

A noninformative Bayesian approach to finite population sampling using auxiliary variables

Radu Lazar, Glen Meeden and David Nelson¹

Abstract

In finite population sampling prior information is often available in the form of partial knowledge about an auxiliary variable, for example its mean may be known. In such cases, the ratio estimator and the regression estimator are often used for estimating the population mean of the characteristic of interest. The Polya posterior has been developed as a noninformative Bayesian approach to survey sampling. It is appropriate when little or no prior information about the population is available. Here we show that it can be extended to incorporate types of partial prior information about auxiliary variables. We will see that it typically yields procedures with good frequentist properties even in some problems where standard frequentist methods are difficult to apply.

Key Words: Sample survey; Noninformative Bayes; Auxiliary variable; Linear constraints and Polya posterior.

1. Introduction

Finite population sampling is one area of statistics where prior information is used routinely when making inferences. In most cases this prior information is not incorporated into the problem in a Bayesian manner. One reason for this is that the prior information usually does not lead, in an obvious way, to a sensible prior distribution. In the predictive approach (see Valliant, Dorfman and Royall 2000) a model is assumed and its unknown parameters are estimated after the sample has been observed. In the usual frequentist theory the prior information is encapsulated in the probability sampling plan or sample design. Basu showed that after the sample has been observed, the sampling design plays no role in the posterior distribution for a Bayesian. (For this fact and more of Basu's thoughts on finite population sampling see Ghosh (1988).) Although this radical conclusion has not been accepted by all Bayesians it is true that the usual frequentist theory and the Bayesian approach to survey sampling do not have a lot in common.

Traditional theory in survey sampling has emphasized estimation of the population mean. When the population mean of an auxiliary variable is known *a priori* the ratio estimator or the regression estimator is often employed. If one wishes to estimate something other than the mean, say a population quantile or the population distribution function, or if one has prior information about the auxiliary variable other than its mean then new methods need to be developed. Recent work along this line can be found in Chen and Qin (1993), Chen and Sitter (1999), Mak and Kuk (1993), Kuk and Mak (1989), Rao, Kovar and Mantel (1990) and Chambers and Dunstan (1986).

One advantage of a Bayesian approach is that a sensible posterior distribution for the population will incorporate the prior information into the estimation of several population parameters. Even if the posterior does not have a closed expression for a particular estimator for any given sample we can find its value approximately. This is done by sampling from the posterior distribution to simulate complete copies of the population and employing Monte Carlo estimation methods. If the posterior does not have a convenient form for sampling one should be able to use Markov Chain Monte Carlo methods to implement the simulation process. For each such simulated copy one computes the value of the parameter of interest. By simulating many such full copies of the population one can find, approximately, the corresponding Bayes point and interval estimates of the given population parameter. The problem then is to find a sensible Bayesian population model which utilizes the type of prior information available for the auxiliary variable.

Often, sensible Bayesian models can be based on the Polya posterior. The Polya posterior is a noninformative Bayesian approach to finite population sampling which uses little or no prior information about the population. A good source for more discussion on this approach is Ghosh and Meeden (1997). It is appropriate when a classical survey sampler would be willing to use simple random sampling as their sampling design. Here we show how it can be extended to cases where prior information about an auxiliary variable is present. For example the mean or median of an auxiliary variable might be known exactly or known to belong to some interval of possible values.

1. Radu Lazar, School of Statistics, University of Minnesota, Minneapolis, MN 55455. E-mail: radu@stat.umn.edu; Glen Meeden, School of Statistics, University of Minnesota, Minneapolis, MN 55455. E-mail: glen@stat.umn.edu; David Nelson, Center for Chronic Disease Outcomes Research, Minneapolis VA Medical Center, Minneapolis, MN 55417. E-mail: Dave.Nelson@med.va.gov.

The basic idea is to constrain or restrict the Polya posterior to put positive probability only on simulated populations which satisfy the constraints specified by the prior information for the auxiliary variables. This appropriately constrained Polya posterior can then be used to make inferences about the population parameters of interest. In Nelson and Meeden (1998) the authors considered several scenarios where a constrained Polya posterior yielded sensible frequentist results. There it was assumed that information about the population quantiles of the auxiliary variable was known *a priori*. Here we assume that we have more general forms of prior information about either the population mean or population quantiles for a set of auxiliary variables. These quantities may either be known exactly or known only to belong to some interval.

In section 2 we review the Polya posterior. In section 3 we introduce the constrained Polya posterior and discuss how to use Markov Chain Monte Carlo methods to find, approximately, the value of an estimate based on this distribution. In section 4 we apply the Polya posterior in a variety of situations and compare it to standard methods. In section 5 we discuss more formally how it relates to some standard frequentist methods. In section 6 we see that the constrained Polya posterior can be used with designs other than simple random sampling. In section 7 we give a few concluding remarks. In the appendix we prove an admissibility result for the constrained Polya posterior which gives a theoretical justification for the methods presented here.

2. The Polya posterior

Consider a finite population consisting of N units. For unit i let y_i , a real number, be the unknown value of some characteristic of interest. We assume the unknown state of nature, $y = (y_1, \dots, y_N)$, belongs to some known subset of N -dimensional Euclidean space. Suppose we wish to estimate some function $\gamma(y)$, of the unknown state of nature. The next step for a proper Bayesian analysis would be to specify a prior distribution over the parameter space. Then, given a sample generated by the sampling design, one would determine the posterior distribution of the unobserved members of the population conditioned on the values of the observed units in the sample. In most cases the posterior will not depend on the sampling design.

The Polya posterior can be used like a proper posterior distribution although it does not arise from a proper Bayesian model. It would be appropriate when there is little known about the population and the sample is assumed to be representative of the population. An example when it would be appropriate is when the sampling design is simple

random sampling. Next, we briefly describe this distribution and outline its theoretical justification.

Given the data, the Polya posterior is a predictive joint distribution for the unobserved units in the population conditioned on the values in the sample. Given a sample we now show how to generate a set of possible values for the unobserved units from this distribution. Consider two urns where the first urn contains the n units in the sample along with their observed y values. The second urn contains the $N - n$ unsampled units. We begin by choosing one unit at random from each of the two urns. We then assign the observed y value of the unit selected from the first urn to the unit selected from the second urn and then place them both in the first urn. The urns now contain $n + 1$ and $N - n - 1$ balls respectively. This process is repeated until all the units have been moved from the second urn to the first and have been assigned a value. At each step in the process all the units in the first urn have the same probability of being selected. That is, the units which have been assigned a value are treated just like the ones that actually appeared in the sample. Once this is done, we have generated one complete realization of the population under the Polya posterior distribution. This simulated, completed copy contains the n observed values along with the $N - n$ simulated values for the unobserved members of the population. Hence, simple Polya sampling yields a predictive distribution for the unobserved given the observed. A good reference for Polya sampling is Feller (1968). The Polya posterior is related to the Bayesian bootstrap of Rubin (1981). See also Lo (1988) and Binder (1982).

This predictive distribution often generates estimators similar to standard frequentist estimators under simple random sampling. Consider, for example, estimation of the population mean. Before continuing we need a bit more notation.

Let s denote a possible sample of size $n(s)$. It is a subset of $\{1, 2, \dots, N\}$, the set of labels for the finite population. If $s = \{i_1, \dots, i_{n(s)}\}$ then $y_s = \{y_{i_1}, \dots, y_{i_{n(s)}}\}$ is the set of observed values for y , the characteristic of interest. We let $z = (s, y_s)$ denote a typical observed sample. Then given $z = (s, y_s)$ we have

$$\bar{z}_s = \sum_{j=1}^{n(s)} y_{i_j} / n(s)$$

and

$$\text{Var}(z) = \sum_{j=1}^{n(s)} (y_{i_j} - \bar{z}_s)^2 / (n(s) - 1)$$

are the sample mean and sample variance. Let $\gamma_{mn}(y) = \sum_{i=1}^N y_i / N$ be the population mean. Under the Polya posterior distribution,

$$E(\gamma_{mn}(y) | z) = \bar{z}_s$$

and

$$\text{Var}(\gamma_{mn}(y) | z) = (1 - f) \frac{\text{Var}(z)n(s) - 1}{n(s)} \frac{1}{n(s) + 1}$$

where $f = n(s)/N$. Note that, except for the last factor in the posterior variance, these two terms are just the sample mean and its variance under simple random sampling. The design probabilities play no explicit role in these calculations. Nonetheless, for the Polya posterior 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. It is in such situations that simple random sampling without replacement is used.

