since order will matter for us another space of interest is

$$Y(0,1,\Pi) = \{(\pi, y_{\pi(u)}) : \text{where } y_u \in Y(0,1) \text{ and } \pi \in \Pi\} \quad (2)$$

For each fixed y_u this set will contain $N!$ points, one for each possible permutation π . For each π the point $(\pi, y_{\pi(u)})$ consists of the permutation along with the order of y_u under this permutation.

Consider a sampling design such as simple random sampling without replacement (srs) where units are selected one at a time without replacement until the desired sample size, say n , is reached. At the point where the units have been selected but before their y values are observed, we can imagine continuing the sampling procedure until all the units from the population have been selected and given an order. This is just a thought experiment and is not something that would be implemented. However we see that the srs design can be extended in a natural way to define a probability distribution on Π . When the design is srs the resulting distribution is just the uniform distribution on Π . Before the labels are selected and the characteristic of interest observed we can think of both π and $y_{\pi(u)}$ as unknown. Observing the data results in partial information about both of them. From the Bayesian perspective this means we could define a joint prior distribution over the pair on the space $Y(0,1,\Pi)$. In the next section we will compare two different approaches using this setup when little is known a priori about the population.

3. Comparing two Bayesians

If we are assuming that $Y(0,1,\Pi)$ is the space of possible values for the unknowns then a prior distribution may be defined on this space. The joint distribution for $(\pi, y_{\pi(u)})$ may be written as a marginal for π and a conditional for $y_{\pi(u)}$ given π . This means that we can use the probability distribution on Π , coming from srs, as the marginal for π . Then given an order, π , it remains to define the conditional distribution of $y_{\pi(u)}$ given π . Under this setup a prior distribution will be of the form

$$p(\pi, y_{\pi(u)}) = p(\pi)p(y_{\pi(u)} | \pi) \quad (3)$$

Then the prior distribution for y is just the marginal distribution for y under the above model.

We now wish to compare a Bayesian who incorporates srs into their prior as described above with the usual Bayesian approach where the design is ignored. We will call them the design Bayesian and the standard Bayesian and denote them by DB and SB respectively.

For the SB her prior will be

$$p(\pi, y_{\pi(u)}) = p(\pi)p(y_{\pi(u)} | \pi) = p(\pi)p(y_u)$$

where $p(\pi) = 1/N!$, but which will play no role in her analysis. Since she knows the labels her prior beliefs about the unknown y_u values are contained in $p(y_u)$. If these beliefs are exchangeable then she will use a prior of the form

$$p(y_u) = p(\theta) / \binom{N}{\theta} \quad \text{when} \quad \sum_{\lambda} y_u = \theta \quad \text{for} \quad \theta = 0, 1, \dots, N \quad (4)$$

where $\sum_{\theta=0}^N p(\theta) = 1$. Suppose in a random sample of size n there were observed x 1's and $n-x$ 0's then her posterior will be of the form

$$p_{SB}(\theta | x) \propto \frac{p(\theta)}{\binom{N}{\theta}} \binom{N-n}{\theta-x} \quad \text{for } \theta = x, x+1, \dots, N-(n-x) \quad (5)$$

We now turn to the DB who assumes that the labels exist but that they are not known to him. For him

$$Y^*(0,1) = \{y_u : \text{where } y_u \text{ has } \theta \text{ 1's followed by } N-\theta \text{ 0's for some } \theta = 0, 1, \dots, N\} \quad (6)$$

is the set of sensible possible values for y_u rather than the set given in equation 1. From this it follows that his prior distribution will be restricted to the subset of $Y(0,1,\Pi)$ of equation 2 given by

$$Y^*(0,1,\Pi) = \{(\pi, y_{\pi(u)}) : \text{where } y_u \in Y^*(0,1) \text{ and } \pi \in \Pi\} \quad (7)$$

Let $p(\theta)$ represent his prior beliefs that $\sum_x y_u = \theta$ and assume he is using srs to define $p(\pi)$. Now for a fixed value of θ , since there are no labels, there are only $\binom{N}{\theta}$ possible different orderings of the population and each such ordering must have the same probability, $p(\theta) / \binom{N}{\theta}$. Note this is exactly the same probability assigned to such an ordering as that assigned by the SB in equation 4.

