

Ordered Designs and Bayesian Inference in Survey Sampling

Glen Meeden*

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

Siamak Noorbaloochi

Center for Chronic Disease Outcomes Research
Minneapolis VA Medical Center
Minneapolis, MN 55417
and Department of Medicine
University of Minnesota
Siamak.Noorbaloochi@med.va.gov

August 05

*Research supported in part by NSF Grant DMS 0406169

Ordered Designs and Bayesian Inference in Survey Sampling

Abstract

Many sampling designs, such as simple random sampling without replacement, can in principle be extended in a natural way so that the units continue to be selected until the population is exhausted. These designs impose a random order on the population. Here we show how such ordered designs can be used to define prior distributions over the population. For such priors the Bayesian analysis uses information that in standard frequentist methods is incorporated in the sampling design. The resulting methods will often have good frequentist properties.

Key Words and phrases: finite population sampling, sampling designs, Polya posterior and noninformative Bayesian.

AMS 1991 Subject Classification: Primary 62D05; Secondary 62C10

1 Introduction

Godambe noted in the beginning of his influential paper (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 prospective it is convenient to ignore this fact and just consider the unordered sample (where the order is ignored). (For samples of size n there are $n!$ possible ordered samples corresponding to each unordered sample.) Murthy (see Murthy (1957)) showed that for any ordered estimator there exists an unordered estimator with the same expectation and smaller variance except when the two estimators are the same. For these reasons most sampling theory has concentrated on unordered designs.

The design probabilities play a fundamental role in standard frequentist theory. Since units in the sample are assumed to be observed without error the only randomness in the model comes from them and it is upon these probabilities that the frequentist properties of estimators are based. However in many situations finding a sensible estimate of variance is far from clear and additional arguments need to be introduced. For example even though systematic samples can work well in practice it is difficult to justify the usual estimate of variance using design based theory.

In the Bayesian approach to statistical inference the posterior distribution summarizes the information about a parameter. This distribution depends on a probability model and a prior distribution and is conditional on the observed data. In finite population sampling the unknown parameter is just the entire population and a prior distribution must be specified over all possible values of the units in the population. Given a sample the posterior 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 which selected 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).

So we have this paradox, for one area in statistics where prior information is routinely used the standard methods cannot be given a Bayesian justification. Here we will consider prior distributions which are defined in two steps. First, we randomly assign an order to the units in the population. Then conditionally on a given order we define a distribution for the possible values of the units. This procedure allows our prior to capture some of

the information that is present in designs that select units sequentially. We will call such designs ordered designs. Our main goal is to study Bayesian models where an ordered design is used in the first step of our process for defining a prior. We will show how the resulting methods are closely related to standard frequentist methods and sometimes yield new procedures which have good frequentist properties. These models are a generalization of the Polya posterior and using them one can objectively incorporate into a prior some of the same kind of information that is encapsulated in a design.

The Polya posterior is a noninformative or objective Bayesian procedure which can be used when little or no prior information is available. The Polya posterior is related to the Bayesian bootstrap of Rubin (1981). See also Lo (1988). One advantage of the Polya posterior is that it has a stepwise Bayes justification (Hsuan (1979)) and leads to admissible procedures. This helps to explain why the Polya posterior often leads to procedures with good frequentist properties. For further discussion of the Polya posterior see Ghosh and Meeden (1997).

In section 2 we review facts about the Polya posterior. In section 3 we introduce a new way to define prior distributions in finite population sampling. In section 4 we introduce the Two Urn Model which given a sample relates the “seen” in the sample to the “unseen” remaining in the population. In section 5 we use a stepwise Bayes argument to prove an admissibility result that ties together sections 3 and 4. In section 6 we find an expression for the posterior variance of the population total for a special case of the Two Urn Model. In section 7 we consider three different applications of our approach. In section 8 we conclude with a brief discussion.

2 The Polya posterior

Consider a finite population containing N units from which we will take a sample of size n using simple random sampling without replacement. Given the sample the “Polya posterior” is a predictive joint distribution for the unobserved or unseen units in the population conditioned on the values in the sample. Given a sample this distribution is constructed as follows.

Consider two urns where the first contains the n observed units and the second the $N - n$ unobserved units. We begin by picking a unit at random from each urn. The unobserved unit selected from the second urn is assigned the value of the observed unit selected from the first urn and then they are

both placed in the first urn. This step is repeated until the second urn is empty and all the unobserved units have been assigned a value that appeared in the sample. In each succeeding step all the units in the first urn are treated identically. At each stage both the units that were in the sample and those that were not in the sample but have received an imputed value have the same probability of being selected from the first urn. Once they have all been assigned a value we have observed one realization of the completed population from the ‘Polya posterior’. Hence by simple Polya sampling we have a predictive distribution for the unseen given the seen. A good reference for Polya sampling is Feller (1968).

It is easy to verify that under this predicted distribution the expected value of the population mean is just the sample mean. With a bit more work one finds that its posterior variance is

$$\left(1 - \frac{n}{N}\right) \frac{v_s}{n} \frac{n-1}{n+1} \tag{1}$$

where v_s is the sample variance. Note that this is approximately the frequentist variance of the sample mean under simple random sampling when the sample size is at least 25. They differ only by the additional factor $(n-1)/(n+1)$ in equation 1. Hence inference for the population mean under the Polya posterior agrees with standard methods. Note the design probabilities play no formal role in the inference based on the Polya posterior. But for it to be appropriate, in the judgment of the survey sampler, the values for the characteristic of interest for the observed and unobserved units need to be roughly exchangeable. This is usually the case when simple random sampling is used to select the sample.

Our goal here is to show how we can generalize the Polya posterior by building Bayesian models that allows one to incorporate design type information into the definition of the prior.

3 Defining a prior

Before discussing our method for defining prior distributions we need to introduce some notation. We need to warn the reader that our notation is not completely standard.

The the number of units in population is N . We assume that the sampling design only assigns positive probability to samples of size n .

$\Lambda = \{\alpha, \beta, \dots, \tau\}$ is a set of N labels which identify the units in the population. We let λ denoting a typical label.

$y = \{y_\lambda : \lambda \in \Lambda\}$ denotes a typical unknown set of population values belonging to the parameter space \mathcal{Y} , where for each λ we have $y_\lambda > 0$. We will specify \mathcal{Y} later.