Under the Polya posterior the Bayesian credible interval for the population mean or point and interval estimates of other population quantities cannot always be found explicitly. In such cases it is easy to find these estimates approximately by repeatedly simulating completed copies of the population. For each simulated copy we calculate the population parameter of interest. Experience has shown that 500 to 1,000 simulated values will usually give good results. The mean of these computed values will be our point estimate and the 0.025 and 0.975 quantiles of these computed values will be our interval estimate.

Since under the Polya posterior the only y values that appear in a simulated completed copy of the population are those that appeared in the sample the Polya posterior is just a way to assign random weights, *i.e.*, probabilities, to the units in the sample. Under the Polya posterior the average weight assigned to each unit in the sample is $1/n(s)$ so, as we have seen, its estimate of the population mean is just the sample mean. It is this relationship and the Bayes like character of the Polya posterior which allows one to prove the admissibility of the sample mean for estimating the population mean under squared error loss. This suggests that inferential procedures based on the Polya posterior will tend to agree with frequentist procedures and will have good frequentist properties.

As further documentation of this point we note that recently two of the authors (Nelson and Meeden 2006) demonstrated that Bayesian credible intervals based on the Polya posterior for the population median agree asymptotically with the standard Woodruff interval (Woodruff 1952). For another example consider estimating either the mean or the total of a subpopulation or domain when a simple random sample from the entire population is used. Here the number of units in the sample which belong to the domain is a random variable. Hence the mean of the units in the sample which fall into the domain is the ratio of two random variables. This estimate is more complicated than the mean of all the units in the sample. To get an estimate of variance

for this estimator the usual frequentist method conditions on the number of units in the sample that are in the domain. However when estimating the domain total this conditional argument does not work and an unconditional method is used to get an estimate of variance. See for example Cochran (1976). Recently one of the authors (Meeden 2005) showed that inferences based on the Polya posterior agree with the usual frequentist answers. Hence the Polya posterior handles both situations with one simple theory. It is important to remember that conditioning in the frequentist approach can be done under simple random sampling but for more complex designs, conditioning is not generally feasible since the conditional randomization distribution is unknown. As a final example note that the usual frequentist two stage cluster sampling procedures can be justified from an extension of the Polya posterior (Meeden 1999).

The Polya posterior is similar in spirit to bootstrap methods for finite population sampling. Both methods use a type of exchangeability argument to generate pseudo-versions of the population. The basic idea for the bootstrap is found in Gross (1980). Suppose we have a simple random sample of size $n(s)$ from the population and suppose $N/n(s) = m$ is an integer. Given the sample we create a good guess for the population by combining m replicates of the sample. We then take repeated random samples of size $n(s)$ from this created population to study the behavior of the estimator of interest. The asymptotic properties of estimators can also be studied (see Booth, Bulter and Hall 1994 for details). This is in contrast to the Polya posterior which for a fixed sample generates complete versions of the population and examines the distribution of the parameter of interest in the population rather than properties of the estimator for the parameter. For the given population quantity of interest the properties of its estimator derive directly from this predictive distribution for the population values.

The Polya posterior is the Bayesian bootstrap of Rubin (1981) applied to finite population sampling. The original Bayesian bootstrap applies to a random sample from an infinite population. Rubin showed that the bootstrap and Bayesian bootstrap are operationally very similar. The same type of analogy holds for the finite population setup. To study the variability of an estimator each repeatedly assigns random weights to the units in the sample. The logic for assigning the weights are different in the two cases as well as their theoretical justifications. The bootstrap has an asymptotic justification under repeated random sampling. The Polya posterior has a decision theoretic justification based on its stepwise Bayes nature (Ghosh and Meeden 1997).

Rather than generating a complete copy of the population it is often more efficient to use a well known approximation

to the Polya posterior. Assume that the sampling fraction f is small. For $j = 1, \dots, n(s)$ let p_j be the proportion of units in a complete simulated copy of the entire population which take on the value y_j . Then, under the Polya posterior, $p = (p_1, \dots, p_{n(s)})$ has approximately a Dirichlet distribution with a parameter vector of all ones, *i.e.*, it is uniform on the $n(s) - 1$ dimensional simplex, where $\sum_{j=1}^{n(s)} p_j = 1$. This approach will be very useful when we consider the constrained Polya posterior.

3. The constrained Polya posterior

3.1 The basic idea

In many situations, in addition to the variable of interest, y , the sampler has in hand auxiliary variables, x , for which prior information is available. For example, the population mean, μ_x , of x could be known. Given a unit in a random sample we observe its pair of values (y, x) . Following our earlier notation we denote the sample by

$$z = (s, (y, x)_s) = (s, \{(y_{i_1}, x_{i_1}), \dots, (y_{i_{n(s)}}, x_{i_{n(s)}})\}).$$

In this situation the regression estimator is often used when estimating the population mean. How should the Polya posterior be adjusted to take into account the fact that the population mean of x is known? The simple answer is to constrain the predictive distribution to put mass only on populations consistent with the prior information. In practice, we would only generate completed copies of the population consistent with the known prior information. To see how this can be done we consider the approximate form of the Polya posterior described at the end of the previous section.

For $j = 1, \dots, n(s)$ let p_j be the proportion of units in a completed copy of the population that have the value (y_j, x_j) . Rather than using the uniform distribution for $p = (p_1, \dots, p_{n(s)})$ over the simplex to generate simulated copies of the population we should use the uniform distribution restricted to the subset of the simplex satisfying

$$\sum_{j=1}^{n(s)} p_j x_j = \mu_x. \quad (1)$$

Before describing how we can generate vectors of p from this constrained Polya posterior we consider how the resulting estimator is related to the regression estimator.

Numerous simulation results (not presented here) show that the constrained Polya posterior behaves very much like the regression estimator under simple random sampling. The following simple argument shows why these two point estimates should often agree even though the Polya posterior makes no assumptions about the relationship between y and x .

Suppose in the population $y_i = a + bx_i + \varepsilon_i$ where ε_i is a random error with expectation zero. Let \bar{X} be the known population mean of x . Then given a sample and p_i 's satisfying $\sum_{i \in s} p_i x_i = \bar{X}$ we have

$$\begin{aligned} E\left(\sum_{i \in s} p_i y_i\right) &= aE\left(\sum_{i \in s} p_i\right) + bE\left(\sum_{i \in s} p_i x_i\right) + \sum_{i \in s} E(p_i \varepsilon_i) \\ &\doteq a + b\bar{X} \\ &\doteq \bar{y}_s - \hat{b}\bar{x}_s + \hat{b}\bar{X} \\ &= \bar{y}_s + \hat{b}(\bar{X} - \bar{x}_s) \end{aligned}$$

where \hat{b} is the least squares estimate of b . Here the sample values are fixed and the p_i 's and ε_i 's are random and the expectation of the p_i 's is with respect to the constrained Polya posterior. The first approximation follows since under simple random sampling we expect to see balanced samples on the average and the p_i 's and ε_i 's to be roughly independent.

3.2 Linear constraints and the Polya posterior

Prior information involving auxiliary variables can arise in many ways. We have already discussed the case where the population mean of an auxiliary variable is known. Another case is knowing a population median. More generally one might only know that a population mean or median belongs to some interval of real numbers. Although such cases are little discussed in the usual design based literature they seem quite realistic. Another case is where a pair of auxiliary variables describe a two way table where each unit must belong to one of the cells and the population row and column totals for the numbers falling into each cell are known. Before describing the constrained Polya posterior approach to such problems we need to mention a minor technical point.

