Using prior information about population quantiles in finite population sampling

David Nelson
University of Minnesota
Minneapolis, MN 55455

Glen Meeden*
School of Statistics
University of Minnesota
Minneapolis, MN 55455

February 1998

*Research supported in part by NSF Grant DMS 9401191

Using prior information about population quantiles in finite population sampling

SUMMARY

Consider the problem of estimating the population mean when it is known a priori that the population median belongs to some interval. More generally consider the problem of estimating the population mean when it is known a priori that the median of some auxiliary variable belongs to some interval. In this note we show how the Polya posterior, which is useful in problems where little prior information is available, can be adapted to these new situations. The resulting point and interval estimators are shown to have good frequentist properties.

Key Words: quantiles, stratification, Polya posterior, finite population sampling, prior information and interval estimation.

1 Introduction

Many classical methods in finite population sampling have been developed to incorporate some fairly standard types of prior information about the population. When the prior information does not fall in the standard categories it can be difficult to find procedures which reflect the information and have good frequentist properties. Recently methods based on the empirical likelihood were considered in Chen and Qin (1993) for estimating the median of the characteristic of interest when the median of an auxiliary variable is known.

The Polya posterior is a noninformative 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 and leads to admissible procedures. For further discussion of the Polya posterior see Ghosh and Meeden (1997). In this note we will consider how prior information about the population median or more generally population quantiles can be used in estimating various functions of the population. In section two we consider the problem of estimating the population mean of a characteristic of interest when it is know a priori that the characteristic's median belongs to some specified interval. We will show how the Polya posterior can be adapted to this problem and give the corresponding stepwise Bayes justification. In section three we assume that an auxiliary variable is present and consider the problem of estimating the mean of the characteristic of interest when it is known a priori that the auxiliary variable's median belongs to some specified interval. Simulation studies are presented to show how these modifications of the Polya posterior work in practice. Section four contains some concluding remarks.

2 Prior information on quantiles as constraints on stratum membership

A special case of the type of problem considered in this section is that of estimating the mean of a population characteristic when it is known or believed a priori that the population median of this characteristic falls in some known interval of real numbers. This information naturally stratifies the population into three strata; those units falling below the interval for the median, those falling in the interval and those falling above the interval. These strata will play an important role in what follows and keeping this example in mind should make the notation easier to understand.

2.1 Some notation

Consider a finite population consisting of N units labeled $1, 2, \ldots, N$. The labels are assumed to be known and often contain some information about the units. For each unit i let y_i, a real number, be the unknown value of some characteristic of interest. We will consider the estimation of a function, $\gamma(\mathbf{y})$, of the unknown state of nature, $\mathbf{y} = (y_1, \ldots, y_N)$. This state of nature is assumed to belong to a set $\mathcal{Y}$ where $\mathcal{Y}$ is a subset of N-dimensional Euclidean space, $\Re^N$. In practice this parameter space, $\mathcal{Y}$, is usually assumed to be all of $\Re^N$ or $\Re^{N^+}$. Typically we will have prior information that will influence the specification of the parameter space. Additionally, we will assume that the values for the characteristic of interest are elements of some specified finite set of k real numbers $\mathbf{b} = \{\mathbf{b}_1, \ldots, \mathbf{b}_k\}$. The parameter space then will be a subset of

$$\mathcal{Y}(\mathbf{b}) = \{\mathbf{y} \mid \text{For } i = 1, \ldots, N, \;\; y_i = b_j \text{ for some } j = 1, \ldots, k\}. \tag{1}$$

Often there is an additional or auxiliary characteristic associated with each element in the population. For each unit i let x_i be the value for this auxiliary characteristic. The vector of these values for the auxiliary characteristic is denoted as $\mathbf{x}$. We will assume that $\mathbf{x}$ is not completely known.

A sample s is a subset of $\{1, 2, \ldots, N\}$. We will let $n(s)$ denote the number of elements in s. Let $\mathcal{S}$ denote the set of all possible samples. A sampling design is a probability measure p defined on $\mathcal{S}$. Given a parameter point $\mathbf{y} \in \mathcal{Y}$ and $s = \{i_1, \ldots, i_{n(s)}\}$, where $1 \leq i_1 < \cdots < i_{n(s)} \leq N$, let $\mathbf{y}(s) = \{y_{i_1}, \ldots, y_{i_{n(s)}}\}$. A sample point consists of the set of observed labels s along with the corresponding values for the characteristic of interest. Such a sample point can be denoted by

$$\begin{aligned} z &= (s, y_s) \\ &= (s, \{y_{s_{i_1}}, \ldots, y_{s_{i_{n(s)}}}\}). \end{aligned}$$

The set of possible sample points depends on both the parameter space and the design. The sample space can be written as

$$Z(\mathcal{Y}, p) = \{(s, y_s) \mid p(s) > 0 \text{ and } y_s = \mathbf{y}(s) \text{ for some } \mathbf{y} \in \mathcal{Y}\}.$$

In what follows the design will always be simple random sampling without replacement of size n. So we can suppress the design p in the notation for the sample space.

2.2 The Polya posterior

Given the data 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 data point (s, y_s) we construct this distribution as follows. We consider an urn that contains n balls, where ball one is given the value $y_{s_{i_1}}$, ball two the value $y_{s_{i_2}}$ and so on. We begin by choosing a ball at random from the urn and assigning its value to the unobserved unit in the population with the smallest label. This ball and an additional ball with the same value are returned to the urn. Another ball is chosen at random from the urn and we assign its value to the unobserved unit in the population with the second smallest label. This second ball and another with the same value are returned to the urn. This process is continued until all $N-n$ unobserved units are assigned a value. Once they have all been assigned a value we have observed one realization 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 has been shown that for a variety of decision problems, procedures based on the 'Polya posterior' are generally admissible because they are stepwise Bayes. In these stepwise Bayes arguments a finite sequence of disjoint subsets of the parameter space is selected, where the order is important. A different prior distribution is defined on each of the subsets. Then the Bayes procedure is found for each sample point that receives positive probability under the first prior. Next the Bayes procedure is found for the second prior for each sample point which receives positive probability under the second prior and which was not taken care of under the first prior. Then the third prior is considered and so on. To prove the admissibility of a given procedure one must select the sequence of subsets, their order, and the sequence of priors appropriately. We will now see how this basic argument can be modified when certain types of prior information are available.

2.3 Constraints on population quantiles and stratification

Recall the problem of estimating the population mean when the population median is believed to belong to some interval. For convenience assume the size of the population, N, is even. This prior information induces a stratification of the population into three strata. The first stratum consists of those units for which the characteristic of interest is strictly less than the lower endpoint of the interval, the second stratum consists of those units for which the characteristic of interest falls in this interval, while the third stratum consists of those elements for which the characteristic of interest is strictly greater than the upper endpoint of the interval. Our prior information is then equivalent to a set of constraints on the numbers of elements in each of the strata. For the median of the characteristic of interest to fall in the specified interval the number of elements falling in the first strata as well as the number of elements falling in the third strata must both be less than $\frac{N}{2}$. More generally, with suitable conditions on N and n, this framework can be applied to situations in which we are willing to specify similar conditions on collections of quantiles for the characteristic of interest.

Suppose we wish to estimate a function $\gamma(\mathbf{y})$. In addition assume that each member of the population belongs to one of L known strata. Let $\mathbf{h}$ denote the stratum membership of the units of the population. So $\mathbf{h}$ is a vector of length N and each h_i belongs to the set $\{1, \ldots, L\}$. Further suppose that the prior information is equivalent to a set of constraints on the number of population elements falling in each of the strata. We assume that these constraints can be formally represented by the condition that $\mathbf{h} \in \mathcal{A}$ where $\mathcal{A}$ is a known subset of $[1, \ldots, \mathrm{L}]^N$.