Λ is an unordered set so the population labels have no order. However order will be important for us. This is achieved by letting $u = (\alpha, \beta, \dots, \tau)$ denote the labels in some fixed order. Then y_u denotes y arranged in this order. So when we write y 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. $y_{\pi(u)}$ denotes the unknown values of y arranged in the order determined by $\pi(u)$. For simplicity we write y_π instead of $y_{\pi(u)}$. We let π_n denote the first n values of π and π_{n+} its remaining $N - n$ values. Similarly we let y_{π_n} be the first n values (in order) of y_π and $y_{\pi_{n+}}$ be the remaining $N - n$ values (in order). Finally we let Π be the set of all $N!$ possible permutations π .

Suppose the design is simple random sampling without replacement (srs). We assume that the units are selected one at a time from the population until our sample of size n is drawn. At this stage we can imagine continuing the sampling procedure until all the units from the population have been selected. This is just a thought experiment and is not something that will be implemented. However we see that the 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 .

Other ordered designs can be given the same treatment. For example if there is a real valued auxiliary variable and at each stage a unit is selected at random with probability proportional to the size of its auxiliary variable (pps) then this also defines a probability distribution on Π . In the following we will assume that a probability distribution has been defined on Π . For now we assume that the probability distribution over Π comes from the ordered design used to select the sample. But as we shall see they do not have to be the same. We are also assuming that the ordered design does not depend on the values of y .

Our key idea is that we can use this probability distribution on Π when defining a prior distribution for y . This defines the marginal distribution for π and then given a order, π , we define the conditional distribution of y_π

Formally the marginal prior distribution for y is

$$\begin{aligned} p(y) &= \sum_{\pi} p(\pi) p(y | \pi) \\ &= \sum_{\pi} p(\pi) p(y_{\pi}) \end{aligned} \tag{2}$$

As far as we know such priors have not been considered in the literature. We will see that this has some interesting consequences.

In the following sometimes order will matter as in the previous equations and at other times it will not. Our notation will try to make this clear.

Let

$$s = (\lambda_1, \lambda_2, \dots, \lambda_n)$$

denote the labels of a subset of size n in order. Let

$$y_s = (y_{\lambda_1}, y_{\lambda_2}, \dots, y_{\lambda_n})$$

denote the corresponding y values of the variable of interest. Again order matters. We let $\{s\}$ denote the unordered version of s and $\{y_s\}$ its unordered set of y values. More generally $\{\}$ applied to an ordered set means that order is no longer taken into account.

Even though we are using an ordered design to define a prior distribution we follow standard practice and assume that the observed data is $(\{s\}, \{y_s\})$.

Let y be consistent with the observed sample $(\{s\}, \{y_s\})$. This means that $y = y_{\pi}$ for some π where

$$\{\pi_n\} = \{s\} \quad \text{and} \quad \{y_{\pi_n}\} = \{y_s\}$$

Our first goal is to find an expression for the probability of the “unseen” given the “seen”. For a fixed observed sample, $(\{s\}, \{y_s\})$, and a y consistent with the sample we need to compute

$$p(y(\{s'\}) | \{s\}, \{y_s\}) = \frac{p(y(\{s'\}) \{s\}, \{y_s\})}{p(\{s\}, \{y_s\})} \tag{3}$$

where $\{s'\}$ is the unordered complement of s . So $y(\{s'\})$ are the unordered y values for the units not in the sample. Similarly $y(\{s\})$ are the unordered y values for the units in the sample.

In the models we will consider $p(\pi)$ and $p(y | \pi)$ will have special forms which will make equation 3 easier to evaluate.

For srs and pps it is true that

$$p(\pi) = p(\pi_n) p(\pi_{n+} | \{\pi_n\}) \quad (4)$$

So for the rest of the manuscript we will assume that our marginal prior distribution on Π satisfies this equation for our fixed sample size n and each choice of π .

For each π and the fixed sample size n we will assume that

$$p(y | \pi) = p(y(\{\pi_n\}) | \pi_n) p(y(\{\pi_{n+}\}) | \{y_{\pi_n}\}, \pi_{n+}) \quad (5)$$

This assumption breaks $p(y | \pi)$ into two factors. The first factor states that given π the distribution of the first n units depends only on π_n and not on anything that comes later. The second factor states that the distribution of the last $N - n$ units can depend on the first n units only through their unordered values. This may seem like a strong assumption but we will see that it is satisfied for some models which are useful in survey sampling.

Under the assumptions in equations 4 and 5 we will see that equation 3 has a particularly nice form. First note that

$$\begin{aligned} p(\{s\}, \{y_s\}) &= \sum_{\pi: \{\pi_n\}=\{s\}} p(\pi_n) \left\{ \sum_{y(\{s\}): \{y(\{s\})\}=\{y_s\}} p(y(\{s\}) | \pi_n) \right\} \\ &= \sum_{\pi: \{\pi_n\}=\{s\}} p(\pi_n) p(\{y_s\} | \pi_n) \end{aligned}$$

In a similar fashion we see that

$$\begin{aligned} p(y(\{s'\}), \{s\}, \{y_s\}) &= \\ \sum_{\pi: \{\pi_n\}=\{s\}} p(\pi_n) p(\{y_s\} | \pi_n) p(\pi_{n+} | \{s\}) p(y(\{s'\}) | \{y_s\}, \pi_{n+}) \end{aligned}$$

Also, if π is such that $\{\pi_n\} = \{s\}$ then

$$p(\pi_n | \{s\}, \{y_s\}) = \frac{p(\pi_n) p(\{y_s\} | \pi_n)}{\sum_{\tilde{\pi}: \{\tilde{\pi}_n\}=\{s\}} p(\tilde{\pi}_n) p(\{y_s\} | \tilde{\pi}_n)}$$

Using the three previous equations and equation 3 we see that

$$\begin{aligned}
& p(y(\{s'\}) \mid \{s\}, \{y_s\}) \\
= & \sum_{\pi: \{\pi_n\}=\{s\}} p(\pi_n \mid \{s\}, \{y_s\}) p(\pi_{n+} \mid \{s\}) p(y(\{s'\}) \mid \{y_s\}, \pi_{n+}) \\
& = \left\{ \sum_{\pi: \{\pi_n\}=\{s\}} p(\pi_n \mid \{s\}, \{y_s\}) \right\} \times \\
& \left\{ \sum_{\pi_{n+}: \{\pi_{n+}\}=\{s'\}} p(\pi_{n+} \mid \{s\}) p(y(\{s'\}) \mid \{y_s\}, \pi_{n+}) \right\} \\
= & \sum_{\pi_{n+}: \{\pi_{n+}\}=\{s'\}} p(\pi_{n+} \mid \{s\}) p(y(\{s'\}) \mid \{y_s\}, \pi_{n+})
\end{aligned}$$