Suppose the population mean of the auxiliary variable x is known to equal $\mu(x)$. There will be samples where the value of x is less than $\mu(x)$ for each unit in the sample. In such cases it would be impossible to use the constrained Polya posterior. But as a practical matter this will hardly ever happen. We will always assume that the sample we are considering is "consistent" with the prior information. This is explained in more detail just below. In our simulation studies we always reject a sample which is not consistent and select another. Again, in most cases, the probability of having to reject a sample is very small.

Each of our examples of prior information can be represented by one or more linear equality or inequality constraints. We have seen that knowing the population mean yields one linear equality constraint. If one knows that the population mean falls in some interval this yields two

linear inequality constraints. We next develop some notation that will allow us to consider a variety of situations where prior information can be described using linear equality and inequality constraints.

We assume that in addition to the characteristic of interest y the population has a set of auxiliary variables x^1, x^2, \dots, x^m . For unit i let

$$(y_i, x_i) = (y_i, x_i^1, x_i^2, \dots, x_i^m)$$

be the vector of values for y and the auxiliary variables. We assume that for any unit in the sample this vector of values is observed. We assume the prior information about the population can be expressed through a set of weighted linear equality and inequality constraints on the distinct auxiliary values in the population with weights corresponding to the proportions of the population taking these individual distinct values. We illustrate this issue more precisely by explaining how we translate this prior information about the population to the observed sample values so that we can construct pseudo-versions of the population consistent with the prior information.

Let s be a sample and, for $j = 1, 2, \dots, n(s)$, let (y_{i_j}, x_{i_j}) be the observed values which for simplicity we assume are distinct. Let $p = (p_1, \dots, p_{n(s)})$ be the proportion of units which are assigned the value (y_{i_j}, x_{i_j}) in a simulated complete copy of the population. Any linear constraint on the population values of an auxiliary variable will translate in an obvious way to a linear constraint on these simulated values. For example, if the population mean of x^1 is known to be less than or equal to some value, say b_1 , then for the simulated population this becomes the constraint

$$\sum_{j=1}^{n(s)} p_j x_{i_j}^1 \leq b_1.$$

If the population median of x^2 is known to be equal to b_2 then the constraint for the simulated population becomes

$$\sum_{j=1}^{n(s)} p_j w_j = 0.5$$

where $w_j = 1$ if $x_{i_j}^2 \leq b_2$ and it is zero otherwise. If the population mean of x^2 is less than or equal to the population mean of x^3 then the simulated population constraint becomes

$$\sum_{j=1}^{n(s)} p_j (x_{i_j}^2 - x_{i_j}^3) \leq 0.$$

Hence, given a family of population constraints based on prior information and a sample we will be able to represent the corresponding constraints on the simulated p by two systems of equations

$$A_{1,s} p = b_1 \tag{2}$$

$$A_{2,s} p \leq b_2 \tag{3}$$

where $A_{1,s}$ and $A_{2,s}$ are $m_1 \times n(s)$ and $m_2 \times n(s)$ matrices and b_1 and b_2 are vectors of the appropriate dimensions. This generalizes the argument leading to equation 1.

We assume the sample is such that the subset of the simplex it defines by equations 2 and 3 is non-empty. For such a sample the asymptotic approximation to the constrained Polya posterior puts a uniform distribution over this subset of the simplex. Before addressing the issue of simulation from this distribution we note that it has a theoretical justification. It can be given a stepwise Bayes justification which guarantees that it will yield admissible procedures. Details are given in the appendix.

3.3 Computation

Let P denote the subset of the simplex which is defined by equations 2 and 3. P is a non-full dimension polytope. We would like to generate independent observations from the uniform distribution over P . Unfortunately we do not know how to do this. Instead, we use Markov chain Monte Carlo (MCMC) methods to generate dependent samples.

In particular we will use the Metropolis-Hastings algorithm which depends on using a Markov chain to generate a dependent sequence of random values for $p \in P$. The process works as follows. We begin by finding a starting point in p_0 in the relative interior of P . This is Step 1 below. Next we choose a random direction d in P . This is a bit tricky because the dimension of P is strictly less than $n(s) - 1$. This is accomplished in Steps 2 and 3 below. Next we find the line segment which is the intersection of the line passing through p_0 in direction d with P . This is Step 4 below. Next we choose a point at random from the uniform distribution over this line segment. This is the first observation in our Markov chain. We then repeat the process with this point playing the role of p_0 to get a second random point. Letting this second random point play the role of p_0 we get a third and so on. More formally our algorithm is:

- Step 1. Choose an initial positive probability vector p_0 such that $A_{1,s} p_0 = b_1$ and $A_{2,s} p_0 < b_2$ and set $i = 0$.
- Step 2. Generate a random direction d_i uniformly distributed over the unit sphere in R^n .
- Step 3. Let d_i^* be the normalized projection of d_i onto the null space of $A_{1,s}$.
- Step 4. Find the line segment $L_i = \{\alpha \in R \mid p_i + \alpha d_i^* \in P\}$ and generate α_i uniformly over the line segment.

Step 5. Set $p_{i+1} = p_i + \alpha_i d_i^*$ and $i = i + 1$ and go back to step 2.

At first glance it might not be clear what role the constraints are playing in this process. They are there however through the definition of P . The Markov chain generated in this way converges in distribution to the uniform distribution over the polytope. The convergence result of such mixing algorithms was proven by Smith (1984). If we wish to approximate the expected value of some function defined on P then the average of the function computed at the simulated values converges to its actual value. This allows one to compute point estimates of population parameters. Finding the 0.95 Bayesian credible interval approximately is more difficult.

One possibility is to run the chain for a long time; for example, we may generate 4.1 million values, throw away the first 100,000 values, and find the 0.025 and 0.975 quantiles of the remaining values. These two numbers will form our approximate 0.95 credible interval. In this manuscript we will only consider sample sizes of less than 100. For such sample sizes we have found that chains of a few million suffice.

How fast a chain mixes can depend on the constraints and the parameter being estimated. It seems to take longer to get good mixing when estimating the median than when estimating the mean. This is not surprising when one recalls that in standard bootstrap methods many more bootstrap samples are required when estimating quantiles rather than means. See for example Efron and Tibshirani (1993).

Another approach which can work well is to run the chain for a long time and then just use every m^{th} point where m is a large integer. Although this is inefficient it can give good answers when finding a 0.95 credible interval for the median.

4. Applications

In this section we show how various types of partial information about auxiliary variables can be incorporated in the estimation of the parameters when the constrained Polya posterior is employed. In many instances, the prior information used in the constrained Polya posterior estimation cannot be utilized by the standard frequentist methods.

4.1 Stratification

Stratification is a type of prior information which is commonly used in finite population sampling. We note that the usual stratified estimator can be thought of as arising from independent Polya posteriors within each stratum. Details can be found in Vardeman and Meeden (1984).

When, in addition to stratification, an auxiliary variable is present a good estimate of the population mean can be found by combining the estimates obtained from the regression estimator within each stratum. For details, see Cochran (1976). If only the population mean of the auxiliary variable is known then under standard approaches it is difficult to combine this information with stratification unless a common model is assumed across strata. The constrained Polya can incorporate both types of information which can lead to improved estimates yet it does not require the common model assumption.

To demonstrate, we constructed a stratified population of size 900 consisting of three strata. The strata sizes were 300, 200 and 400. There were two auxiliary variables, say x_1 and x_2 . In stratum one the $x_{1,i}$'s were a random sample from a gamma(10,1) distribution and the $x_{2,i}$'s were a random sample from a gamma(2,1) distribution. In the second stratum the $x_{1,i}$'s and the $x_{2,i}$'s were generated by the gamma(15,1) and the gamma(7,1) distributions respectively. In the third stratum the $x_{1,i}$'s and the $x_{2,i}$'s were generated by the gamma(5,1) and the gamma(3,1) distributions respectively. The characteristic of interest for the population was generated as follows:

$$\text{stratum 1: } y_i = 1 + x_{1i} x_{2i} + \varepsilon_i$$

$$\text{stratum 2: } y_i = 3 + x_{1i} + x_{1i} x_{2i} + \varepsilon_i$$

$$\text{stratum 3: } y_i = 2 + x_{2i} + x_{1i} x_{2i} + \varepsilon_i$$

where in stratum one the ε_i 's were normal(0,1), while in stratum two they were normal(0,1.5²), and in stratum three they were normal(0,3.5²). All the ε_i 's were independent.