Given such a setup how can the Polya posterior be modified to yield sensible estimates? Consider again the special case where we wish to estimate the population mean and believe the population median belongs to some interval. Assume we have a sample which is consistent with our prior beliefs, that is some members of the sample fall in each of the three intervals defined by the prior information. Since the stratum membership is known only for the units in the sample the first step is to use the Polya posterior on the strata in the sample to simulate strata membership for the unobserved units in the rest of the population so that the simulated population for $\mathbf{h}$ belongs to $\mathcal{A}$. Next for each stratum we independently use the Polya posterior on the values appearing in the sample for that stratum to generate possible values for the unobserved units which have been assigned to that stratum by the simulation for strata membership.

What if we observe a sample which is not consistent with our constraints? For example, suppose all the observations fall into the lower interval. Then there is no way just using the Polya posterior to generate simulated copies of the entire population that satisfy the constraints of the prior information. One possibility is to assume the prior information is incorrect and just use the usual Polya posterior. Another is to modify the Polya posteriors using a prior guess for the population to seed the urn in which the sample values are placed to generate simulated copies of the population that satisfy the constraints. Both approaches will be considered in the following and each of them can be given a stepwise Bayes interpretation.

2.4 Proving Admissibility

In this section we will prove that the inferential procedure which uses the modified Polya posterior for samples which are consistent with the constraints and ignores the constraints for samples which are not consistent with the constraints yields admissible estimators.

We begin with some more notation. For each $k = 1, \ldots, L$ let $\mathbf{b}^k$ be the finite set of possible values for the characteristic of interest for those population members which belong to the k^{th} stratum. Let $|\mathbf{b}^k|$ denote the number of elements belonging to $\mathbf{b}^k$. Let $\mathbf{b} = \{\mathbf{b}^1, \ldots, \mathbf{b}^L\}$. Since we are assuming the stratum membership is unknown for particular units a typical parameter point is now the pair $(\mathbf{y}, \mathbf{h})$. We will consider the parameter space

$$\begin{aligned} \mathcal{Y}(\mathbf{b}, L) &= \{(\mathbf{y}, \mathbf{h}) \,|\, \text{for } i = 1, \ldots, N \ y_i = b_j^k \\ &\qquad \text{for some } j = 1, \ldots, |\mathbf{b}^k| \text{ when } h_i = k\}. \end{aligned} \tag{2}$$

Let $\mathcal{Y}(\mathbf{b}, L, \mathcal{A})$ denote the subset of $\mathcal{Y}(\mathbf{b}, L)$ where the constraints given by the prior information are satisfied. We assume that $\mathcal{Y}(\mathbf{b}, L, \mathcal{A})$ is nonempty and

$$\mathcal{Y}(\mathbf{b}, L, \mathcal{A}) = \{(\mathbf{y}, \mathbf{h}) \in \mathcal{Y}(\mathbf{b}, L) \,|\, \mathbf{h} \in \mathcal{A}\} \tag{3}$$

That is a parameter point belongs to the restricted parameter space if only if $h \in \mathcal{A}$ and there are no constraints on the values of the characteristic of interest beyond those given in $\mathbf{b}$.

For a sample s let $\mathbf{h}(s)$ and h_s be defined similarly as $\mathbf{y}(s)$ and y_s. Hence h_s is just the strata memberships for the units in the sample. That is, if

$s = \{i_1, \ldots, i_n\}$ then h_{s_j} is the stratum which contains unit i_j. We now denote a typical point of our sample space by $z = (s, y_s, h_s)$.

As we noted earlier in any stepwise Bayes admissibility argument the first step is to select in proper order the subsets of the parameter space to be considered at each stage. In this case we will need to consider all the possible subsets of the $\mathbf{b}^j$'s. To this end we consider all possible collections $\mathbf{c} = \{\mathbf{c}^1, \ldots, \mathbf{c}^L\}$ with $\mathbf{c}^j \subseteq \mathbf{b}^j$ for each $j = 1, \ldots, L$ where at least one of the $\mathbf{c}^j$'s is not the empty set. Note that $\mathcal{Y}(\mathbf{c}, L)$ and $\mathcal{Y}(\mathbf{c}, L, \mathcal{A})$ are defined in the obvious manner. Even though $\mathcal{Y}(\mathbf{b}, L, \mathcal{A})$ is assumed to be nonempty it is possible for certain choices of $\mathbf{c}$ for $\mathcal{Y}(\mathbf{c}, L, \mathcal{A})$ to be empty. We need to consider such cases so we can handle samples from which it is not possible to construct models for the population from the values in the sample which satisfy our prior information. Such samples will need to be handled differently from those samples for which it is possible to construct appropriate models for the population using just the observed values. Let

$$\Gamma = \{\mathbf{c} \subseteq \mathbf{b} \,|\, \mathcal{Y}(\mathbf{c}, L, \mathcal{A}) \neq \emptyset\}$$

and

$$\Gamma' = \{\mathbf{c} \subseteq \mathbf{b} \,|\, \mathcal{Y}(\mathbf{c}, L, \mathcal{A}) = \emptyset\}$$

In addition, for each collection $\mathbf{c}$ let

$$J_c = \left\{\mathbf{c}^i \in \mathbf{c} \,|\, \mathbf{c}^i \neq \emptyset\right\}$$

and let

$$I_c = \left\{i \in \{1, \ldots, L\} \,|\, \mathbf{c}^i \in J_c\right\}.$$

The elements of both J_c and I_c are ordered according to the indices of the strata.

For each $\mathbf{c} \in \Gamma$ let

$$\begin{aligned}
\mathcal{Y}^*(\mathbf{c}; \mathcal{A}) \;=\;& \{(\mathbf{y}, \mathbf{h}) \in \mathcal{Y}(\mathbf{b}, L) \,|\, \text{for } i = 1, \ldots, N,\ h_i \in I_c; \\
& \text{for } i = 1, \ldots, N \text{ and for each } j \in I_c \\
& \;\text{if } h_i = j \text{ then } y_i \in \mathbf{c}^j; \\
& \text{for each } j \in I_c \text{ and for } i_j = 1, \ldots, |\mathbf{c}^j| \ \exists\, l_{i_j} \\
& \text{such that } h_{l_{i_j}} = j \text{ and } y_{l_{i_j}} = \mathbf{c}^j_{i_j};\ \mathbf{h} \in \mathcal{A}\}.
\end{aligned}$$

For each $\mathbf{c} \in \Gamma'$ let

$$\begin{aligned}\mathcal{Y}^*(\mathbf{c};\,\mathcal{A}) \;=\;& \{(\mathbf{y},\mathbf{h}) \in \mathcal{Y}(\mathbf{b},L) \,|\, \text{for } i=1,\ldots,N,\ h_i \in I_c; \\ & \text{for } i=1,\ldots,N \text{ and for each } j \in I_c \\ & \ \text{if } h_i = j \text{ then } y_i \in \mathbf{c}^j; \\ & \text{for each } j \in I_c \text{ and for } i_j = 1,\ldots,|\mathbf{c}^j| \ \exists\, l_{i_j} \\ & \text{such that } h_{l_{i_j}} = j \text{ and } y_{l_{i_j}} = \mathbf{c}^j_{i_j}\}.\end{aligned}$$

For $\mathbf{c} \in \Gamma'$ the definition of $\mathcal{Y}^*(\mathbf{c};\,\mathcal{A})$ depends on $\mathcal{A}$ only through the fact that $\mathbf{c}$ belongs to Γ'. Note that in each $\mathcal{Y}^*(\mathbf{c};\mathcal{A})$ we are considering only populations for which the number of elements in each stratum is greater than the corresponding $|\mathbf{c}^j|$.