But given a sample of size n with x 1's his posterior distribution will be

$$p_{DB}(\theta | x) \propto p(\theta) \binom{\theta}{x} \binom{N-\theta}{n-x} \quad \text{for } \theta = x, x+1, \dots, N-(n-x) \quad (8)$$

where he is using the design probabilities to compute his posterior. Now assuming that the two Bayesians are using the same $p(\theta)$ these two posteriors will be identical if

$$\binom{N-n}{\theta-x} / \binom{N}{\theta} \propto \binom{\theta}{x} \binom{N-\theta}{n-x} \quad \text{for } \theta = x, x+1, \dots, N-(n-x)$$

for fixed x , n and N . But this is easy to verify.

So the SB who assumes that the labels exist but carry no information and the DB who does not know the labels but makes use of the simple random sampling design when making his inferences have exactly the same posterior distribution given the sample.

4. Discussion

We have argued that once a Bayesian considers the sampling design as part of the prior distribution then in the situation where little is known about the population it makes sense that their inferences can be based on the fact that the sample was drawn at random. Furthermore their posterior distribution is identical to a posterior of a standard Bayesian

who ignores the sampling design and uses an exchangeable prior distribution. Here we have only considered the special case where 0 and 1 are the two possible values for units in the population. However the argument easily generalizes to the case where each unit in the population can take on a finite set of possible values. We believe that this framework allows one to demonstrate the formal equivalence between between a subjective exchangeable prior distribution and an objective sampling design thus giving an explicit link to the "intimate similarities" noted in Ericson (1969).

Now some might object that the DB is not really a Bayesian at all. In response we recall a result from Diaconis (1977) where he discusses a probability model for a set of N finitely exchangeable random variables which only take on the two values 0 or 1. He considers $N+1$ urns where urn θ contains θ 1's and $N-\theta$ 0's. He selects urn θ with probability $p(\theta)$ and then uses simple random sampling without replacement to exhaust the chosen urn to generate possible values for y_1, y_2, \dots, y_N . He shows that this model characterizes the family of finitely exchangeable random variables. Note the items in the urns have no labels assigned to them and the indices of the units denotes only when they were selected. Under this model the posterior distribution of θ given that we see x 1's in the first n units drawn from the selected urn is just $p_{DB}(\theta|x)$ given in equation 8. Although this result is well known as far as we know the explicit link to sample survey discussed here has never been noted.

References

1. Diaconis, P. (1977). Finite forms of de finetti's theorem on exchangeability. *Synthese*, 36:271-281.
2. Ericson, W. A. (1969). Subjective bayesian models in sampling finite populations (with discussion). *Journal of the Royal Statistical Society, Series B*, 31:195-233.
3. Ghosh, J. K. (1988). *Statistical Information and Likelihood: A Collection of Critical Essays* by D. Basu. Springer-Verlag, New York.
4. Godambe, V. P. (1955). A unified theory of sampling from finite populations. *Journal of the Royal Statistical Society, Series B*, 17:269-278.
5. Meeden, G. and Noorbaloochi, S. (2010). Ordered designs and bayesian inference in survey sampling. *Sankhya*, 72:119-135.
6. Murthy, M. N. (1957). Ordered and unordered estimators in sampling without replacement. *Sankhya*, 18:379--390.
7. Pathak, P. K. (1961). Use of order-statistic in sampling without replacement. *Sankhya A*, 23:409-414.
8. Raj, D. (1956). Some estimators in sampling with varying probabilities without replacement. *Journal of the American Statistical Association*, 51:269-284.
9. Zacks, S. (1969). Bayesian sequential designs for sampling finite populations. *Journal of the American Statistical Association*, 64:1342-1349.