Hence under our assumptions on $p(\pi)$ and $p(y \mid \pi)$ we have shown that our posterior is of the form

$$p(y(\{s'\}) \mid \{s\}, \{y_s\}) = \sum_{\pi_{n+}: \{\pi_{n+}\}=\{s'\}} p(\pi_{n+} \mid \{s\}) p(y(\{s'\}) \mid \{y_s\}, \pi_{n+}) \quad (6)$$

This has the same form as the prior in equation 2 with each factor updated in a sensible manner. We will see that it can be be easy to work with. When estimating the population total we will find explicit expressions for the corresponding point estimate along with its posterior variance in several models of interest. When no closed form expressions are available they can often be found through simulation. Generating simulated copies of $y(\{s'\})$ from this distribution can be done in two steps. First we simulate a set of values for π_{n+} from the first factor. Then we plug these values into the second factor and use it to get a set of simulated values for the unseen members of the population.

It seems natural that $p(\pi)$, our prior distribution over Π , should be the sampling design which is actually selecting the units which appear in the sample. Although this is a useful way to think about it, turns out that there is no theoretical reason why they should always be the same. In fact we shall see that it is often convenient to let them be different. We should also point out that except in section 7.3 all the sampling designs we are considering never depend on the unknown values of y .

In the next section we will consider some specific examples of the general scheme described here.

4 Some Two Urn Models

We will now introduce some Bayesian models which relate the unseen to the seen in finite population sampling. This will be done conditionally upon a fixed sample. In section 5 we will show how the conditional distributions defined here are related to the family of priors we defined in section 3.

In this section we assume that associated with each unit is a positive auxiliary variable which is positively correlated with its y value. We let x

$$x = \{x_\lambda : \lambda \in \Lambda\}$$

denote this set of values which is assumed to be known. For definiteness assume the sampling design is srs.

Suppose now we have observed our sample $(\{s\}, \{y_s\})$ along with the values $\{x_s\}$. These are the values of the auxiliary variable for the units in the sample. In this section the sample $(\{s\}, \{y_s\})$ is assumed to be fixed.

Using the ideas of the previous sections we now describe a Conditional Two Urn Bayesian model which leads to the ratio estimator. We begin with an informal description which is suggested by our discussion of the Polya posterior. As was the case there we consider two urns. Into the first urn we place the n sampled units. Attached to each unit is their known x value and their observed y value. Into the second urn we place the $N - n$ unsampled units. Attached to each of them is their known x value. To get a simulated copy of the entire population we proceed as follows. We randomly select a unit from the second urn using srs. The next step is to use the sampled units in the first urn to randomly assign a value to this randomly selected unit. To do this we select a unit from the first urn using pps sampling on the x values and compute the ratio of its y and x values, say r . To get the simulated y value for the unit selected from the second urn we take the product of its x value and r . This unit, with its simulated y value and known x value, is then placed in the first urn. Of course, just as before the unit selected from the first urn which was used in constructing this simulated value is also returned to the first urn. For the rest of the steps this unsampled unit with its simulated y value is treated exactly the same as the n original sampled units. This process is repeated until each unobserved unit has been assigned a y value. Thus we have constructed one simulated copy of the entire population given the sample using this predictive distribution for the unseen given the seen. Note if x is the vector of all 1's then this distribution is just the Polya posterior.

We will now prove that under this model, which for every sample $(\{s\}, \{y_s\})$ gives a probability distribution for the unseen given the seen, the resulting Bayes estimator is just the ratio estimator. In fact we will prove a more general result for which this is a special case. To do that we need to introduce

$$w = \{w_\lambda : \lambda \in \Lambda\}$$

a known set of positive weights. So associated with the unit labeled λ we have y_λ (the variable of interest), x_λ (the known value of the auxiliary variable) and $w_\lambda > 0$ a known weight. We can now generalize our Conditional Two Urn Model.

As before we place the n sampled units into the first urn. Attached to each unit is their observed y value, along with their known x value and w value. Into the second urn we place the $N - n$ unsampled units. Attached to each of them is their known x and w values. To get a simulated copy of the entire population we proceed as follows. We randomly select a unit from the second urn using srs. The next step is to use the sampled units in the first urn to assign a value to this randomly selected unit. To do this we select a unit from the first urn using pps sampling on the w values (not the x values as before). Given a selected unit we find the ratio of its y and x values and call it r . To get the simulated y value for the unit selected from the second urn we take the product of its x value and r . This unit, with its simulated y value and known x and w values, is then placed in the first urn. (Of course, just as before the unit selected from the first urn which was used in constructing this simulated value is also returned to the first urn.) For the rest of the steps this unsampled unit with its simulated y value is treated exactly the same as the original sampled units. This process is repeated until each unobserved unit has been assigned a y value. For a fixed sample and known auxiliary variable x and weight function w we will call this the Conditional Two Urn Model and denote it by $CTUM(\{s\}, \{y_s\}, x, w)$.

We first prove that under this model the marginal distribution of y_λ/x_λ is the same for each $\lambda \notin s$.

Lemma 1. *Let $s = (\lambda_1, \lambda_2, \dots, \lambda_n)$ be the labels of a sample of size n and consider the Conditional Two Urn Model, $CTUM(\{s\}, \{y_s\}, x, w)$. For $i = 1, 2, \dots, n$ let $r_i = y_{\lambda_i}/x_{\lambda_i}$. Let w_1, \dots, w_n be the corresponding w values and $sw = \sum_{i=1}^n w_i$. Then for $\lambda \notin s$*

$$p(y_\lambda = x_\lambda r_i) = \frac{w_i}{sw} \quad \text{for } i = 1, 2, \dots, n$$

Proof. Let X be the random variable which denotes the x value of the first unit selected at random from the second Urn which contains the unsampled units. Let R be the random variable which denotes the y/x ratio of the first unit selected from the first urn (which contains the sampled units) using pps defined by w_s . Clearly

$$\begin{aligned} P(X = x_\lambda, R = r_i) &= p(x_\lambda | s)p(r_i | s, w_s) \\ &= \frac{w_i}{(N-n)sw} \quad \text{for } \lambda \notin s \text{ and } i = 1, 2, \dots, n \end{aligned}$$