Next we need to consider in which order the $\mathcal{Y}^*(\mathbf{c};\,\mathcal{A})$'s will appear in the proof. This is done by ordering the $\mathbf{c} = (\mathbf{c}^1,\ldots,\mathbf{c}^L)$'s using the usual lexicographic order on the sizes of the $\mathbf{c}^j$, $j=1,\ldots,L$ with the order of lower strata increasing faster than the order of higher strata.

The final preliminary step is to specify the prior distribution which will be used on $\mathcal{Y}^*(\mathbf{c};\mathcal{A})$ at each stage of the argument. For such a $\mathbf{c} = \{\mathbf{c}^1,\ldots,\mathbf{c}^L\}$, for each $i=1,\ldots,|I_c|$, and for a $(\mathbf{y},\mathbf{h}) \in \mathcal{Y}^*(\mathbf{c};\mathcal{A})$ let $c_h(i)$ be the number of elements, h_j of $\mathbf{h}$, for which $h_j = I_c^i$ where I_c^i is the i^{th} element of I_c. For each i in I_c and for $j_i = 1,\ldots,|\mathbf{c}^i|$ let $c_y^i(j_i)$ be the number of elements, for which $y_{l_{j_i}} = \mathbf{c}^i_{j_i}$. For a $\mathbf{c}$ with $|I_c|$ greater than one let

$$B_{I_c} = \{\beta \in \Re^{|I_c|} \,|\, \text{for } i=1,\ldots,|I_c|,\ 0<\beta_i<1,\ \textstyle\sum_{i=1}^{|I_c|}\beta_i = 1\}$$

and $d\beta_{I_c} = \prod_{i=1}^{|I_c|-1} d\beta_i$. For each $\mathbf{c}^j$ contained in J_c with $|\mathbf{c}^j|$ greater than one let

$$D_{\mathbf{c}^j} = \{\theta_j \in \Re^{|\mathbf{c}^j|} \,|\, \text{for } i_j=1,\ldots,|\mathbf{c}^j|,\ 0<\theta_{j,i_j}<1,\ \textstyle\sum_{j_i=1}^{|\mathbf{c}^j|}\theta_{j,i_j} = 1\}$$

and $d\theta_{\mathbf{c}^j} = \prod_{i_j=1}^{|\mathbf{c}^j|-1} d\theta_{j,i_j}$. Finally, for each $\mathbf{c}$ let $I_{\mathbf{c}}^+ = \{i \in I_{\mathbf{c}} \,|\, |\mathbf{c}^i| > 1\}$.

Now we consider a collection of prior distributions $\{\pi_{\mathbf{c}}(\cdot,\cdot)\}$ as $\mathbf{c}$ ranges over all possible values. We begin by considering a $\mathbf{c} \in \Gamma'$. Then for $(\mathbf{y},\mathbf{h})$ in $\mathcal{Y}^*(\mathbf{c};\,\mathcal{A})$

$$\begin{aligned}\pi_{\mathbf{c}}(\mathbf{y},\mathbf{h}) \;=\;& \pi_{\mathbf{c}}(\mathbf{h})\pi_{\mathbf{c}}(\mathbf{y}|\mathbf{h}) \qquad (4) \\ \propto\;& \int_{B_{I_c}} \prod_{i=1}^{|I_c|} \beta_i^{c_h(i)-1} d\beta_{I_c} \times \prod_{j\in I_c^+} \int_{D_{\mathbf{c}^j}} \prod_{i_j=1}^{|\mathbf{c}^j|} \theta_{j,i_j}^{c_y^j(i_j)-1} d\theta_{\mathbf{c}^j}\end{aligned}$$

$$\propto \prod_{i=1}^{|I_c|} \Gamma(c_h(i)) \times \prod_{j \in I_c^+} \prod_{i_j=1}^{|\mathbf{c}^j|} \Gamma(c_y^j(i_j))$$

while $\pi_{\mathbf{c}}(\mathbf{y}, \mathbf{h}) = 0$ for $(\mathbf{y}, \mathbf{h})$ not in $\mathcal{Y}^*(\mathbf{c};\, \mathcal{A})$. If there is only one stratum observed in the sample, that is if $|I_c| = 1$, then the first factor is understood to be constant for all elements of $\mathcal{Y}^*(\mathbf{c};\, \mathcal{A})$. Similarly, if I_c^+ is empty then the second factor is understood to be constant for all elements of $\mathcal{Y}^*(\mathbf{c};\, \mathcal{A})$.

These distributions are expressed as the product of the kernels for the marginal distribution of the strata membership and the conditional distribution for the values of the characteristic of interest given this strata membership. Conditional on the distribution of the strata membership the distribution for the characteristic of interest within a stratum is independent of the distributions in the other strata. Furthermore since we are considering a $\mathbf{c} \in \Gamma'$ the constraint $\mathbf{h} \in \mathcal{A}$ plays no role.

Next we suppose $\mathbf{c} \in \Gamma$. Now the constraint will play a role but only in the marginal distribution of $\mathbf{h}$. In this case the marginal distribution $\mathbf{h}$ will have the same kernel as in the previous case except now it will be restricted to just those $\mathbf{h}$ which belong to $\mathcal{A}$. By assumption the constraint does not depend on what values appear in the strata, only the number in each strata, and so the conditional distribution of $\mathbf{y}$ given $\mathbf{h}$ will be just the same as in the previous case. Hence for $\mathbf{c} \in \Gamma$ we take as our prior

$$\pi_{\mathbf{c}}^{\mathcal{A}}(\mathbf{y}, \mathbf{h}) = \pi_{\mathbf{c}}^{\mathcal{A}}(\mathbf{h}) \pi_{\mathbf{c}}(\mathbf{y}|\mathbf{h}) \tag{5}$$

where

$$\pi_{\mathbf{c}}^{\mathcal{A}}(\mathbf{h}) = \pi_{\mathbf{c}}(\mathbf{h}) / \pi_{\mathbf{c}}(\mathcal{A}).$$

for $\mathbf{h} \in \mathcal{A}$.

We now are ready to state and prove a theorem which states that the family of priors defined in equations 4 and 5 are indeed stepwise Bayes. But before that we need one more bit of notation. For a given $\mathbf{c}$ and sample point $z = (s, y_s, h_s)$ let $c_{y_s}^i(j_i)$ and $c_{h_s}(i)$ be defined analogously to $c_y^i(j_i)$ and $c_h(i)$. For example $c_{h_s}(i)$ is the number of elements, h_{s_j} of h_s, for which $h_{s_j} = I_c^i$ where I_c^i is the i^{th} element of I_c.

Theorem 2.1 *For a fixed* $\mathbf{b}$ *and* L *consider the finite population with parameter space given in equation 2. Let* $\mathcal{A}$ *represent prior information about strata membership which satisfies equation 3. Suppose the sampling plan is simple random sampling without replacement of size* n *where* $2 \leq n < N$.

