Interval estimators for the population mean for skewed distributions with a small sample size

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

July 1997

*Research supported in part by NSF Grant SES 9201718

Interval estimators for the population mean for skewed distributions with a small sample size

SUMMARY

In finite population sampling it has long been known that for small sample sizes, when sampling from a skewed population, the usual frequentist intervals for the population mean cover the true value less often than their stated frequency of coverage. Recently a noninformative Bayesian approach to some problems in finite population sampling has been developed which is based on the 'Polya posterior'. For large sample sizes these methods often closely mimic standard frequentist methods. In this note a modification of the 'Polya posterior' which employs the weighted Polya distribution is shown to give interval estimators with improved coverage properties for problems with skewed populations and small sample sizes. This approach also yields improved tests for hypotheses about the mean of a skewed distribution.

Key Words: Polya posterior, estimating a mean, finite population sampling, interval estimation, hypotheses testing and skewed distributions.

AMS 1991 Subject Classification: Primary 62D05; Secondary 62F15, 62F25.

1 Introduction

It has long been known that when sampling from a skewed population, with small sample sizes, the usual frequentist confidence intervals for the population mean have poor coverage properties. In Johnson (1978) a modification of the usual t confidence interval was proposed for asymmetrical populations. He mainly considered the linear term from a Cornish-Fisher approximation to the numerator of the t statistic. Kleijnen et al. (1986) noted that this often could have little affect and studied in more detailed what could happen if the quadratic term in the approximation was used as well. Recently in Sutton (1993) it was noted that for testing problems when the skewness is severe and the sample size is small that Johnson's test can also be appreciably inaccurate. Chen (1995) gives another modification of the t statistic based on the Edgeworth expansion which improves on the earlier modifications. For the finite population version of this problem it was pointed out in section 2.15 of Cochran (1977) that the usual 95% confidence interval for the mean, which is based on normal theory, will cover less that 95% of the time. Moreover for populations which are skewed to the right, intervals which do not contain the true mean are more likely to lie to the left of the true mean than to the right. Hence to improve the coverage probability one needs to find an interval which is shifted to the right and perhaps longer as well. The amount of "adjustment" clearly will depend on the sample size and the amount of skewness present in the population. For large sample sizes it was seen in Meeden and Vardeman (1991) that for interval estimation of the mean the standard interval behaves much like Bayesian credible intervals for the mean based on the 'Polya posterior'.

The 'Polya posterior' is a noninformative approach to finite population sampling. It is appropriate when little is known about the population except that one's prior beliefs about the units in the population are roughly exchangeable. For large sample sizes, i.e. greater than thirty, it will yield inferences very similar to standard frequentist inferences based on simple random sampling without replacement. However the design probabilities play no formal role in inferences based on the 'Polya posterior'. For further discussion of the 'Polya posterior' see Ghosh and Meeden (1997). In this note we will introduce an extension of the 'Polya posterior', the weighted Polya posterior. It will be shown how credible intervals and tests of hypotheses based on the weighted Polya posterior will have reasonable frequentist properties. In section 2 the weighted Polya posterior will be introduced and discussed.

In section 3 some simulation studies will be introduced to see how it works in practice for the interval estimation problem. In section 4 the hypotheses testing problem will be considered.

2 The weighted Polya posterior

2.1 Notation

Let $\mathcal{U}$ denote a finite population which consists of N units labeled $1, 2, \ldots, N$. Attached to unit i let y_i be the unknown value of some characteristic of interest. In this note we will only consider the simplest situation where y_i is a real number. For this problem $\mathbf{y} = (y_1, \ldots, y_N)^T$ is the unknown parameter and is assumed to belong to to $\mathcal{Y}$ a subset of N-dimensional Euclidean space. In what follows we will sometimes assume that $\mathcal{Y}$ is of a particular special form. If $\mathbf{b} = (b_1, \ldots, b_k)^T$ is a k-dimensional vector of distinct real numbers then we let

$$\begin{aligned} \mathcal{Y}(\mathbf{b}) \;=\; & \{\, \mathbf{y} : \text{such that for } i = 1, \ldots, N \; y_i = b_j \\ & \text{for some } j = 1, \ldots, k \,\} \end{aligned} \tag{2.1}$$