Now for the second draw from each urn let X^* and R^* be defined analogously to X and R for the first draw. Then for $\lambda \notin s$ and $\gamma \notin s$ we have

$$\begin{aligned} P(X^* = x_\gamma, R^* = r_j | X = x_\lambda, R = r_i) &= \\ &\begin{cases} w_j / ((N-n-1)(sw + w_\lambda)) & \text{if } i \neq j \\ (w_j + w_\lambda) / ((N-n-1)(sw + w_\lambda)) & \text{if } i = j \end{cases} \end{aligned}$$

Letting

$$p(x_\gamma, r_j) = P(X^* = x_\gamma, R^* = r_j)$$

we see that

$$\begin{aligned} p(x_\gamma, r_j) &= \sum_{\lambda \notin s \text{ and } \lambda \neq \gamma} \left\{ \sum_{i=1}^n P(X = x_\lambda, R = r_i, X^* = x_\gamma, R^* = r_j) \right\} \\ &= \frac{1}{(N-n)(N-n-1)ws} \sum_{\lambda \notin s \text{ and } \lambda \neq \gamma} \left\{ \sum_{i \neq j} \frac{w_i w_j}{sw + w_\lambda} \right. \\ &\quad \left. + \frac{w_j(w_j + w_\lambda)}{sw + w_\lambda} \right\} \\ &= \frac{1}{(N-n)(N-n-1)ws} \sum_{\lambda \notin s \text{ and } \lambda \neq \gamma} \frac{w_j(sw + w_\lambda)}{sw + w_\lambda} \\ &= \frac{w_j}{(N-n)sw} \end{aligned}$$

Hence (X^*, R^*) has the same joint distribution as (X, R) .

Now let (X^*, R^*) be the random variables which denote the values which appear on the k th draw where $2 < k \leq N-n$. It follows by induction that

their joint distribution is the same as that of (X, R) . One just conditions on the values of (X, R) and then uses the induction hypothesis and the above calculation. Finally, since under the Conditional Two Urn Model the probability that x_λ appears on any given draw is just $1/(N - n)$ the lemma is proved. \square

We see by the lemma that

$$E(y_\lambda \mid s, y_s) = x_\lambda \sum_{i=1}^n \frac{r_i w_i}{s w} = x_\lambda \sum_{i=1}^n \frac{y_{\lambda_i} w_i}{x_i s w} \quad (7)$$

Since we can write the population total as

$$\sum_{i=1}^n y_{\lambda_i} + \sum_{\lambda \notin \{s\}} y_\lambda \quad (8)$$

we find under squared error loss that the corresponding Bayes estimator is

$$\sum_{i=1}^n y_{\lambda_i} + \left(\sum_{\lambda \notin \{s\}} x_\lambda \right) \sum_{i=1}^n \frac{y_{\lambda_i} w_i}{x_i s w} \quad (9)$$

If we take the weight function w to be x then the estimator in equation 9 is the ratio estimator. For another example suppose x is such that $0 < x_\lambda < 1$ for all $\lambda \in \Lambda$ and their sum is n . If we let $w_\lambda = 1 - x_\lambda$ then the estimator in equation 9 is the Horvitz-Thompson estimator. In this case we can think of x as the inclusion probabilities of the design. Note that the actual design leading to s plays no role in our Conditional Two Urn Model. But given the sample the Conditional Two Urn Model posterior for the unseen given the seen is making use of information contained in x .

5 Admissibility

So far we have presented the Conditional Two Urn Model as just “a thing to do”. In the discussion we have talked about posterior distributions without specifying a prior distribution. That is because there is no known single prior distribution which leads to the posteriors in the Conditional Two Urn Model. In section 3.2 of Ghosh and Meeden (1997) we discussed the family

of estimators given in equation 9. For the situation when the sample was unordered and $n = 2$ we showed that estimators in this family were all stepwise Bayesian and admissible. However we could not prove admissibility when $n > 2$. We will now show that the estimator in equation 9 is stepwise Bayes and hence admissible.

In stepwise Bayes arguments a finite sequence of disjoint subsets of the parameter space is selected, where the order of the subsets is important. A different prior distribution is defined on each of the subsets. First, the Bayes procedure is found for each sample point that receives positive probability under the prior defined on the first subset. Next the Bayes procedure is found for each sample point which receives positive probability under the prior defined on the second subset and which was not considered in the first step. Then for the prior defined on the third subset the Bayes estimate is found for all sample points which are assigned positive probability and which were not considered in the first two steps. This process is continued until all possible sample points have been considered. For a particular sample point the value of its stepwise Bayes estimate comes from the prior associated with the step where it was first assigned positive probability. It is the stepwise Bayes nature of the posteriors in the Two Urn Model that explains their somewhat paradoxical nature. Given a sample each behaves just like a proper Bayesian posterior but the collection of possible posteriors that arise from all possible samples comes from a family of priors and not a single prior.

Before stating the theorem we need a bit more notation. Let q be a sampling design and $UCTUM_n(q, x, w)$ denote the Unconditional Two Urn Model. In this setup the design q is used to select the sample $\{s\}$ of size n . Given the sample $CTUM(\{s\}, \{y_s\}, x, w)$ is used to form an estimate of the population total.

Let $b = (b_1, b_2, \dots, b_k)$ be a vector of known positive real numbers. For a given auxiliary variable x we define a parameter space for y by

$$\mathcal{Y}_x(b) = \{y : \text{such that for } \lambda \in \Lambda \text{ there exists} \\ \text{some } j = 1, 2, \dots, k \text{ for which } y_\lambda/x_\lambda = b_j\}$$

Theorem 1. *Let x be a positive auxiliary variable, w a positive weight function and q an ordered design. Consider the problem of estimating the population total with squared error as the loss function. Then for any positive vector b the estimator arising from the Unconditional Two Urn Model, $UCTUM_n(q, x, w)$, is stepwise Bayes and hence admissible when the parameter space for y is $\mathcal{Y}(b)$.*

Proof. There are many stages in the proof and except for the first stage each have many similar steps. For each step within each stage we need to select a subset of $\mathcal{Y}(b)$ and define a prior distribution over the subset of the form

$$p(y) = p(\pi)p(y | \pi)$$

At every step in each stage the distribution $p(\pi)$ is just the uniform distribution over the set of all possible permutations on a set of N objects.

In the first stage $p(y | \pi)$ puts mass $1/k$ on each of the k points of $\mathcal{Y}(b)$ where the ratios $y_\lambda/x_\lambda = b_j$ for all $\lambda \in \Lambda$. It is easy to check that the resulting Bayes estimator agrees with equation 9 in this trivial case. This takes care of the samples where all the observed ratios are identical.