Then the family of priors given in equations 4 and 5 under the lexicographic order on the $\mathbf{c}$*'s yield a sequence of unique stepwise Bayes posteriors. These posteriors will give admissible decision rules for any decision problem with a loss function whose expected loss has a unique minimum against any prior.*

For a fixed $\mathbf{c} \in \Gamma$ *and a sample point* $z = (s, y_s, h_s)$ *which appears in the sample space of* $\mathcal{Y}^*(\mathbf{c}, \mathcal{A})$ *but not in the sample space of any* $\mathbf{c}'$ *which precedes* $\mathbf{c}$ *in the lexicographic order we have for any* $(\mathbf{y}, \mathbf{h}) \in \mathcal{Y}^*(\mathbf{c}, \mathcal{A})$ *which is consistent with the sample point* z *that*

$$\pi_{\mathbf{c}}^{\mathcal{A}}(\mathbf{y}, \mathbf{h} \,|\, y_s, h_s) \propto \prod_{i=1}^{|I_c|} \frac{\Gamma(c_h(i))}{\Gamma(c_{h_s}(i))} \times \prod_{j \in I_c^+} \prod_{i_j=1}^{|c^j|} \frac{\Gamma(c_y^j(i_j))}{\Gamma(c_{y_s}^j(i_j))}. \tag{6}$$

Proof. We will assume we are in the stage of the stepwise Bayes argument where the parameter space $\mathcal{Y}^*(\mathbf{c}, \mathcal{A})$ is being considered. This means that the sample points to be considered at this stage contain units which fall in every strata belonging to I_c. Moreover for each $i \in I_c$ all the values in $\mathbf{c}^i$ are observed. Let $z = (s, y_s, h_s)$ denote such a sample point.

Since all the posteriors being considered are Polya posteriors or restricted Polya posteriors we note a few facts about them. If $\mathbf{c} \in \Gamma'$ then the marginal prior distribution for $\mathbf{h}$ given in equation 4 is just the usual one leading to the Polya posterior. It is easy to check that under this prior the marginal the marginal distribution for h_s given the labels in the sample satisfies

$$\pi(h_s) \propto \prod_{i=1}^{|I_c|} \Gamma(c_{h_s}(i)).$$

This follows since the prior is proportional to a Dirichlet integral. For more details see page 34 of Ghosh and Meeden (1997).

We next prove a few facts about the priors and posteriors used in the Theorem. We always assume that any parameter point, $(\mathbf{y}, \mathbf{h})$, considered is consistent with the sample point $z = (s, y_s, h_s)$. For $\mathbf{c} \in \Gamma'$ we have

$$\begin{aligned}
\pi_{\mathbf{c}}(\mathbf{y}, \mathbf{h} \,|\, y_s, h_s) &= \frac{\pi_{\mathbf{c}}(\mathbf{y}, \mathbf{h})}{\pi_{\mathbf{c}}(y_s, h_s)} \\
&= \frac{\pi_{\mathbf{c}}(\mathbf{y}, \mathbf{h})}{\pi_{\mathbf{c}}(y_s, \mathbf{h})} \frac{\pi_{\mathbf{c}}(y_s, \mathbf{h})}{\pi_{\mathbf{c}}(y_s, h_s)} \\
&= \pi_{\mathbf{c}}(\mathbf{y} \,|\, \mathbf{h}, y_s) \frac{\pi_{\mathbf{c}}(\mathbf{h})}{\pi_{\mathbf{c}}(h_s)} \frac{\pi_{\mathbf{c}}(y_s \,|\, \mathbf{h})}{\pi_{\mathbf{c}}(y_s \,|\, h_s)} \\
&= \pi_{\mathbf{c}}(\mathbf{y} \,|\, \mathbf{h}, y_s) \pi_{\mathbf{c}}(\mathbf{h} \,|\, h_s)
\end{aligned} \tag{7}$$

where the last step follows since $\mathbf{h}$ is consistent with h_s and the last ratio is one. This follows in much the same way that the expression for $\pi(h_s)$ given above was found.

We now prove an analogous expression for $\mathbf{c} \in \Gamma$. In this case we have

$$\pi_{\mathbf{c}}^{\mathcal{A}}(\mathbf{y}, \mathbf{h} \mid y_s, h_s) = \frac{\pi_{\mathbf{c}}(\mathbf{y}, \mathbf{h} \mid y_s, h_s)}{\pi_{\mathbf{c}}(A \mid h_s)} \tag{8}$$

for $\mathbf{h}$ consistent with h_s and $\mathcal{A}$ and $\mathbf{y}$ consistent with y_s.

We begin by noting

$$\pi_{\mathbf{c}}^{\mathcal{A}}(\mathbf{y}, \mathbf{h} \mid y_s, h_s) = \pi_{\mathbf{c}}^{\mathcal{A}}(\mathbf{h} \mid y_s, h_s)\pi_{\mathbf{c}}^{\mathcal{A}}(\mathbf{y} \mid \mathbf{h}, y_s, h_s)$$

and proceed by finding alternative expressions for the factors in the right hand side of the above. Since

$$\begin{aligned} \pi_{\mathbf{c}}^{\mathcal{A}}(\mathbf{h} \mid y_s, h_s) &= \frac{\pi_{\mathbf{c}}^{\mathcal{A}}(y_s, \mathbf{h})}{\pi_{\mathbf{c}}^{\mathcal{A}}(y_s, h_s)} = \frac{\pi_{\mathbf{c}}^{\mathcal{A}}(\mathbf{h})\pi_{\mathbf{c}}^{\mathcal{A}}(y_s \mid \mathbf{h})}{\pi_{\mathbf{c}}^{\mathcal{A}}(h_s)\pi_{\mathbf{c}}^{\mathcal{A}}(y_s \mid h_s)} \\ &= \frac{\pi_{\mathbf{c}}^{\mathcal{A}}(\mathbf{h})}{\pi_{\mathbf{c}}^{\mathcal{A}}(h_s)} = \frac{\pi_{\mathbf{c}}(\mathbf{h})}{\pi_{\mathbf{c}}(h_s \cap \mathcal{A})} = \frac{\pi_{\mathbf{c}}(\mathbf{h} \mid h_s)}{\pi_{\mathbf{c}}(\mathcal{A} \mid h_s)} \end{aligned}$$

and

$$\pi_{\mathbf{c}}^{\mathcal{A}}(\mathbf{y} \mid \mathbf{h}, y_s, h_s) = \pi_{\mathbf{c}}(\mathbf{y} \mid \mathbf{h}, y_s)$$

equation 8 follows from the preceding two equalities and equation 7.

Suppose now we are at a typical stage of the stepwise Bayes proof of admissibility for a fixed $\mathbf{c} \in \Gamma$. Let $z = (s, y_s, h_s)$ be a sample point which appears in the sample space of $\mathcal{Y}^*(\mathbf{c}, \mathcal{A})$ but not in the sample space of any $\mathbf{c}'$ which precedes $\mathbf{c}$ in the lexicographic order. This implies that each member of I_c appears at least once in h_s and for each $\mathbf{c}^i \in J_c$ all the values of $\mathbf{c}^i$ appear in y_s in stratum i. If only some of the strata appear or only some of the corresponding values appear such a sample point would have been considered for some $\mathbf{c}'$ preceding $\mathbf{c}$ in the lexicographic order.