In addition to the strata sizes we assumed that the population median of x_1 and the population mean of x_2 were known. We generated 500 random samples according to our sampling plan drawing 75 units such that 25 units were in the first stratum, 20 units were in the second stratum and 35 units were in the third stratum. For each sample we computed the sample mean, the usual stratified estimate which is the sum of the sample means within each stratum adjusted for the size of all strata, the constrained Polya estimate, and the corresponding 95% confidence intervals and 0.95 credible intervals for these estimates.

The results of the simulations are given in Table 1. From the table, we see that the constrained Polya estimator on average agrees with the usual stratified estimator and is essentially unbiased. But its average absolute error is much smaller than the average errors of the other two. This is to be expected since the more information an estimator uses the better it should perform and the constrained Polya estimates are using information from the auxiliary variables that is ignored by the estimates which just use stratification. Note that the constrained Polya made no assumptions about how y and x_1 and x_2 were related. Furthermore it is not clear

how standard methods could make use of knowing the population median of x_1 and the population mean of x_2 . If just information about means is available then the empirical likelihood based methods of Chen and Sitter (1999) and Zhong and Rao (2000) could be used. The results clearly show that the constrained Polya posterior is utilizing this additional information in a sensible manner.

Table 1
Simulation results for the stratified example where the median of the first auxiliary variable and the mean of the second are known

Method	point estimate		95% confidence or credible intervals		
	Ave. of estimate	Ave. of absolute error	Ave. of lower bound	Ave. length	Freq. of coverage
Meanest	47.978	4.821	36.44	23.09	1.000
Strataest	43.395	2.072	38.22	10.35	0.942
Polyaest	43.355	1.516	40.19	6.75	0.936

In this example the constrained Polya estimates were obtained using Markov chains of length 4,000,000 after the initial 100,000 points were discarded.

4.2 Categorical auxiliary variables

Assume that the elements of a population of known size N are associated with the elements of k categorical auxiliary variables. For simplicity, we consider $k = 2$ but the theory applies to more than two categorical variables. If one auxiliary variable takes on r distinct values and the other takes on c distinct values they allow the elements of the population to be classified into a two-way table with $r \times c$ cells. Let N_{ij} be the number of elements in the population that belong to the ij -cell, for i in $\{1, \dots, r\}$ and j in $\{1, \dots, c\}$, then $\sum_{i=1}^r \sum_{j=1}^c N_{ij} = N$. If the N_{ij} 's are known and s is a random sample with n_{ij} elements from the ij -cell then a good estimate of the population mean is given by

$$\sum_{i=1}^r \sum_{j=1}^c \frac{N_{ij}}{N} \bar{y}_{ij}^s$$

where \bar{y}_{ij}^s is the mean of the n_{ij} elements from the ij -cell in the sample. This is the usual stratified estimator where the cells in the table are considered the strata.

A harder problem is the estimation of the population mean when the counts, N_{ij} 's, are not known but the marginal counts are known. Let $N_{i\cdot} = \sum_{j=1}^c N_{ij}$ denote the marginal row counts, for i in $\{1, \dots, r\}$ and $N_{\cdot j} = \sum_{i=1}^r N_{ij}$ denote the marginal column counts, for j in $\{1, \dots, c\}$. In such cases, one way of estimating the population mean is the frequentist procedure called calibration or raking. In this procedure, given a sample s , the estimator is given by $\sum_{k \in s} \hat{w}_k y_k$, where the \hat{w}_k 's are not the design weights but are new weights assigned to the units in the sample. A good set of weights needs to satisfy two conditions. The first is

that the weights must preserve the known marginal counts, for example, $\sum_{k \in s(\cdot, j)} \hat{w}_k = N_{\cdot j}$ where $s(\cdot, j)$ is the portion of the sample falling in the j^{th} column of the two-way table. The second is that the weights should in some sense be close to the sampling design weights $1/\pi_k$, where $\pi_k = P(k \in s)$. Depending on the function used to measure the distance different calibration estimators can be obtained. Although this is a sensible idea, selecting the right distance measure and then getting a sensible estimate of variance for the resulting estimator has no standard frequentist answer. For details, see Deville and Särndal (1992).

The Polya posterior gives an alternative approach to this problem since the information provided by the known marginal totals determines a set of linear constraints on the random weights it assigns to the units in the sample. If there are continuous auxiliary variables for which we have prior information then additional constraints can be added. To see how this could work in practice we considered a simple example with two dichotomous variables so each unit can be classified into a cell of a 2×2 table, together with a third continuous auxiliary characteristic. We assumed four different levels of prior information.

1. The marginal counts for the 2-way table are known.
2. The marginal counts and the mean of the continuous auxiliary variable are known.
3. The marginal counts and the median of the continuous auxiliary variable are known.
4. The marginal counts are known and the mean of the continuous auxiliary variable is known to lie between two bounds. We chose the 45th and 65th quantiles of its population of values to specify these bounds.

For each case we formed a population using the following model where all the random variables are independent.

Cell 1,1 $x_i \sim \text{gamma}(8,1)$, $\varepsilon_i \sim \text{normal}(0,7^2)$ and $y_i = 25 + 3x_i + \varepsilon_i$ for i in $\{1, \dots, 150\}$.

Cell 1,2 $x_i \sim \text{gamma}(10,1)$, $\varepsilon_i \sim \text{normal}(0,7^2)$ and $y_i = 25 + 3x_i + \varepsilon_i$ for i in $\{1, \dots, 350\}$.

Cell 2,1 $x_i \sim \text{gamma}(6,1)$, $\varepsilon_i \sim \text{normal}(0,4^2)$ and $y_i = 25 + 2x_i + \varepsilon_i$ for i in $\{1, \dots, 250\}$.

Cell 2,2 $x_i \sim \text{gamma}(4,1)$, $\varepsilon_i \sim \text{normal}(0,4^2)$ and $y_i = 25 + 2x_i + \varepsilon_i$ for i in $\{1, \dots, 250\}$.

For each of the cases we generated a population and took 500 random samples of size 80 with 20 units from each cell. For each sample we computed the sample mean and the stratification estimate, assuming that the true population cell counts were known, and their corresponding 95% confidence intervals. We also computed the constrained Polya

estimates along with their 0.95 credible intervals. The constrained Polya estimates were obtained from the last 4,000,000 points of a Markov chain of size 4,100,000. The results of the simulations are given in Tables 2 through 5. The results in the tables show that the constrained Polya estimates based on known marginal counts and a known mean, median or known interval about the mean are better than the strata estimates based on known cell counts. The stratified estimates are better than the constrained Polya posterior only when the constrained Polya posterior only makes use of the known marginal counts.