In the next stage we take care of all samples where the observed ratios in the sample take on just two values. This is done in $\binom{k}{2}$ steps by considering samples where just (b_1, b_2) appear, then samples where just (b_1, b_3) appear and so on until the final step which considers samples where just (b_{k-1}, b_k) appear. In the next stage we take care of all samples where the observed ratios in the sample take on just three values. We continue to the last stage where the samples take on $\min\{k, n\}$ different values.

We now consider some specific step in some specific stage. Let b^* be a subset of b of size $1 < m \leq \min\{k, n\}$. Without loss of generality we can assume that it is the first m elements of b . Let $\mathcal{Y}^*(b^*)$ be the subset of $\mathcal{Y}(b)$ where each of the values of b^* appear at least once. At this stage we only consider samples where the observed ratios in the sample consists of all the values in b^* . Samples where only some of these ratios appear have been considered in an earlier stage. Since we have already defined $p(\pi)$ it only remains to define $p(y | \pi)$ for $y \in \mathcal{Y}(b^*)$.

We see from equation 6 that

$$p(y(\{s'\}) | \{s\}, \{y_s\}, \pi_{n+}) \tag{10}$$

is what remains to be specified since $p(\{s\}, \{y_s\})$ plays no role in the calculation of our estimate once $(\{s\}, \{y_s\})$ has been observed. Any choice of $p(\pi_n)p(y(\{\pi_n\}) | \pi_n)$ which is consistent with the rest of our definition is fine. But for the distribution in equation 10 we take the distribution given by $CTUM(s, y_s, x, w)$. This is a well defined distribution and as we have seen yields the estimator in equation 9 \square

The proof of this theorem is quite similar to the stepwise Bayes proof of the admissibility of the usual estimator $(N/n) \sum_{i=1}^n y_{\lambda_i}$ of the population

total. Using this theorem it is straight forward to show that the estimator given in equation 9 is admissible when the parameter space is $(0, \infty)^N$. See Ghosh and Meeden (1997) for more details about these proofs.

Joshi (see Joshi(1966)) proved the admissibility the ratio estimator, the Horvitz-Thompson estimator and the regression estimator when the parameter space was N -dimensional Euclidian Space.

Note the sampling design q plays no direct role in the proof. Scott (see Scott (1975)) showed quite generally that the admissibility of an estimator does not really depend on the sampling design. However we have shown here that using an ordered design in the definition of a prior is a way to incorporate into a Bayesian analysis information which is typically included in a design.

Actually the Two Urn Model considered in the theorem is just a rather special case of such models. It follows from equation 6 and a careful reading of the proof that any specification of

$$p(\pi_{n+} | \{s\}) \quad \text{and} \quad p(y(\{y'\}) | \{y_s\}, \pi_{n+})$$

for every $(\{s\}, \{y_s\})$ will yield an admissible estimator. This is a reflection that admissibility is a rather weak property. However this flexibility does allow for the easy incorporation of various kinds of information into a Bayesian analysis.

So far we have assumed that the sampling from the second urn which contains the unsampled units was just simple random sampling without replacement. Clearly other sampling plans can be used as well. Also one can change the way a unit is selected from the first urn at each stage to randomly assign a y value to the unsampled unit selected from the second urn. However in many such models we are unable to find explicit forms for the resulting estimator and its variance. In practice this may not be a problem because often, given a sample, it is easy to simulate completed copies of the population using equation 6. In such cases one can find approximately the resulting estimator of many population quantities of interest. One can also find approximately the corresponding 0.95 Bayesian credible interval whose frequentist coverage will often be approximately 95%.

6 The posterior variance

In most situations in addition to a point estimate one wants some measure of its precision. For frequentists this is usually an estimate of variance which can be used to construct confidence intervals. For a Bayesian it is the posterior variance. In this section we will find expressions for the posterior variance for the models in section 4

We begin with a lemma which is a companion to lemma 1 It is easy to extend the argument in the proof of that lemma to see that for any $\lambda \notin s$ and $\gamma \notin s$ the joint distribution of $(y_\lambda/x_\lambda, y_\gamma/x_\gamma)$ does not depend on λ and γ . Using this fact we can find the posterior variance of the estimator in equation 9. The key step is proved in the next lemma.

Lemma 2. *Let $s = (\lambda_1, \lambda_2, \dots, \lambda_n)$ be the labels of a sample of size n and consider the Conditional Two Urn Model, $CTUM(\{s\}, \{y_s\}, x, w)$. For $i = 1, 2, \dots, n$ let $r_i = y_{\lambda_i}/x_{\lambda_i}$. Let w_1, \dots, w_n be the corresponding w values and $sw = \sum_{i=1}^n w_i$. Suppose $\lambda \notin s$ and $\gamma \notin s$ then the expected value of the product of their y values*

$$E(y_\lambda y_\gamma) = \frac{x_\lambda x_\gamma}{2} \sum_{i=1}^n (E_{i,\lambda}(R) + E_{i,\gamma}(R)) \frac{r_i w_i}{sw}$$

where

$$E_{i,\lambda}(R) = \frac{1}{sw + w_\lambda} (r_i(w_i + w_\lambda) + \sum_{j \neq i} r_j w_j)$$

and $E_{i,\gamma}(R)$ is defined similarly.

Proof. Suppose that the unit labeled λ appears on draw t_1 from the second urn and the unit labeled γ appears on draw t_2 . Then we see that

$$\begin{aligned} E(y_\lambda y_\gamma \mid t_1 < t_2) &= x_\lambda x_\gamma \sum_{i=1}^n \frac{r_i w_i}{sw} \left[\frac{r_i(w_i + w_\lambda)}{sw + w_\lambda} + \sum_{j \neq i} \frac{r_j w_j}{sw + w_\lambda} \right] \\ &= x_\lambda x_\gamma \sum_{i=1}^n \frac{r_i w_i}{sw} E_{i,\lambda}(R) \end{aligned}$$

Next we condition on the event $t_1 > t_2$ and the result follows. \square

This makes it easy to write an expression for the posterior variance of the estimator given in equation 9. Using lemma 1 we can find the variance of any $y_\lambda \notin s$ and using lemma 2 we can find the covariance of any pair.