Let n be a fixed positive integer lying between 4 and N. A subset s of $\{\, 1, 2, \ldots, N \,\}$ of size n will be a sample. We assume the sample was selected by simple random sampling without replacement. Given a sample $s = \{\, i_1, \ldots, i_n \,\}$, the labels of the units in the observed sample, let $z_s = (\, z_{i_1}, \ldots, z_{i_n} \,)^T$ be the observed values of the characteristic of interest for the units in the sample. That is z_{i_j} is the observed value for unit i_j. We will denote a typical data point by the pair $(\, s, z_s \,)$.

2.2 Definition

With this notation in hand we can now easily describe 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 $(\, s, z_s \,)$ we construct this distribution as follows. We consider an urn that contains n balls, where ball one is given the value z_{i_1}, ball two the value z_{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 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 created a predictive distribution for the unseen given the seen. A good reference for Polya sampling is Feller (1968). In addition, it can be shown that the 'Polya Posterior' has a stepwise Bayes justification so that for many problems, decision procedures based on it will be admissible. For further discussion see Ghosh and Meeden (1997).

In the construction of the 'Polya posterior' each ball in the original collection of the n balls in the urn, which represented the sample, was given weight one. Also each additional ball added to the urn during process of generating a simulated copy of the entire population was also given weight one. For defining a sampling strategy from the urn, this is not really necessary and as is well known any collection of nonnegative weights assigned to the original n balls in the urn could do as well. However finding collection of weights which can be given a stepwise Bayes justification is more difficult. With this in mind we now introduce the weighted Polya posterior which is a simple modification of the 'Polya posterior'. Let $\mathbf{b} = (b_1, \ldots, b_k)^T$ be the k distinct values that appear in z_s and for $i = 1, \ldots, k$ let n_i be the number of z_j in z_s which equal b_i. Therefore $n_i \geq 1$ and $\sum_{i=1}^{k} n_i = n$. Let $\alpha < 1$ be fixed. Now when we construct the original configuration of balls in the urn to represent the observed sample each ball with the value b_i is given the weight $1 - \alpha/n_i$. Hence the total weight associated with the n original balls is $n - k\alpha$, not n as in the earlier case.

Next we consider the predictive distribution of the unseen given the seen. As before a ball is picked at random, following the distribution defined by the weights, from the urn and its value is assigned to the unobserved unit with the smallest label. Then it is returned to the urn along with another ball with the same value and weight <u>one</u>. A second ball is picked from the urn containing now $n+1$ balls following the new distribution of weights and its value is assigned to the unobserved unit with the second smallest label. This second ball is returned to the urn along with another ball of weight one, with the same value. This is repeated until all $N-n$ unseen units have been assigned values.This process produces one possible realization of the entire population from the weighted Polya posterior.

Note that the distribution of the unseen given the seen under the weighted Polya posterior is exchangeable. As in the above, for a given data point (s, z_s), let $\mathbf{b}$ be the vector of distinct values that appear in z_s and n_i the number of times that b_i appears. Then the probability that any unseen unit takes on the value b_i is just $(n_i - \alpha)/(n - k\alpha)$. This makes it easy to calculate the conditional expectation of the population mean, say $\mu = \mu(y)$, since

$$\begin{aligned} E(\mu \mid (s, z_s)) &= \frac{1}{N}\left\{\sum_{i \in s} z_i + (N-n)\sum_{j=1}^{k} b_j \frac{n_j - \alpha}{n - k\alpha}\right\} \\ &= \bar{z}_s + \frac{k\alpha(N-n)}{(n-k\alpha)N}\left\{\bar{z}_s - \bar{b}\right\} \end{aligned} \tag{2.2}$$

where $\bar{z}_s$ denotes the mean of the observed data (s, z_s) and $\bar{b} = \sum_{j=1}^{k}(b_i/k)$. Note that when $\alpha = 0$ the weighted Polya posterior becomes just the regular 'Polya posterior' and the above expectation becomes just $\bar{z}_s$. In one other important special case the above expectation is $\bar{z}_s$, that is when all the values of the characteristic in the sample are distinct. This is true for any value of $\alpha < 1$ since in this case $k = n$ and $n_i = 1$ for all i.