Table 2
Simulation results for the categorical example when just the marginal cell counts are assumed known

Method	point estimate		95% confidence or credible intervals		
	Ave. of estimate	Ave. of absolute error	Ave. of lower bound	Ave. length	Freq. of coverage
Meanest	43.805	0.919	41.107	5.396	0.976
Strataest	44.355	0.846	42.259	4.191	0.940
Polyaest	43.909	0.896	41.863	4.197	0.922

Table 3
Simulation results for the categorical example when the marginal cell counts and the mean of the auxiliary variable are assumed known

Method	point estimate		95% confidence or credible intervals		
	Ave. of estimate	Ave. of absolute error	Ave. of lower bound	Ave. length	Freq. of coverage
Meanest	43.804	0.922	41.063	5.482	0.964
Strataest	44.399	0.862	42.272	4.256	0.942
Polyaest	44.506	0.510	43.257	2.497	0.960

Table 4
Simulation results for the categorical example when the marginal cell counts and the mean of the auxiliary variable are assumed known

Method	point estimate		95% confidence or credible intervals		
	Ave. of estimate	Ave. of absolute error	Ave. of lower bound	Ave. length	Freq. of coverage
Meanest	43.439	0.877	40.783	5.312	0.986
Strataest	43.927	0.884	41.804	4.244	0.940
Polyaest	43.784	0.785	42.032	3.640	0.920

Table 5
Simulation results for the categorical example when the marginal cell counts are assumed known and the mean of the auxiliary variable is known to lie between its known 45th and 65th quantiles

Method	point estimate		95% confidence or credible intervals		
	Ave. of estimate	Ave. of absolute error	Ave. of lower bound	Ave. length	Freq. of coverage
Meanest	43.463	0.840	40.789	5.348	0.978
Strataest	43.948	0.865	41.825	4.245	0.948
Polyaest	43.519	0.829	41.555	4.029	0.938

4.3 An example

In this section we consider data from the Veterans Health Administration. In 1998 the VA Upper Midwest Health Care Network administered a functional status survey of the veteran users of the VA facilities within the network (Singh, Borowsky, Nugent, Murdoch, Zhao, Nelson, Petzel and Nichol 2005). Veterans eligible for this survey were those with any outpatient encounter or inpatient stay between October 1997 and March 1998 at any one of the five VA facilities in the network. In addition to basic demographic measures, such as age and sex, the primary component of the survey was the SF36-V (Kazis, Miller, Clark, Skinner, Lee, Rogers, Spiro, Payne, Fincke, Selim and Linzer 1998). This health-related quality of life survey instrument consists of eight sub-scales of physical functioning, role limitations due to physical problems, bodily pain, general health, energy/vitality, social functioning, role limitations due to emotional problems, and mental health. These scales are combined to form physical (PCS) and mental (MCS) component summary scores. Larger scores represent better health status. VHA administrative data measuring major comorbid conditions present in the year before the survey were also collected.

From the population of one of the five facilities we selected a subpopulation comprising all of the women and a random subset of the men to form a population of 2,500 individuals. For purposes of this example the number of comorbidities was categorized into three categories to represent measures of good, average and poor health. We then selected 200 stratified random samples of size 100 from the population. The strata sizes along with the sample sizes are given in table 6. Our sampling plan over sampled the women. Such unbalanced sampling plans can often occur in practice.

Table 6
The strata sizes along with the sample sizes for the Veterans Administration data

	Good	Average	Poor
F	353(20)	155(10)	117(10)
M	890(30)	493(20)	492(10)

We compared three different estimators of the mean PCS score for this population of 2,500; the sample mean which ignores the stratification, the usual stratified estimator which assume the strata sizes are known, and a constrained Polya posterior estimator which assumes that the marginal row and column totals of table 6 are known along with the average age of the individuals in the population. The population correlation between *PCS* and age is -0.22. The correlations of *PCS* with gender and with categorized comorbidity-based state of health are -0.13 and -0.28. From

the results in table 7 we see that the constrained Polya estimator performs about the same as the stratified estimator and both are a bit better than the sample mean. To compute the constrained Polya estimator we generated Markov Chains of length 7,000,000.

Table 7
Results for estimating PCS in the Veterans Administration data. The constrained Polya estimator assumes the row and column totals are known along with the average age of the individuals in the population

Method	point estimate		95% confidence or credible intervals		
	Ave. of estimate	Ave. of absolute error	Ave. of lower bound	Ave. length	Freq. of coverage
Meanest	37.235	1.040	34.907	4.650	0.938
Strataest	36.648	0.925	34.322	4.651	0.948
Polyaest	36.644	0.925	34.344	4.605	0.958

5. Relation to empirical likelihood methods

In this section we review some frequentist methods for problems where constraints are involved and discuss their relationship to the constrained Polya posterior.

Chen and Qin (1993) considered an empirical likelihood approach to estimation in survey sampling when prior information about an auxiliary characteristic is available. To construct estimators after the sample has been observed the units in the sample are weighted to reflect the prior information. For example, suppose that the sample mean is less than the known population mean of the x values. Then positive weights, which sum to one, are selected for the sampled units such that the mean of the x_s values under the probability distribution given by the weights satisfies the known constraint. Although these weights can not be found explicitly they are easy to compute. When estimating the population mean of y the resulting estimator was first noted in Hartley and Rao (1968) and shown to be asymptotically equivalent to the regression estimator. If the population median of x is known then the units in the sample less than the known population median are given equal weights which sum to 0.5 and similarly for the sampled units with x values larger than the known population median. When estimating the population median the resulting estimator is one proposed by Kuk and Mak (1989).

An advantage of the constrained Polya posterior, and more generally of a Bayesian approach, is that it is straightforward to estimate many population quantities besides the mean without developing any new theory or methods. Given a simulated copy of the entire population which satisfies the constraints one just calculates the population parameter of interest. Then one uses such simulated values just as when one is estimating the mean.

To compare the Chen and Qin estimator of the population median of y with the constrained Polya posterior estimator when the population mean of x is known eight different populations were constructed. In half of the populations one would expect the regression estimator to do well in estimating the population mean while the remaining half did not satisfy the usual super-population model assumptions associated with the regression estimator. For each population 500 random samples of sizes 30 and 50 were taken, subject to satisfying the constraint that the sample contained values for x greater and less than the known mean. In all cases the two estimators using the prior information performed better than the sample median. These results were consistent with the simulation results of Chen and Qin. We calculated the average absolute error for the two estimators using the mean constraint. In each of the 16 different sets of simulations we then calculated the ratio of the constrained Polya posterior absolute error to that of the estimator of Chen and Qin. The range of these 16 values was 0.85 to 1.00 with a mean of 0.91. So in terms of absolute error, the constrained Polya posterior performed about 10% better, on average, than the estimator of Chen and Qin.

Suppose now that the population median of x is known. To simplify matters suppose that none of the actual values are equal to the population median of x . Let n_l be the number of units in the sample whose x values are less than the known population median of x . Then $n_u = n(s) - n_l$ is the number of units in the sample which are on the other side of the known median. Let $p_l = (p_1, \dots, p_{n_l})$ and $p_u = (p_1, \dots, p_{n_u})$ be two probability vectors. Intuitively, a sensible posterior distribution given the sample and the known population median would be for p_l and p_u to be independent Dirichlet distributions with all parameters equal to one with each of them assigned a weight of one half so that their total sum is one. It follows from the Theorem proved in the appendix that under our sampling plan these posteriors are stepwise Bayes. Note that under these posteriors the expected values of the proportions assigned to each unit in the sample are the weights assigned to the sample by Chen and Qin. This proves the admissibility of their estimator of the population median and consequently of Kuk and Mak's. Simulation results show that this constrained Polya posterior's 0.95 credible intervals cover approximately 95% of the time except in one special case. If the sample size is small and y and x are highly correlated then the medians for the simulated populations under the constrained Polya posterior do not vary enough and the resulting intervals are too short and their coverage frequency may be considerable less than 95%.

This close relation between the empirical likelihood approach and the Polya posterior is not surprising when one

notes that in the unconstrained case the sequence of priors leading to the Polya posterior can be used to prove the admissibility of the maximum likelihood estimator for the probability vector of a multinomial distribution.

6. Other sampling designs

All the simulation results presented thus far have used (stratified) simple random sampling without replacement (SRS) as the sampling design. In an earlier version of the manuscript a referee wanted to know how much the behavior of estimators based on the constrained Polya posterior depended on using this design. The answer is there is some dependence but not as much as you might initially believe.

We have seen under SRS that the constrained Polya posterior (CPP) estimator behaves much like the regression estimator (REG). Formally, the regression estimator depends only on knowing the population mean of the auxiliary variable. Its properties are usually studied under simple random sampling and the estimator of its variance is only valid for large samples.