In many situations, given the sample, frequentists talk about the weight assigned to each unit in the sample. These weights reflect how much of the population each of the units should represent. Under srs all the units in the sample are given the same weight, one. We can think of the vector w in the Two Urn Model as playing a similar role. To study this further we will consider now the special case where the auxiliary x is identically equal to one.

Given the sample $\{w_s\}$ are the weights associated with the observed y values $\{y_s\}$. We are assuming that we wish to treat all the unsampled units in an exchangeable manner so we will set all their weights equal to sw/n . This will keep the scale between the unseen and seen approximately correct. But to allow for a final adjustment we will select a real number c close to one and then multiply each member of $\{w_s\}$ by c . We will denote this adjusted collection by w^c and its values are given by

$$w_\lambda^c = \begin{cases} cw_\lambda & \text{if } \lambda \in \{s\} \\ sw/n & \text{if } \lambda \notin \{s\} \end{cases} \quad (11)$$

Note the point estimate of the population total under w and w^c must be the same, just their posterior variances will be different.

For w^c it is easy to see that for $\lambda \notin s$

$$E_{i,\lambda}(R) = \frac{n}{(nc+1)sw} \left\{ c \sum_{j=1}^n w_j y_j + y_i \frac{sw}{n} \right\}$$

So for $\lambda \notin s$ and $\gamma \notin s$

$$E(y_\lambda y_\gamma) = \frac{n}{(nc+1)sw} \left\{ \frac{c}{sw} \left(\sum_{j=1}^n w_j y_j \right)^2 + \frac{1}{n} \sum_{j=1}^n w_j y_j^2 \right\}$$

Next some easy algebra gives

$$\begin{aligned} cov(y_\lambda y_\gamma) &= \frac{1}{nc+1} \left\{ \frac{\sum_{j=1}^n w_j y_j^2}{sw} - \left\{ \frac{\sum_{j=1}^n w_j y_j}{sw} \right\}^2 \right\} \\ &= \frac{1}{nc+1} V(y_\lambda) \end{aligned}$$

A final bit of algebra and the usual formula for the variance of a sum yields the expression

$$V\left(\sum_{\lambda \notin s} y_{\lambda} \mid \{s\}, \{y_s\}\right) = \frac{N - (1 - c)n}{nc + 1} (N - n)V(y_{\lambda}) \quad (12)$$

In the special case where w is identically one and $c = 1$ then w^c is identically one and our Two Urn Model is just the Polya posterior and the posterior variance given above agrees with the Polya posterior variance. One of the authors (see Meeden (1999)) recently considered another somewhat different generalization of the Polya posterior which he called the weighted Polya posterior. There the weights were assigned to the different y values that appeared in the sample not to the units themselves. When all the values in the sample are distinct then equation 12 agrees with one found in Meeden (1999). Although there is some formal overlap between the two approaches we believe that assigning weights to units that appear in the sample is much closer to survey practice.

7 Some applications

Here we will present some applications of our approach. As we saw in the discussion following equation 6 given the sample $(\{s\}, \{y_s\})$ it will often be possible to simulate from the posterior for the Two Urn Model. This allows one to find approximately point estimates and Bayesian credible intervals for population quantities of interest. In the following we will concentrate on estimating the population total.

7.1 The Horvitz-Thompson estimator

The Horvitz-Thompson estimator has been much discussed in the literature and is a classic example of unequal probability sampling. However expect for small sample sizes it is often difficult to construct designs with specified inclusion probabilities. Moreover finding a sensible estimate of its variance can also be a problem. In practice it is often recommended to compute the estimate of the variance under the assumption that the sampling was done with replacement even though sampling without replacement was used. It is well known that this will tend to overestimate the variance and result in con-

servative confidence intervals. Using the results of section 6 we will compare our Two Urn Model estimate of variance to the recommended estimate.

We constructed three populations as follows. We began by taking a random sample of size 200 from a gamma distribution with shape parameter 5 and scale parameter 1. To get x we added 15 to each member of the sample. In the first population the distribution of y_λ given x_λ was normal with mean $2x_\lambda$ and standard deviation 2.5. The correlation between the two was 0.89. In the second population the distribution of y_λ given x_λ was normal with mean $2x_\lambda$ and standard deviation 4.5. The correlation between the two was 0.78. In the third population the distribution of y_λ given x_λ was normal with mean $3x_\lambda$ and standard deviation $3\sqrt{x_\lambda}$. The correlation between the two was 0.53. Note these are populations where we would expect the Horvitz-Thompson estimator to work well. Since our sample size $n = 30$ we rescaled x so that it sums to 30. Recall to get the Horvitz-Thompson estimator $w_\lambda = 1 - x_\lambda$ for $\lambda \in \Lambda$.

We took 1,000 samples of size 30 from each population. We used pps sampling (proportion to x) without replacement. For each sample we computed the estimated variance and the resulting confidence interval using the standard conservative method. We did the same for the Two Urn Model using the w^c of equation 12 with $c = (n - 4)/n$. We also found the 0.95 credible interval approximately by simulating completed copies of the population using equation 6. This gave very similar results to using the posterior variance and they are omitted. The results are given in table 1

The Two Urn Model estimate of variance seems to be a slight improvement over the the conventional conservative estimate. Simulations for other sample sizes and different populations yield similar results. The choice of $c = (n - 4)/n$ seems to work well in these examples.

7.2 Smooth populations

For some populations, units with similar labels tend to be more alike than units whose labels are quite different. One situation where this happens is when x and y are highly correlated. We now consider a Two Urn Model which should be useful for such problems.

Given the sample we select an unsampled unit from the second urn using srs. Suppose this unit has label λ . We then select a unit at random from the

Table 1: A Comparison of the usual conservative 95% confidence interval and the approximate Two Urn Model 0.95 credible interval for the Horvitz-Thompson estimator when estimating the population mean. The results are based on 1000 samples of size 30.

Pop	Est	Ave len	Freq of coverage
1	HT	394	0.97
	TUM	374	0.96
2	HT	651	0.96
	TUM	616	0.96
3	HT	1834	0.95
	TUM	1736	0.94

first urn containing the sampled units using pps with weights

$$\exp -a|x_{\lambda_i} - x_{\lambda}| \quad \text{for } i = 1, 2, \dots, n \quad (13)$$

where $a > 0$ is specified by the statistician. The y value of this unit is assigned to the unit with label λ . This unit along with its assigned y value and know x is place in the first urn. The unit selected at random from the first urn is also returned to the first urn. This process is continued until all the unobserved units have been move to the first urn and have been assigned a y value. Clearly this defines a probability distribution for the “unseen” given the “seen” which makes use of the idea that we expect units whose x values are close tend to have similar y values. Note that this model is somewhat different than the previous Two Urn Models because at each step the draws from the two urns are no longer independent. But as we have noted this causes no problems in the proof of the Theorem 1 so the model based on the specification of equation 13 has a stepwise Bayes justification.