We are now ready to compute the posterior distribution given $z = (s, y_s, h_s)$ where h_s is consistent with $\mathcal{A}$ since $\mathbf{c} \in \Gamma$. Following the argument on page 34 of Ghosh and Meeden (1997) for $\mathbf{h} \in \mathcal{A}$ we have

$$\pi_{\mathbf{c}}^{\mathcal{A}}(\mathbf{h} \mid h_s) \propto \frac{\int_{B_{I_c}} \prod_{i=1}^{|I_c|} \beta_i^{c_h(i)-1} d\beta_{I_c}}{\int_{B_{I_c}} \prod_{i=1}^{|I_c|} \beta_i^{c_{h_s}(i)-1} d\beta_{I_c}}.$$

In essentially the same way $\pi_{\mathbf{c}}(\mathbf{y} \mid \mathbf{h}, y_s)$ can be found. Then equation 6 follows easily from equations 7 and 8. If $\mathbf{c} \in \Gamma'$ the proof is very similar and hence the theorem is proved.

It should be noted that the sampling design played no role in the proof and the theorem is true for a large class of designs. For more details on this point see Scott (1995).

The collection of posterior distributions in the theorem not only provides us with a method for incorporating the prior information that $\mathbf{h}$ falls in $\mathcal{A}$ they are easy to simulate from so that point and set estimators of population parameters can easily be found approximately. Given a sample point $z = (s, y_s, h_s)$ which is consistent with $\mathcal{A}$ we first use the Polya posterior for the strata labels in the sample to simulate a full population of strata labels. If the resulting simulated set of labels does not belong to $\mathcal{A}$ we simulate another set of possible labels, continuing until we get a set which falls in $\mathcal{A}$. Given a set of satisfactory labels we then use the Polya posterior independently within each stratum to simulate the values for unobserved members of the stratum from the observed sample values of that stratum. For a sample point which is not consistent with $\mathcal{A}$ the process is even easier since we do not need to check that the simulated copy of labels belongs to $\mathcal{A}$ since any set is acceptable.

In some cases it does not make much sense to ignore the prior information contained in $\mathcal{A}$ when the sample is inconsistent with it. For example consider the situation where we we wish to estimate the population mean and we know a priori that the median value for the characteristic of interest falls in a specified closed interval. If the population size is odd the population then must contain at least one element for which the characteristic of interest falls in this closed interval. The likelihood of observing a sample which does not contain any members of the population belonging to the interval may be relatively large if the interval is quite small. For such a sample it is not possible to simulate populations which satisfy the prior information. These difficulties can be eliminated if we include a prior guess for the population. We will show how this can be done in the next section where we assume that prior information is available for an auxiliary variable rather than the characteristic of interest.

3 Constraints on an auxiliary variable

The estimation techniques of the preceding section were developed to incorporate information concerning the characteristic of interest in the estimation process. Often in finite population sampling problems we possess information concerning an auxiliary variable rather than the characteristic of interest. We now turn to developing extensions of the Polya posterior which incorporate prior information concerning such an auxiliary variable. We again will focus on prior information which will allow us to stratify the population and place a set of constraints on the number of elements in each of the strata.

3.1 The posterior

The characteristic of interested is denoted by $\mathbf{y}$ and the auxiliary variable by $\mathbf{x}$, both of which are assumed to be unknown. We assume that we have some additional prior information involving the x_i's which allows us to stratify the population into L strata where $\mathbf{h}$ denotes strata membership for units in the population. The prior information is expressed by the set $\mathcal{A} \subset [1, \ldots, L]^N$. As before we assume that a parameter point $(\mathbf{y}, \mathbf{x}, \mathbf{h})$ satisfies the constraints of our prior information if and only if $\mathbf{h} \in \mathcal{A}$.

In addition we assume we have a prior guess for what are typical values of (x_i, h_i) of the auxiliary variable. Let $(\mathbf{a}, \mathbf{t}) = \{(a_j, t_j)\}_{j=1}^k$ be such a prior guess for possible values of (x_i, h_i) in the population. Let $\mathbf{v} = \{v_j\}_{j=1}^k$ be a set of respective probabilities for the possible values. This prior guess could be a quantification of any additional prior information we may have regarding the population. Such a guess will often contain just a few such pairs and can be much smaller than the population although one would hope that it is representative of the full population. For convenience we will assume that for each stratum the prior guess has at least one pair of values that falls in that stratum. For the prior guess to be compatible with the prior information that $\mathbf{h}$ is in $\mathcal{A}$ assume that the probabilities of each of the strata with respect to $\mathbf{v}$, multiplied by the population size, satisfy the constraints on the sizes of the strata. The guess $\{(\mathbf{a}, \mathbf{t}), \mathbf{v}\}$ will be said to be $\mathcal{A}$-*consistent* if, given any sample, it is possible to construct a model for the population from the values in the sample and the guess that satisfies the prior constraint $\mathbf{h} \in \mathcal{A}$.

For each $i = 1, \ldots, L$ let w_i be a weight we attach to our prior guess for typical values for the auxiliary variable in the i^{th} stratum. Let $\mathbf{w} = \{w_i\}_{i=1}^L$. These weights may be thought of as measures of how good our guess or prior information may be in each of the strata. For convenience we will call the collection $\{(\mathbf{a}, \mathbf{t}), \mathbf{v}, \mathbf{w}\}$ a *strata weighted prior guess* for the typical values of the pairs (x_i, h_i).

For future use, for $i = 1, \ldots, L$ we let

$$v^i = \sum_{j:t_j=i} v_j$$

be the total probability our prior guess assigns to stratum i.

Let $z = (s, y_s, x_s, h_s)$ denote a possible sample point. Without formally stating and proving an admissibility result for the modified Polya posterior for this problem we will just give the form of the stepwise Bayes posterior for such a sample point.

For $i = 1, \ldots, L$ let Λ_s^i denote the distinct values of x_s, the values of the auxiliary variable in the sample, which appear in stratum i. Note this set may be empty. Let Λ_g^i denote the distinct values of a_j for which $t_j = i$. By assumption Λ_g^i is nonempty and the two sets need not be disjoint. Let $\Lambda^i = \Lambda_s^i \cup \Lambda_g^i$. If the j_i^{th} member of Λ^i belongs to Λ_g^i let $v_{j_i}^i$ be the conditional probability our prior guess assigns to this value, i.e., it is just its original assigned probability divided by v^i. If it does not belong to Λ_g^i then $v_{j_i}^i$ is set equal to zero.

The posterior is restricted to parameter points $(\mathbf{y}, \mathbf{x}, \mathbf{h})$ which are consistent with our prior guess and the sample. This means there are no restrictions on $\mathbf{h}$ beyond the fact it must belong to $\mathcal{A}$. Within stratum i the values of $\mathbf{x}$ must belong to Λ^i. Finally the ratios of the y_i/x_i's not in the sample are restricted to the observed values of the sample ratios. Let $\mathbf{r}$ denote this set of distinct observed values of y_s/x_s. For such a parameter point let $c_h(i)$ be the number of elements of $\mathbf{h}$ which fall in stratum i and $c_{h_s}(i)$ be the corresponding number for the sample h_s. Let $c_x^i(j_i)$ be the number of elements of $\mathbf{x}$ which fall in stratum i and take on the j_i^{th} value of Λ^i and $c_{x_s}^i(j_i)$ the corresponding sample quantity. Let $c_{y,x}(j)$ be the number of pairs of elements of $\mathbf{y}$ and $\mathbf{x}$ for which the ratio of y_i/x_i is equal to the the j^{th} member of $\mathbf{r}$ and $c_{y_s,x_s}(j)$ the corresponding sample quantity. With this notation the expression for the posterior is