For a general design the Horvitz-Thompson estimator (HT) is often used. It is unbiased but computing the exact inclusion probabilities can be difficult. This is true, for example, if the sampling is done with selection probability proportional to the size for an auxiliary variable x , say $PPS(x)$. In practice one simply assumes that the inclusion probability of a unit is proportional to its value of x and the resulting estimator will be approximately unbiased.

We implemented several simulation studies comparing these three methods for estimating a population total. For brevity, we present the results of two of the studies. In these studies we constructed two populations of size 500. The auxiliary variable is the same in both populations and is a random sample from a gamma distribution with shape parameter 5 and scale parameter 1. Plots of the two populations are given in figure 1. We are not suggesting that in practice one would be likely to use the regression estimator in the second population. It is presented here simply to illustrate what can happen.

For each population we took 400 random samples of size 30 and 60 under two different sampling designs. They were $PPS(x)$ and $PPS(1/(x+5))$. We assumed that the population mean of x was known. For each sample we calculated the three estimates of the population total. The results from the first design are given in table 8. We see that CPP is the clear winner. The HT interval estimator's observed frequency of coverage is closest to the nominal level of 0.95. But the interval length is ridiculously long. This occurs because the reciprocals of the inclusion probabilities vary greatly. For the first population, which is

roughly linear, REG and CPP behave similarly. However, for the second population, CPP performs better than REG. It's only shortcoming here is that it under covers with the smaller sample size. Under $PPS(1/(x+5))$ the story is much the same although the difference between REG and CPP is much smaller for the second population. For example, when the sample size is 30 the average absolute error and frequency of coverage for REG is 131.9 and 0.875 compared to 124.3 and 0.908 for CPP. When the sample size is 60 the numbers for REG are 88.4 and 0.905 compared to 90.1 and 0.958 for CPP. The average length of their intervals are 384 and 560 respectively.

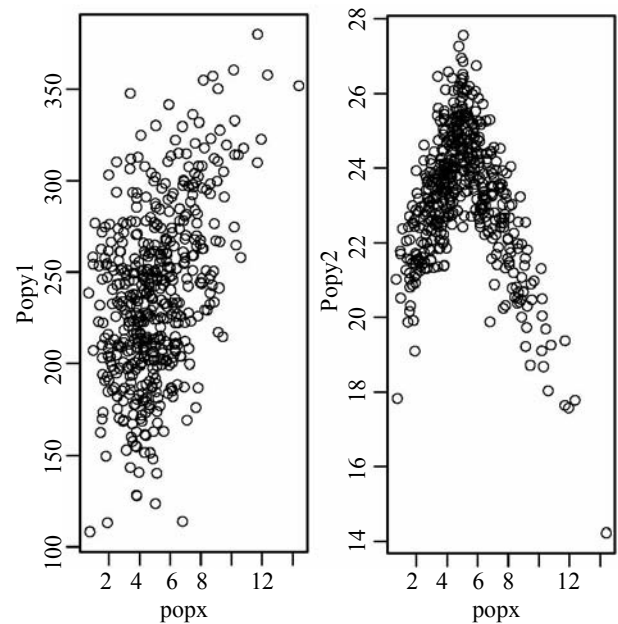


Figure 1 Plots of the two populations used in the simulations in table 8. The correlations for the two populations are 0.47 and -0.22 and their totals are 118,210 and 11,648.7

For the second population we did a second set of simulations using the $PPS(x)$ design for sample sizes of 30 and 60. This time we assumed that the population means of x and x^2 are both known. We then compared the CPP estimator which incorporates constraints on both x and x^2 with the regression estimator which assumes a quadratic function of x as the model. These estimators are denoted by CPP2 and REG2 in the table. At first glance it might seem surprising that the results for CPP and CPP2 are essentially the same. But upon reflection it is what one should expect. The constrained Polya is simulating full copies of the population that are “balanced” with respect to x , that is agree with its known population mean. The additional constraint that a simulated copy of the population must be “balanced” with respect to x^2 as well adds little

information. On the other hand with a sensible model the regression estimator can exploit the additional information. This results in an improved point estimator but its interval estimates still under cover.

Table 8
Simulating results for the two populations in figure 1 when estimating the population total. In each case they are based on 400 samples which were select using PPS(x) as the design. Note abserr is the absolute value of the difference between the estimate and the true population total

Method	Ave. of estimates	Ave. of abserr	Ave of length	Freq. of coverage
For population 1 with total = 118,210.2 for a sample size of 30				
HT	118,803.1	8,095.3	38,696.6	0.905
REG	116,838.1	3,355.3	14,136.4	0.905
CPP	117,515.7	3,277.3	14,330.7	0.905
for a sample size of 60				
HT	119,139.2	5,395.6	28,233.3	0.952
REG	117,041.4	2,213.2	9,561.0	0.910
CPP	118,041.4	2,195.3	11,836.5	0.938
For population 2 with total = 11,648.7 for a sample size of 30				
HT	11,737.2	783.5	4,012.0	0.945
REG	11,800.3	179.7	533.0	0.745
CPP	11,689.9	122.4	535.4	0.900
REG2	11,660.0	97.2	382.3	0.862
CPP2	11,689.9	122.4	535.4	0.900
CPPbd	11,683.2	116.5	537.0	0.918
for a sample size of 60				
HT	11,774.2	564.8	2,908.2	0.955
REG	11,795.8	155.2	373.1	0.635
CPP	11,647.9	80.4	524.4	0.978
REG2	11,663.1	66.7	266.2	0.895
CPP2	11,651.2	88.4	523.6	0.962
CPPbd	11,644.6	83.9	552.1	0.978

For the second population we did a third set of simulations using PPS(x) as the design for sample sizes of 30 and 60. In this case we assumed that the population mean of x was contained in the interval (4.45, 5.53). These are the 0.45 and 0.65 quantiles of the x population. The mean of this population is 5.02. The results are in table 8 under the label CPPbd. We see that the results are very similar to those where the population mean of x was assumed to be known.

All three estimators are using the information contained in the auxiliary variable x but the HT estimator is the only one that depends on knowing the sampling design. As we have noted, it is well known that Bayesian estimators do not use the design probabilities in their computation. In these examples we see that CPP is making effective use of the information contained in the auxiliary variable. In general, the Polya posterior and variations on the Polya posterior,

like the Constrained Polya posterior, do not rely directly upon simple random sampling, stratified random sampling, or any other design. Their suitability and their performance are dependent upon the agreement of the structure underlying the population and the structure specified in the chosen predictive distribution.

The basic idea underlying the CPP is that one should use the sample and the available auxiliary information to simulate complete representative copies of the population. In simple examples like those given above we see that its point estimator should have excellent frequentist properties for a wide class of designs and the performance of its interval estimator will be adequate if the sample size is not too small. Does this mean that it can automatically adjust "bad" samples to get good estimates? Not really since with a very bad sample, one that agrees poorly with the known prior information, two bad things can happen. First, extremely unbalance or biased samples will introduce some bias into the point estimate. Second, they will severely constrain the possible values of p under the CPP and result in a posterior variance that is too small, which will lead to interval estimates that are too short and under cover the quantity of interest. In more complicated situations further study needs to be done to discover when the CPP can profitably be employed.

7. Final remarks

One problem with standard frequentist methods is that each different problem demands its own solution. Estimating the population median of y when the population mean of x is known is a different problem than estimating the mean of y when the mean of x is known. Also, if the population mean of x is not known exactly but is only known to belong to some interval of values then the standard frequentist methods cannot make use of this information. A strength of a Bayesian approach is that once you have a posterior distribution which sensibly combines the sample with the prior information inference can be done for many population parameters of interest simply by simulating completed copies of the population.