To see how the resulting estimator could work in practice we construct another population. We began by taking a random sample of size 200 from the uniform(10,14) distribution. Next we subtracted 12 from each of them and cubed the differences. Finally we added the same constant to each so that the minimum value of the set was 5. The resulting set was our set of x values. Then the distribution of y_{λ} given x_{λ} was normal with mean x_{λ} and

standard deviation 0.8. The correlation between x and y was 0.88 and the true total of y was 2635.

For this population a frequentist could either stratify the population or use the regression estimator or perhaps do both. When using the regression estimator with stratification it is always a question of whether you should use one regression estimator for the entire population or use different regression estimators within the strata. Here we did both and denoted the two estimators by `reg` and `regstr`. Another frequentist estimator we considered was the method of collapsed strata. One might think that this could be a competitor to the Two Urn Model procedure because in some sense both are more local than the others. This estimator was denoted by `colspstrat` while `strat` denoted the usual estimator based on the stratification. Finally, `tum` denotes the Two Urn Model estimator with $a = 12$. We constructed 1000 simulated copies of the population when computing `tum`.

We considered two different sampling designs. Using the ordered x values for the first design we constructed 4 strata of size 50 and selected 10 units from each strata using `srs`. For the second design we constructed 10 strata of size 20 and selected 4 units from each strata using `srs`. In each case we took 500 random samples. The results are given in table 2.

The Two Urn Model procedure is the clear winner. As a point estimator it is beaten in only one case and then by just a bit and its interval estimates are superior across the board. However an important questions remains how was the choice of $a = 12$ determined.

We see from equation 13 that as the value of a increases in our Two Urn Model the resulting simulated populations have less variation. In particular we see that for a close to zero the Two Urn Model will behave much like the Polya posterior which yields the standard estimator. To see how sensitive the model is to the choice of a we generated another population using the same procedure that generated the one use for table 2. In this case the correlation between x and y was 0.89 and the population total was 2573. In this case we took 500 `srs` of size 50 and compute the usual estimator and the regression estimator. We also compute the Two Urn Model estimator for five choices of a . The results are given in table 3

We see for a large range of values for a the behavior of the Two Urn Model point estimators do not vary much. For a smaller range of a their 0.95 credible sets will yield approximate 95% confidence sets for the population total. For many problems one should be able to choose a reasonable value of a . This Two Urn Model is using more information than the regression estimator since

Table 2: A comparison of a Two Urn Model procedure with standard frequentist procedures for two different stratified designs. The results are based on 500 samples.

Est	Ave value	Ave aberr	Ave lowbd	Ave len	Freq of coverage
4 strata of size 50					
strat	2637	36.6	2550	174	0.93
colspstr	2637	36.6	2579	115	0.78
reg	2635	28.7	2559	153	0.97
regstr	2637	21.0	2593	87	0.89
tum $a = 12$	2637	21.4	2585	104	0.95
10 strata of size 20					
strat	2636	21.5	2580	111	0.94
colspstr	2636	21.5	2581	109	0.95
reg	2636	19.6	2558	156	1.00
regstr	2635	24.6	2599	72	0.74
tum $a = 12$	2636	17.9	2586	100	0.96

it assumes that the values of x are known for the population rather than just the mean. In some sense it is a kind of post stratification. Also note that its posterior variance does not depend on any model assumptions about how the y and x values are related. Instead this comes from our Two Urn Model posterior which just uses the assumption that units with x values which are close should tend to have similar y values.

We see that as the value of a increases in our Two Urn Model based on equation 13 the resulting simulated populations have less variation. This is clear from equation 13. In particular we see that for a close to zero the Two Urn Model will behave much like the Polya posterior which yields the standard estimator. But for a large range of values for a the resulting Two Urn Model estimators will do significantly better than the regression estimator. For a smaller range of a their 0.95 credible sets will yield approximate 95% confidence sets for the population total which are significantly shorter than the regression interval. For many problems one should be able to choose a reasonable value of a . This Two Urn Model is using more information than the regression estimator since it assumes that x is known for the population

Table 3: A comparison of a Two Urn Model procedure for five different choices of a with standard frequentist procedures. The results are based on 500 random samples of size 50.

Est	Ave value	Ave aberr	Ave lowbd	Ave len	Freq of coverage
usual	2578	67.8	2415	326	0.94
reg	2574	30.8	2498	152	0.94
Two Urn Model of equation 13					
$a = 4$	2576	25.0	2495	164	0.99
$a = 8$	2575	17.3	2518	114	0.98
$a = 12$	2573	17.4	2527	93	0.96
$a = 16$	2573	18.0	2532	82	0.92
$a = 20$	2572	17.4	2534	76	0.91

rather than just the mean. In some sense it is a kind of post stratification.

7.3 Designs depending on y

In certain situations, like observational studies, the design generating the units in the sample may not be under the control of the statistician. In such cases it might be reasonable to assume that the probability a unit appears in the sample depends on its y value. Andreatta and Kaufman (see Andreatta and Kaufman (1986)) considered a problem in geology where the probability of discovering an oil field is proportional to the size of the field.

Here we assume that the sample is drawn using pps sampling with weights proportional to $\phi(y)$ where ϕ is a strictly increasing positive function. For now we will assume that ϕ is known but in practice it will only be known approximately. Given such a sample we wish to estimate the population total.

Suppose we are given the sample $(\{s\}, \{y_s\})$ how should we select a Two Urn Model that makes use of the knowledge of ϕ ? Since we are assuming $x \equiv 1$ we must make a sensible choice of w . Making $w \equiv 1$ cannot be right since this will just give us the usual estimator which we know will be biased upwards. Intuitively we want to weight the units in the sample with large

y values less than those with the smaller values. This suggest the following sample dependent weight function w^c .

$$w_\lambda^c = \begin{cases} c(1/\phi(y_\lambda)) & \text{if } \lambda \in \{s\} \\ (\sum_{i=1}^n (1/\phi(y_{\lambda_i}))/n & \text{if } \lambda \notin \{s\} \end{cases} \quad (14)$$

Ignoring the factor c we see that the first line of equation 14 essentially weights a sample value inversely proportional to its probability of being in the sample and the second line gives an unsampled unit the average weight of the sampled units.