$$\begin{aligned}
\pi_{\mathbf{r}}^{\mathcal{A}}(\mathbf{y}, \mathbf{x}, \mathbf{h} \mid y_s, x_s, h_s) &= \pi_{\mathbf{r}}^{\mathcal{A}}(\mathbf{h} \mid h_s)\pi_{\mathbf{r}}^{\mathcal{A}}(\mathbf{x} \mid x_s, \mathbf{h})\pi_{\mathbf{r}}^{\mathcal{A}}(\mathbf{y} \mid y_s, \mathbf{x}, \mathbf{h}) \\
&\propto \prod_{i=1}^{L} \frac{\Gamma(c_h(i) + w_i v^i)}{\Gamma(c_{h_s}(i) + w_i v^i)}
\end{aligned}$$

$$\times \prod_{i=1}^{L} \prod_{j_i=1}^{|\Lambda^i|} \frac{\Gamma(c_x^i(j_i) + w_i v_{j_i}^i)}{\Gamma(c_{x_s}^i(j_i) + w_i v_{j_i}^i)}$$
$$\times \prod_{j=1}^{|\mathbf{r}|} \frac{\Gamma(c_{y,x}(j))}{\Gamma(c_{y_s,x_s}(j))}. \tag{9}$$

One can easily simulate from this posterior. Given the sample one does Polya sampling from an urn which contains the observed sample strata labels and which has been "seeded" by our prior guess and weighted by our prior weights until one gets a simulated copy of the entire population of labels. Given such a simulated population of labels then within each stratum one simulates possible values for the auxiliary variable using the observed sample values and our weighted prior guess. The final step is to use the simulated copy of the auxiliary variable along with Polya sampling from the observed ratios of y_i/x_i to get a simulated ratio value for each unobserved unit. For each such unit we then multiply the simulated ratio and auxiliary value together to get a simulated value for the characteristic of interest.

Actually one can consider more general versions than the one just outlined. For example one can include in your prior guess possible values for the characteristic of interest. One can also assume exchangeability of the ratios within each stratum or no exchangeability on the ratios at all. The notation and proofs follow those given here. Although the basic idea is the same the proofs can become much more complicated. Simulation studies were done to see how these methods would perform in practice. Some of these are presented in the next section. For additional discussion on these points see Nelson (1998).

3.2 Some simulation results

Simulations were conducted to examine the performance of these restricted Polya posteriors in situations where we possess prior information concerning an auxiliary characteristic. In particular we examined the estimation of both the mean and the median for the characteristic of interest in situations where we possess prior information allowing us to specify an interval to contain the median for the auxiliary characteristic along with a prior guess for the typical values of the auxiliary characteristic.

Six constructed populations were considered in the simulations. Various superpopulation models were used to generate the populations. Each population consisted of 500 elements. The first four distributional relationships were considered in Meeden (1995) in simulations investigating the use of the Polya posterior to estimate the median value of the characteristic of interest. The first population, *gam5a*, was constructed by first generating the x_i's using a random sample from a gamma distribution with shape parameter 5 and scale parameter 1 shifted to the right by 10. Given the x_i's the conditional distribution of the y_i's was normal with mean $3x_i$ and variance x_i. The second population, *gam5b*, was constructed similarly but with conditional variance x_i^2. The third population, *exp1*, was generated using a random sample from an exponential distribution with parameter 1 shifted to the right by 50. The conditional distribution of the y_i's was normal with mean $80 - x_i$ and variance $\{.6 \ln(x_i)\}^2$. The fourth population, *lnnrma* was generated using a log-normal distribution with mean and variance of the log equal to 4.9 and 0.586 to generate the auxiliary characteristic with a normal conditional distribution for the characteristic of interest with mean $x_i + 2\ln(x_i)$ and variance x_i^2. The fifth population, *gam2*, was constructed by generating the auxiliary characteristic using a random sample from a gamma distribution with shape parameter 2 and scale parameter 1 and then generating the characteristic of interest using a normal conditional distribution with mean $2x_i + 3\sqrt{x_i}$ and variance $(.75x_i)^2$. The last population, *lnnrmb* was generated using a log-normal(2,.6) distribution for the auxiliary variable with mean and variance of the log equal to 2.6 and 0.6 and a normal conditional distribution for the characteristic of interest with mean $3x_i - \sqrt{x_i}$ and variance $2x_i$. Table 1 presents some key summary measures for these constructed populations.

Place Table 1 about here

We begin by considering point estimation for the population median of y for these populations. We will compare the method proposed here to several others which make various assumptions about what is known about the auxiliary variable. The first was the proposed in Meeden (1995). It assumes that the auxiliary variable is known for all units in the population and assumes that the ratios y_i/x_i are roughly exchangeable. This assumption is satisfied under the super population model $y_i = \beta x_i + x_i e_i$. The resulting estimator is denoted by *estpp*.

Chambers and Dunstan (1986) presented a model based estimator for the population distribution function. Actually they presented a family of estimators corresponding to different superpopulation models underlying the data. We will consider their estimator derived from the superpopulation model $y_i = \beta x_i + \sqrt{x_i} e_i$. We will denote their estimator by *estcd*

In a somewhat similar approach Rao, Kovar, and Mantel (1990) presented an approach to the estimation of population quantiles that is based on a design unbiased, asymptotically model unbiased method for estimating the cumulative distribution function. The form of the estimator depends on both the sampling design and the superpopulation model. We will consider the estimator based on simple random sampling and the superpopulation model $y_i = \beta x_i + \sqrt{x_i} e_i$. This estimator can be time consuming to compute so a method suggested in Kuk and Mak (1993) was used to approximate the value of the estimate in the following simulations. The notation *estrkm* will be used to refer to this estimation method.

Kuk and Mak (1989) proposed three methods for estimating the median value for the characteristic of interest that incorporate the sample values for the auxiliary variable along with the population median for the auxiliary variable. One uses a cross-classification of the values for the auxiliary characteristic and the characteristic of interest. This approach is based upon the level of concordance between the two variables rather than some superpopulation model. Kuk and Mak also presented a simple ratio estimator for the median that is directly analogous to the standard ratio estimator for estimating the mean. In the following discussion the notation *estkm* and *estratmd*, respectively, will be used to refer to these estimation techniques. Their third estimator which performed similarly to *estkm* becomes for this problem an estimator discussed in Chen and Qin (1993).

To examine the performance of the restricted Polya posterior the population thirty-fifth percentile and sixty-fifth percentile for the auxiliary variable were used to form the interval the prior information would specify as containing the median value for the auxiliary variable. Additionally, two prior guesses were considered in these simulations. The first prior guess consisted of every fifth population percentile for the auxiliary variable starting from the fifth percentile. The probability distribution specified for the prior guess placed equal weight on each percentile. The second prior guess consisted of every tenth population percentile starting from the fifth percentile. The probability distribution specified for this guess also placed equal weight on each percentile. In both cases the value $w_i = 5$ was used for weighting the prior guesses for the three strata. The notation *estrpp1* and *estrpp2*, respectively, will be used to refer to the use of this estimation technique with these two levels of prior information.

All of these estimators are based on the possession of some level of prior information concerning the auxiliary variable. The Polya posterior based method proposed by Meeden and the cumulative distribution based methods proposed by Chambers and Dunstan and by Rao, Kovar, and Mantel require that the value of the auxiliary variable is known for each member of the population. The analogue of the ratio estimator and the concordance based estimation technique both require that the median value for the auxiliary variable be known. Although knowledge of the median value is less complete than knowledge of all the values for the auxiliary variable such knowledge of the median value could be considered to be both precise and extensive knowledge. The restricted Polya posterior estimation technique developed here uses less extensive levels of prior information. The prior information needed consists of a specified interval to contain the median value for the auxiliary variable and a prior guess for typical values in the population. Both the length of the interval and the weight assigned to the prior guess can be adjusted to match the quality of the information used to construct these elements.