Here we have argued that the constrained Polya posterior is a sensible method of introducing objective prior information about auxiliary variables into a noninformative Bayesian approach to finite population sampling. The resulting point estimators have a stepwise Bayes justification which guarantees their admissibility. Their 0.95 credible intervals will usually be approximate 95% confidence intervals and they give sensible answers for problems where there are no standard frequentist procedures available. This demonstrates an important strength of the Polya posterior. Once you can simulate sensible copies of the

entire population inference for a variety of problems becomes straightforward. On the downside, one needs to use MCMC methods for their calculation. All our computations were done in R (R Development Core Team 2005). Two of the authors have recently released an R package `polyapost` which makes it easy for others to use our methods. Here we have restricted ourselves to samples of less than 100. This was just a matter of convenience so we could do our simulations in a reasonable amount of time. In practice for a larger specific sample one just needs to run a longer chain. Then one can use some of the standard diagnostics to decide whether or not it seems to have converged.

Appendix

An admissibility proof

The basic theoretical justification for point estimators arising from the Polya posterior is that are admissible. The proofs of admissibility use the stepwise Bayes nature of the Polya posterior. This section presents a proof for point estimators based on the constrained Polya posterior.

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 each sample point which receives positive probability under the second prior and which was not considered under the first prior. Then, the third prior is considered and so on. For a particular sample point the value of the stepwise Bayes estimate is the value of the Bayes procedure from the step at which it was considered. It is the stepwise Bayes nature of the Polya posterior that explains its somewhat paradoxical properties. Given a sample, it behaves just like a proper Bayesian posterior but one never has to explicitly specify a prior distribution. For more details and discussion on these points see Ghosh and Meeden (1997).

To prove the admissibility of the estimators arising from the Polya posterior for the parameter space $[0, \infty)^N$ the main part of the stepwise Bayes argument first assumes that the parameter space is Λ^N , where Λ is an arbitrary finite set of positive real numbers. Once admissibility has been demonstrated for such general Λ , admissibility for the parameter space $[0, \infty)^N$ follows easily. A similar argument will be used for the constrained Polya posterior.

Dealing with constraints on finite populations introduces some technical problems which are difficult to handle. For this reason, we will suppose that the population is large enough compared to the sample size that the approximate

form of the Polya posterior involving the Dirichlet distribution is appropriate. For simplicity we assume that the population U is infinite.

We assume that for all j in U , $(y_j, X_j) = a_i$ for some i in $\{1, \dots, k\}$, where $a_i = (a_{i1}, \dots, a_{i(m+1)})$ are distinct vectors in R^{m+1} and where k can be very large. That is, the vectors (y_j, X_j) can take on only a finite number of values. If p_i is the proportion of (y_j, X_j) 's in the population which are equal to a_i , for i in $\{1, \dots, k\}$, then the population mean of Y is $\sum_{i=1}^k p_i a_{i1}$.

We assume that there is prior information available about the auxiliary variables $X^i := \{x_j^i \mid j \in U\}$ for i in $\{1, \dots, m\}$, which gives rise to linear equalities and inequalities involving the proportions p of the form

$$A_1 p = b_1 \quad (4)$$

$$A_2 p \leq b_2 \quad (5)$$

where A_1, A_2 are $m_1 \times k$ and $m_2 \times k$ matrices and b_1, b_2 vectors of appropriate dimensions. In this setting, for instance, we may want to estimate

$$\mu(p) = \sum_{i=1}^k p_i a_{i1}$$

subject to the constraints in equations 4 and 5 and where $\sum_{i=1}^k p_i = 1$ with $p_i \geq 0$, for all i in $\{1, \dots, k\}$.

Consider a sample s of size n which for notational convenience we will assume consists of n distinct a_i 's. Let a_s denote this set of values. We then let $A_{1,s}$ and $A_{2,s}$ be the $m_1 \times n$ and $m_2 \times n$ matrices which are just A_1 and A_2 restricted to the columns corresponding to the members of a_s . Let p_s be p restricted to the members of a_s . Then the constraints on the population given in equations 4 and 5 translate into the following constraints

$$A_{1,s} p_s = b_1 \quad (6)$$

$$A_{2,s} p_s \leq b_2 \quad (7)$$

for the random weights assigned to members of the sample. That is, given a sample the constrained Polya posterior is just the uniform distribution over the subset of the simplex defined by equations 6 and 7.

A technical difficulty when proving admissible under constraints is that even when the population satisfies the stated constraints it is always possible to get a sample which fails to satisfy them. There are several ways one can handle such cases. One possibility is to assume that the constraints are wrong and just ignore them. This tactic was used in Nelson and Meeden (1998). Another possibility is to use prior information to augment the sample so that it satisfies the constraints. This can be messy and your answer can depend strongly on how you adjust the sample. We will take a third approach here.

We will assume the sampling design is simple random sampling and that the our prior information must be correct. In such a situation it might make sense to reject any sample which does not satisfy the constraints since it is clearly an unrepresentative sample. More specifically, suppose we take a simple random sample of size n from the population and observe all x_j^i 's in the sample. Let $p^s = (p_1^s, \dots, p_k^s)$ be the proportions of the possible vectors for the x_j that are observed in the sample. The element p_i^s is zero whenever the vector $a_i = (a_{i2}, \dots, a_{i(m+1)})$ does not appear in the sample. If p^s satisfies equations 6 and 7 we keep the sample, if not we discard it and try again. We will call this sampling plan constraint restricted random sampling. In practice, for typical constraints, it will almost never be necessary to discard a sample. Although this is a sampling plan that would never be used it is not a bad approximation to what is actually done.

More formally, let Z_i be the number of (y_j, X_j) 's in the sample that equal a_i , for i in $\{1, \dots, k\}$, then (Z_1, \dots, Z_k) is Multinomial(n, p_1, \dots, p_k) where the parameter values belong to

$$P := \left\{ (p_1, \dots, p_k) \left| \begin{array}{l} A_1 p = b_1, A_2 p \leq b_2, \sum_{i=1}^k p_i = 1, \\ \text{and } p_i \geq 0 \forall i \in \{1, \dots, k\} \end{array} \right. \right\}. \quad (8)$$

For a given sample $s = (z_1, \dots, z_k)$ let

$$P^s := \{p \mid p \in P \text{ and } p_i = 0 \text{ whenever } p_i^s = 0 \text{ for } i=1, \dots, k\}. \quad (9)$$

We see that we keep a sample if and only if P^s is not empty.

Denote the $k - 1$ dimensional simplex by

$$F := \left\{ (p_1, \dots, p_k) \left| \sum_{i=1}^k p_i = 1, p_i \geq 0 \forall i \in \{1, \dots, k\} \right. \right\}.$$

For $i = 1, \dots, k$ let e_i denote the vertices of F . The e_i 's are the unit vectors whose i^{th} value is 1 and is 0 elsewhere.

Now P is a convex polytope which is the intersection of F with the space

$$G := \{(p_1, \dots, p_k) \mid A_1 p = b_1, A_2 p \leq b_2\}.$$

A partition of the parameter space P can be found in the following way. Let F_j denote the set of faces of dimension j of the simplex F , $j = 0, 1, \dots, k - 1$. Then F_0 is the set of its vertices, F_j is the collection of the convex hulls of all combinations of $j + 1$ vertices, for $j = 1, \dots, k - 2$ and F_{k-1} is the simplex F . If $\text{int}(F_j)$ is the set of the interiors of the faces of dimension j , for $j = 1, \dots, k - 1$, then $\{F_0, \text{int}(F_1), \dots, \text{int}(F_{k-1})\}$ determines a partition of the simplex F . If $G_0 := F_0 \cap G$ and $G_j := \text{int}(F_j) \cap G$ for $j = 1, \dots, k - 1$ then $\{G_0, G_1, \dots, G_{k-1}\}$ is a partition of the parameter space P . Note that some of G_j 's might be

empty. The stages of the stepwise Bayes argument follow the nonempty members of the G_j 's.