With this choice of w^c we see from equation 9 that our estimator will be

$$\sum_{i=1}^n y_{\lambda_i} + (N - n) \left\{ \left(\sum_{i=1}^n y_{\lambda_i} / \phi(y_{\lambda_i}) \right) / \left(\sum_{i=1}^n 1 / \phi(y_{\lambda_i}) \right) \right\} \quad (15)$$

Note that the estimate is the same for all choices of $c > 0$. It is the posterior variance which depends on c . We will want to select c so that the Bayesian credible intervals will have good frequentist properties.

Note this estimator can be written as

$$\sum_{i=1}^n y_{\lambda_i} + (N - n)(E(y/\phi(y))/E(1/\phi(y)))$$

where by a slight abuse of notation these expected values are with respect to the uniform distribution on the set $\{y_s\}$. Since $1/\phi$ is strictly decreasing the covariance of $\phi(y)$ and y is negative. This implies that the second factor in the second term of the previous equation is less than $E(y) = \sum_{i=1}^n y_{\lambda_i}/n = \bar{y}_s$. This in turn implies that the estimator in equation 15 is less than $N\bar{y}_s$. In other words this estimator should have less bias then the usual estimator.

To see how this estimator works in practice we constructed a population by taking a random sample of size 300 from a log-normal distribution with mean and standard deviation of the log 5 and 0.6 respectively. Our estimator will always be computed under the assumption that the sample is drawn using pps sampling with probabilities proportional to $\phi(y) = y$. Sometimes this will be the case. But we will consider two other situations where the selection probabilities are proportional to $y^{0.8}$ and $y^{1.2}$ respectively. For each of the three scenarios we will consider two different choices of w^c in equation 14. They are $c = 0.75$ and $c = 0.5$. For the six different situations we took a 1000 samples of size 30 and found the Two Urn Model estimates. The results are

given in table 4. Note we are assuming that the population size $N = 300$ is known.

We see from the first line of the table that when we assume $\phi(y) = y$ and that is the actual sampling design then our method works very well. It is almost unbiased and its 0.95 credible interval contains the true population total 95.4% of the time. This is in contrast to the usual estimator which in this case had an average absolute error of 16,546. The average length of its 95% confidence interval was 23,525 and it contained the true population total only 18.8% of the time. In the second line we see that setting $c = 0.5$ results in too conservative of an interval.

In the next four lines of table 4 we see that the behavior of our procedure degrades somewhat when our choice of the sampling design is incorrect. Assuming the sampling design is proportional to y when in fact it is proportional to $y^{1.2}$ is a worse mistake than assuming it is proportional to y when in fact it is proportional to $y^{0.8}$. Selecting a smaller c cannot correct for the bias of the point estimator under misspecification of the sampling design but it can improve the frequency of coverage of the interval estimator. For moderate sample sizes, say from 20 to 60 choosing a c between 0.5 and 0.75 will usual give sensible results as long as the sampling design is not too misspecified. For distributions less skewed than the log-normal, e.g. the gamma distribution, slightly large values of c can be used. However the most important choice is that of the function ϕ . If that is poorly specified then our procedure will preform poorly.

When estimating the population total our point and interval estimators are very easy to compute since they just use equations 15 and 12. This is in contrast to other methods which can be quite computationally involved. If one is interested in estimating other quantities, say a population quantile, then it is easy to simulate completed copies of the population under our Two Urn Model and use them to compute the corresponding Bayes estimators.

8 Discussion

Here we have proposed a new Bayesian way of incorporating prior information into a finite population sampling problem. It allows one to define Two Urn Models which use prior information and the information in the sample to relate the “seen” to the “unseen”. These models are very flexible and allow one to incorporate design type information into a Bayesian analysis.

Table 4: A comparison of a Two Urn Model procedure when the design depends on y . The population total is 49,479. The results are based on 1000 random samples of size 30.

design	value of c	Ave value	Ave aberr	Ave lowbd	Ave len	Freq of coverage
y	0.75	50,925	4,408	40,129	21,593	0.954
y	0.50	50,923	4,336	38,108	25,631	0.983
$y^{0.8}$	0.75	47,718	4,502	37,453	20,531	0.919
$y^{0.8}$	0.50	47,667	4,517	35,227	24,880	0.969
$y^{1.2}$	0.75	53,952	5,735	42,737	22,429	0.892
$y^{1.2}$	0.50	53,980	5,837	40,513	26,934	0.943

Because these are stepwise Bayes procedures, not proper Bayes procedures, they avoid the difficult problem of specifying a proper prior. Given a sample it is often easy to simulate completed copies of the population from the resulting posteriors. This is a real strength of the Bayesian approach since using these simulated copies one can find approximately point and interval estimates of a variety of population parameters. Since these models can sensibly express available prior information they will often have good frequentist properties.

References

- [1] G. Andreatta and G. M. Kaufman. Estimation of finite population properties when sampling is without replacement and proportional to magnitude. *Journal of the American Statistical Association*, 81:657–666, 1986.
- [2] J. K. Ghosh. *Statistical Information and Likelihood: A Collection of Critical Essays by D. Basu*. Springer-Verlag, New York, 1988.
- [3] M. Ghosh and G. Meeden. *Bayesian Methods for Finite Population Sampling*. Chapman and Hall, London, 1997.
- [4] V. P. Godambe. A unified theory of sampling from finite populations. *Journal of the Royal Statistical Society, Series B*, 17:269–278, 1955.
- [5] Francis C. Hsuan. A stepwise Bayesian procedure. *Annals of Statistics*, 7:860–868, 1979.
- [6] V. M. Joshi. Admissibility and Bayes estimation in sampling finite populations–iv. *Annals of Mathematical Statistics*, 37:1658–1670, 1966.
- [7] A. Y. Lo. A Bayesian bootstrap for a finite population. *Annals of Statistics*, 16:1684–1988, 1988.
- [8] G. Meeden. Interval estimators for the population mean for skewed distributions with a small sample size. *Journal of Applied Statistics*, 26:81–96, 1999.
- [9] M. N. Murthy. Ordered and unordered estimators in sampling without replacement. *Sankhyā*, 18:379–390, 1957.
- [10] Donald B. Rubin. The Bayesian bootstrap. *Annals of Statistics*, 9:130–134, 1981.
- [11] A. J. Scott. On admissibility and uniform admissibility in finite population sampling. *Annals of Statistics*, 3:489–491, 1975.