In each simulation, five hundred simple random samples of size 30 were obtained from the population under consideration and the estimators under study were either found directly or approximately by simulation. For each estimator its average value and mean absolute deviation of the estimates from the true median were found. As a bench mark these quantities were also found for the sample median, denoted by *estsmd*. When computing the point estimate for the procedures based on the Polya posterior the mean of the medians of 500 simulated copies of the population was used. The results are given in Table 2.

Place Table 2 about here

In the simulations considered by Meeden (1995) the Polya posterior estimation technique and the technique proposed by Chambers and Dunstan tended to result in better overall performance than the other methods. Surprisingly, in the current simulations the performance of the restricted Polya posterior estimation technique was comparable to the performance of these two estimation techniques in four of the six populations. It is important to remember that these two competitors assume that all the x_i are known and hence need more prior information to be used. Other than these two competitors the restricted Polya posterior performed much better than the other methods considered.

How should one rate the amount of prior information needed to use the method we are proposing here? Assuming that the median lies within the thirty-fifth percentile and sixty-fifth percentile seems to be something that could often happen in practice. Taking as the prior guess the true set of quantiles is perhaps more difficult to justify but is certainly much less stringent than assuming all the values of the auxiliary variable are known. Finally, we need to evaluate how much influence the w_i's have on our estimator. Recall in these simulations the sample size was 30 and $w_i = 5$ for $i = 1, 2, 3$. In the posterior given in equation 9 the w_i's appear in the first two of the three main factors. For the factor corresponding to $\pi_{\mathbf{r}}^{\mathcal{A}}(\mathbf{h} \mid h_s)$ we have $w_i v^i \doteq 5(1/3)$ and each $c_h(i)$ should on the average be about 10. So for stratum membership the sample values were weighted roughly six times the guessed values. In $\pi_{\mathbf{r}}^{\mathcal{A}}(\mathbf{x} \mid x_s, \mathbf{h})$ for a fixed stratum i we have $\sum_{j_i} w_i v_{j_i}^i \doteq 5$ and $\sum_{j_i} c_{x_s}^i(j_i)$ should on the average be about 10. So for values of the characteristic of interest within a stratum the sample values were given about twice as much weight on the average as the prior guesses. Even though the guessed values within a stratum are weighted more heavily than the guessed stratum memberships in these simulations it was the later which played the more crucial role in the performance of the proposed estimators. So in this instance the prior information played a relative minor role.

Interval estimates were formed only for the three Polya posterior based estimation techniques. The method used followed that of Meeden (1995) where the interval estimates obtained from the Polya posterior technique based on a known auxiliary variable were shown to have good frequentist properties. In general, the interval estimates formed here with the restricted Polya posterior covered as well as these earlier intervals but did tend to yield somewhat longer intervals.

Simulations were done to study the behavior of the methods proposed here for estimating the mean when prior information about median of the auxiliary variable was available. The Polya posterior methods were the same as for the previous problem except that rather than calculating the median value for each of the simulated populations the mean was calculated. Those three estimators were compared to the ratio estimator, the regression estimator and just the sample mean which ignores the auxiliary variable. These estimators are denoted by *estrat*, *estreg* and *estsmn*. Again 500 random samples of size 30 were taken from each population. The results are given in Table 3 and

Table 4.

$\boxed{\text{Place Table 3 and Table 4 about here}}$

In general, the restricted Polya posterior provided an improvement over the sample mean in estimating the population mean but did not perform as well as the other methods. This shows that prior information about the median of the auxiliary variable when estimating the mean is not as useful as when estimating the median.

How sensitive are these methods to the interval restriction placed on the median? To get some sense of this we return to the setup of the Theorem proved in the last section where we wish to estimate the mean of $\mathbf{y}$ and there is a restriction placed on the median of $\mathbf{y}$. Three populations of 500 units each were considered. The first population, *gam2*, was constructed as 50 plus 100 times a random sample from a gamma distribution with shape parameter 2 and scale parameter 1. The result was then truncated to the nearest integer. Population *gam5* was generated in the same manner using a random sample from a gamma distribution with shape parameter 5 and scale parameter 1. The last population, *lnrm6*, was constructed using the same shifting, rescaling and truncating of a random sample from a log-normal distribution with location parameter 0 and shape parameter 0.6. This last population is denoted by *lnrmc* For each of these populations the 35^{th} and 65^{th} percentiles were found and the resulting interval was assumed to be the interval specified to contain the median. Next these intervals were shifted to both the left and to the right by one-quarter, one-half, and three-quarters the distance between the median and the 35^{th} percentile. Each of the resulting intervals contained the median value for the characteristic of interest with larger shifts placing the median nearer the edge of the resulting interval. For each of the populations and each of the intervals 500 random samples were taken. For *gam2* and *gam5* the sample size was 20 while for *lnrmc* it was 30. The results are given in Table 5 and are fairly representative of the results observed across the group of simulations performed for these and other populations

$\boxed{\text{Place Table 5 about here}}$

As to be expected shifting the specified interval results in a shift of the estimator. Even though some bias is introduced the estimator proposed here still had smaller mean absolute deviation than the usual sample mean. The most negative effect of a poorly specified interval is an interval estimator with poor frequentist coverage properties. Although the average lengths of the usual normal theory intervals are not given in the table our simulations found for the scenario described in the theorem that the the restricted Polya posterior typically gave interval estimates that on the average were at least 20% to 25% shorter than the normal theory interval estimates. Hence if one can specify a good interval for the median the methods proposed here can result in significantly improved point and interval estimators.

4 Concluding remarks

Survey sampling is one area of statistics that routinely makes use of various types of prior information. However the methods are often quite case specific and in new situations it can be difficult to find procedures that make use of the prior information and still have good frequentist properties. On the other hand the Polya posterior in its simpliest form gives a noninformative Bayesian justification for some standard methods used in finite population sampling. In addition it can be adapted to incorporate different levels of prior information and the Bayesian formalism allows one to compute both point and interval estimates by simulation. In this note we have shown that this approach can be used when prior information about population quantiles for the characteristic of interest or an auxiliary variable is available. In some situations it does not lead to an improvement over standard methods but in others it gives answers with good frequentist properties which are difficult for standard methods to handle. Furthermore it seems to be reasonably robust against misspecification.

References

[1] R. L. Chambers and R. Dunstan. Estimating distribution functions from survey data. *Biometrika*, 73:597–604, 1986.

[2] Jiahua Chen and Jing Qin. Empirical likelihood estimation for finite populations and the effective usage of auxiliary information. *Biometrika*, 80:107–116, 1993.

[3] William Feller. *An introduction of probability theory and its applications, volume I.* Wiley, New York, 1968.

[4] Malay Ghosh and Glen Meeden. *Bayesian Methods for Finite Population Sampling.* Chapman and Hall, London, 1997.

[5] Anthony Y. C. Kuk and T. K. Mak. Median estimation in the presence of auxiliary information. *Journal of the Royal Statistical Society, Series B*, 51:261–269, 1989.

[6] Albert Lo. A Bayesian bootstrap for a finite population. *Annals of Statistics*, 16:1684–1695, 1988.