If Z is the sample space of the counts (Z_1, \dots, Z_k) then for $p \in P$ the distribution of the counts, say $f_p(z \mid p)$ is Multinomial(n, p_1, \dots, p_k) when the sample size is n . Let P_F be the restriction of the parameter space P to F , where F is any subset of P and Z_{P_F} be the restriction of the sample space Z determined by P_F .

We are now ready to prove the admissibility of the constrained Polya posterior estimator of $\mu(p)$ over P . Suppose we are at the stage where we are considering G_j , for some $j \geq 0$. Assume $G_j = G \cap F$ for some subset is nonempty. There are two possible cases.

Case 1. If the dimension of G_j is zero, i.e., it consists of one vector, say p^0 , then we take the prior that puts unit mass on this vector. The posterior also then puts unit mass on this vector and if z is the unique member of Z_{P_F} then the Bayes estimator is $\delta_{\pi_{P_F}}(z) = E(\mu(p) \mid z) = \mu(p^0)$.

Case 2. If the dimension of G_j is greater than zero then the distribution of (Z_1, \dots, Z_k) restricted to Z_{P_F} is

$$f_{P_F}(z \mid p) = \frac{f_p(z \mid p)}{\sum_{z \in Z_{P_F}} f_p(z \mid p)}.$$

The prior we consider on P_F is

$$\pi_{P_F}(p) \frac{\sum_{z \in Z_{P_F}} f_p(z \mid p)}{\prod_{\{i \mid p \in P_F, p_i > 0\}} p_i},$$

which can be normalized to be a proper prior since $\sum_{z \in Z_{P_F}} f_p(z \mid p)$ can be written as $g(p) \prod_{\{i \mid p \in P_F, p_i > 0\}} p_i$ where $g(p)$ is a bounded function of p . With this prior, the posterior distribution is the Dirichlet density kernel restricted to P_F ,

$$f_{P_F}(p \mid z) \propto f_p(z \mid p) \pi_{P_F}(p) \propto \prod_{\{i \mid p \in P_F, p_i > 0\}} p_i^{z_i - 1}.$$

The Bayes estimator of $\mu(p)$ against π_{P_F} , where p belongs to P_F , is $\delta_{\pi_{P_F}}(z) = E(\mu(p) \mid z)$ for all z in Z_{P_F} . Hence, if we use the sequence of priors

$$\{\pi_{P_F \mid F \in G_0}\}, \{\pi_{P_F \mid F \in G_1}\}, \{\pi_{P_F \mid F \in G_2}\}, \dots, \{\pi_{P_F \mid F \in G_{j-1}}\},$$

ignoring the empty sets at each step, then the estimator $\delta(z)$ defined by

$$\delta(z) = \delta_{\pi_{P_F}}(z) \text{ for } z \in Z_{P_F}, F \in G_i, i = 1, \dots, \gamma - 1, \quad (10)$$

where $\gamma = k$ if $k < n$ and n if $k \geq n$, is an admissible estimator for $\mu(p)$. This concludes the proof of the following theorem.

Theorem 1. Under the constraint restricted random sampling plan defined by equations 2 and 3 with parameter

space defined in equation 8 the constrained Polya posterior estimator given in equation 10 for estimating the population mean is stepwise Bayes and hence admissible under squared error loss.

Acknowledgements

The research of Glen Meeden and Radu Lazar was supported in part by NSF Grant DMS 0406169. The authors would like to thank the referees and associate editor for several helpful comments.

References

- Binder, D.A. (1982). Non-parametric Bayesian models for samples from finite populations. *Journal of the Royal Statistical Society, Series B*, 44, 388-393.
- Booth, J.G., Bulter, R.W. and Hall, P. (1994). Bootstrap methods for finite population sampling. *Journal of the American Statistical Association*, 89, 1282-1289.
- Chambers, R.L., and Dunstan, R. (1986). Estimating distribution functions from survey data. *Biometrika*, 73, 597-604.
- Chen, J., and Sitter, S.S. (1999). A pseudo empirical likelihood approach to the effective use of auxiliary information in complex surveys. *Statistica Sinica*, 9, 385-406.
- Chen, J., and Qin, J. (1993). Empirical likelihood estimation for finite populations and the effective usage of auxiliary information. *Biometrika*, 80, 107-116.
- Cochran, W. (1976). *Sampling Techniques*. New York: John Wiley & Sons, Inc., 3rd Edition.
- Deville, J.-C., and Särndal, C.-E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87, 376-382.
- Efron, B., and Tibshirani, R. (1993). *An Introduction to the Bootstrap*. Chapman and Hall, London.
- Feller, W. (1968). *An Introduction of Probability Theory and its Applications*. New York: John Wiley & Sons, Inc., Volume I.
- Ghosh, J.K. (1988). *Statistical Information and Likelihood: A Collection of Critical Essays*. By (Ed. D. Basu). New York: Springer-Verlag.
- Ghosh, M., and Meeden, G. (1997). *Bayesian Methods for Finite Population Sampling*. London: Chapman and Hall.
- Gross, S. (1980). Median estimation in survey sampling. In *Proceedings of the Section of Section on Survey Research Methods*, American Statistical Association, 181-184.
- Hartley, H.O., and Rao, J.N.K. (1968). A new estimation theory for sample surveys. *Biometrika*, 55, 159-167.
- Kazis, L.E., Miller, D.R., Clark, J., Skinner, K., Lee, A., Rogers, W., Spiro, A. 3rd, Payne, S., Fincke, G., Selim, A. and Linzer, M. (1998). Health related quality of life in patients served by the department of veterans affairs. *Archives of Internal Medicine*, 158, 626-632.
- Kuk, A.Y.C., and Mak, T.K. (1989). Median estimation in the presence of auxiliary information. *Journal of the Royal Statistical Society, Series B*, 51, 261-269.
- Lo, A. (1988). A Bayesian bootstrap for a finite population. *Annals of Statistics*, 16, 1684-1695.
- Mak, T.K., and Kuk, A. (1993). A new method for estimating finite population quantiles using auxiliary information. *Canadian Journal of Statistics*, 21, 29-38.
- Meeden, G. (1999). A noninformative Bayesian approach for two-stage cluster sampling. *Sankhyā, Series A*, 61, 133-144.
- Meeden, G. (2005). A noninformative bayesian approach to domain estimation. *Journal of Statistical Planning and Inference*, 129, 85-92.
- Nelson, D., and Meeden, G. (1998). Using prior information about population quantiles in finite population sampling. *Sankhyā, Series A*, 60, 426-445.
- Nelson, D., and Meeden, G. (2006). Noninformative nonparametric finite population quantile estimation. *Journal of Statistical planning and Inference*, 136, 53-67.
- R Development Core Team (2005). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- Rao, J.N.K., Kovar, J.G. and Mantel, H.J. (1990). On estimating distribution functions and quantiles from survey data using auxiliary information. *Biometrika*, 77, 365-375.
- Rubin, D. (1981). The Bayesian bootstrap. *Annals of Statistics*, 9, 130-134.
- Singh, J.A., Borowsky, S.J., Nugent, S., Murdoch, M., Zhao, Y., Nelson, D., Petzel, R. and Nichol, K.L. (2005). Health related quality of life, functional impairment, and health care utilization by veterans: Veterans' quality of life study. *Journal of American Geriatric Society*, 53, 108-113.
- Smith, R.L. (1984). Efficient Monte Carlo procedures for generating points uniformly distributed over bounded regions. *Operations Research*, 32, 1296-1308.
- Valliant, R., Dorfman, A. and Royall, R. (2000). *Finite Population Sampling and Inference, a Predictive Approach*. New York: John Wiley & Sons, Inc.
- Vardeman, S., and Meeden, G. (1984). Admissible estimators for the total of a stratified population that employ prior information. *Annals of Statistics*, 12, 675-684.
- Woodruff, R.S. (1952). Confidence intervals for the median and other positive measures. *Journal of the American Statistical Association*, 47, 635-646.
- Zhong, B., and Rao, J.N.K. (2000). Empirical likelihood inference under stratified random sampling using auxiliary population information. *Biometrika*, 87, 920-938.