[7] Tak K. Mak and Anthony Kuk. A new method for estimating finite-population quantiles using auxiliary information. *Canadian Journal of Statistics*, 21:29–38, 1993.

[8] Glen Meeden. Median estimation using auxiliary information. *Survey Methodology*, 21:71–77, 1995.

[9] David Nelson. *Stepwise Bayes methods for incorporating prior information in finite population sampling.* PhD thesis, University of Minnesota, 1998.

[10] J. N. K. Rao, J. G. Kovar, and H. J. Mantel. On estimating distribution functions and quantiles from survey data using auxiliary information. *Biometrika*, 77:365–375, 1990.

[11] Donald Rubin. The Bayesian bootstrap. *Annals of Statistics*, 9:130–134, 1981.

[12] A. J. Scott. On admissibility and uniform admissibility in finite population sampling. *Annals of Statistics*, 3:489–491, 1975.

Table 1: Summary of Populations for Examining Mean and Median Estimation with Prior Information about the Median of an Auxiliary Variable

Population	Auxiliary Variable				Characteristic of Interest	
	Mean	Median	Q35	Q65	Mean	Median
gam5a	14.91	14.62	13.85	15.38	44.51	43.73
gam5b	15.00	14.53	13.86	15.47	45.12	43.82
exp	50.97	50.71	50.49	51.05	28.87	28.94
lnnrma	167.18	134.62	113.05	170.84	168.72	128.01
gam2	150.66	140.27	129.83	158.71	342.37	332.08
lnnrmb	8.67	7.25	5.88	9.28	23.34	19.05

Table 2: Summary of Median Point Estimation Performance

	gam5a		*gam5b*		*exp1*	
	Ave	*MAD*	*Ave*	*MAD*	*Ave*	*MAD*
estsmd	43.44	1.54	43.31	2.48	28.80	0.51
estratmd	43.65	1.14	43.39	2.40	28.80	0.52
estkm	43.96	1.32	43.62	2.61	28.99	0.50
estrkm	43.69	1.15	43.93	2.27	28.91	0.52
estcd	43.67	0.59	44.19	2.20	28.96	0.48
estpp	43.53	0.61	43.80	2.15	28.95	0.46
estrpp1	43.58	0.80	43.80	2.27	28.94	0.44
estrpp2	43.58	0.81	43.80	2.26	28.94	0.44

	lnnrma		*gam2*		*lnnrmb*	
	Ave	*MAD*	*Ave*	*MAD*	*Ave*	*MAD*
estsmd	123.20	27.75	326.04	23.81	19.01	2.13
estratmd	124.74	27.45	326.60	22.31	19.22	1.25
estkm	135.22	29.67	335.33	23.56	19.74	1.44
estrkm	129.48	25.54	331.73	21.91	19.24	1.13
estcd	145.73	26.41	333.22	15.74	19.43	0.64
estpp	129.43	21.25	327.96	15.54	18.86	0.63
estrpp1	131.09	22.21	327.23	16.13	18.99	1.28
estrpp2	130.57	22.04	326.44	16.19	19.00	1.31

Table 3: Summary of Mean Point Estimation Performance

	gam5a		*gam5b*		*exp1*	
	Ave	*MAD*	*Ave*	*MAD*	*Ave*	*MAD*
estsmn	44.51	1.10	45.23	2.52	28.86	0.37
estrat	44.46	0.58	45.25	2.20	28.85	0.40
estreg	44.46	0.60	45.21	2.26	28.88	0.34
estpp	44.46	0.59	45.11	2.17	28.88	0.40
estrpp1	44.37	0.84	45.00	2.32	28.87	0.37
estrpp2	44.40	0.84	45.02	2.33	28.87	0.37

	lnnrma		*gam2*		*lnnrmb*	
	Ave	*MAD*	*Ave*	*MAD*	*Ave*	*MAD*
estsmn	169.37	33.95	342.76	16.83	23.39	2.58
estrat	169.98	31.00	343.14	14.90	23.35	0.69
estreg	171.56	33.45	343.30	15.14	23.36	0.70
estpp	178.09	26.69	347.37	15.13	22.89	0.80
estrpp1	174.62	27.95	344.74	15.51	22.52	1.91
estrpp2	175.79	28.16	344.48	15.34	22.70	1.86

Table 4: Summary of Mean Interval Estimation Performance

	gam5a		*gam5b*		*exp1*	
	Ave Length	*Coverage Rate*	*Ave Length*	*Coverage Rate*	*Ave Length*	*Coverage Rate*
estsmn	5.35	92.4	11.86	95.0	1.79	93.0
estrat	2.74	94.0	10.60	94.6	1.96	92.8
estreg	2.75	93.0	10.54	93.4	1.63	93.0
estpp	2.68	91.2	10.20	93.6	1.87	92.6
estrpp1	4.28	94.5	10.76	94.8	1.88	94.1
estrpp2	4.34	95.1	10.79	93.7	1.88	93.7
	lnnrma		*gam2*		*lnnrmb*	
	Ave Length	*Coverage Rate*	*Ave Length*	*Coverage Rate*	*Ave Length*	*Coverage Rate*
estsmn	154.5	90.8	85.6	94.6	11.51	91.6
estrat	138.5	90.2	77.8	97.0	3.13	90.8
estreg	134.6	86.0	75.6	95.8	3.00	90.0
estpp	111.4	88.8	73.4	95.0	3.43	89.8
estrpp1	126.0	91.3	84.3	97.0	7.94	91.0
estrpp2	128.0	92.5	84.7	97.6	8.19	93.2

Table 5: Effect of Shifting Prior Interval for Median on Performance of Restricted Polya Posterior. Population 35^{th} and 65^{th} Percentiles used as Interval Specified to Contain the Median.

gam2	Restricted Polya				Normal Theory	
Shift	Ave.	M.A.D.	Coverage	Length	M.A.D.	Coverage
−.75	234.91	20.32	86.2	85.67	23.82	93.0
−.5	240.40	19.08	88.6	89.06	24.56	91.4
−.25	241.87	18.14	89.0	88.24	23.85	90.6
0	247.44	19.33	89.6	89.18	25.48	89.4
.25	249.03	18.31	91.0	90.62	23.63	91.2
.5	256.63	19.28	94.4	93.14	23.75	91.4
.75	260.23	21.39	90.0	91.94	24.15	90.8

gam5	Restricted Polya				Normal Theory	
Shift	Ave.	M.A.D.	Coverage	Length	M.A.D.	Coverage
−.75	517.26	37.72	81.6	129.77	38.70	93.2
−.5	529.17	30.25	88.4	135.14	39.33	91.6
−.25	540.05	27.59	94.0	143.74	39.23	93.4
0	551.24	28.09	94.8	143.49	40.76	92.6
.25	557.12	26.32	94.8	142.20	37.06	94.4
.5	564.49	30.74	93.4	140.34	40.03	93.0
.75	570.65	33.08	90.8	138.43	41.76	91.8

lnrmc	Restricted Polya				Normal Theory	
Shift	Ave.	M.A.D.	Coverage	Length	M.A.D.	Coverage
−.75	156.76	5.19	91.2	24.02	6.12	93.6
−.5	157.71	4.74	94.2	24.34	6.00	93.6
−.25	158.74	4.82	96.8	24.89	6.27	94.4
0	159.42	4.88	95.6	25.14	6.23	93.6
.25	160.53	4.73	95.8	24.69	6.15	93.0
.5	160.89	5.01	96.0	24.43	6.34	93.2
.75	162.29	5.06	94.6	24.13	6.19	92.6