Next we find the conditional variance of the population mean under the weighted Polya posterior. This is found by a straightforward but somewhat more involved calculation.

$$\begin{aligned} \mathrm{Var}(\mu|(s, z_s)) &= \frac{1}{N^2}\mathrm{Var}(\sum_{j \notin s} y_j|(s, z_s)) \\ &= \frac{1}{N^2}\Big\{\sum_{j \notin s}\mathrm{Var}(y_j|(s, z_s)) \\ &\quad + \sum_{i<j\,:\, i \notin s \text{ and } j \notin s} 2\mathrm{Cov}((y_i, y_j)|(s, z_s))\Big\} \\ &= \frac{1}{N^2}\Big\{(N-n)\mathrm{Var}(y_1|(s, z_s)) + 2\binom{N-n}{2}\mathrm{Cov}((y_1, y_2)|(s, z_s))\Big\} \end{aligned}$$

where for notational convenience we assume that neither unit 1 nor unit 2 appeared in the sample.

Now $\mathrm{Var}(y_1|(s, z_s))$ is just the variance of the original configuration based on the sample and the value of α for the distribution of balls placed in the urn which defines the weighted Polya. That is, this variance is given by

$$V_{b,\alpha} = (n-k\alpha)^{-1}\sum_{i=1}^{k} b_i^2(1-\alpha/n_i) - \Big((n-k\alpha)^{-1}\sum_{i=1}^{k} b_i(1-\alpha/n_i)\Big)^2 \quad (2.3)$$

The next step is to find the covariance term.

$$\begin{aligned}
\mathrm{Cov}((y_1,y_2)|(s,z_s)) &= E(y_1y_2|(s,z_s)) - E(y_1|(s,z_s))^2 \\
&= E(y_1E(y_1y_2|y_1,(s,z_s))|(s,z_s)) - E(y_1|(s,z_s))^2 \\
&= E(y_1\frac{y_1+\sum_{i=1}^{k}(1-\alpha/n_i)b_i}{n-k\alpha}|(s,z_s)) \\
&\qquad\qquad - E(y_1|(s,z_s))^2 \\
&= (n-k\alpha)^{-1}V_{b,\alpha}
\end{aligned}$$

where the last equation follows from some easy algebra. With these expressions for the variance and covariance terms under the weighted Polya some more easy algebra yields

$$V(\mu \mid (s,z_s)) = \frac{N-n}{N^2}\frac{N-k\alpha}{1+n-k\alpha}V_{b,\alpha} \quad (2.4)$$

Let $Var(z_s)$ be the variance of the observed sample z_s, with divisor $n-1$. In the special case when all the values in the sample are distinct, i.e. when $k=n$, we have $V_{b,\alpha} = ((n-1)/n)\mathrm{Var}(z_s)$ and so the conditional variance becomes

$$V(\mu \mid (s,z_s)) = (1-f)\frac{\mathrm{Var}(z_s)}{n}\frac{n-1}{1+n(1-\alpha)}\frac{N-n\alpha}{N} \quad (2.5)$$

where $f = n/N$ is the sampling fraction. Perhaps it needs to be emphasized, that the above variance is conditioned on the observed sample and the design probabilities play no role in its computation.

In the special case when all the values in the sample are distinct and $\alpha = 0$ then the weighted Polya becomes just the usual Polya posterior. In this case the posterior variance of μ is just $(1-f)(\mathrm{Var}(z_s)/n)([n-1]/[n+1])$. Except for the factor $[n-1]/[n+1]$ this is just the estimated variance of the sample mean under the usual normal theory and simple random sampling. It also indicates why for n large the Bayesian credible interval under the Polya posterior agrees closely with the standard confidence interval. Now the value of α which makes the product of the last two factors in equation 2.5 equal to one, i.e. the value of α which makes the posterior variance of the population mean under the weighted Polya equal to the estimated variance of the sample mean under the usual theory, is $2N/(n[N-n+1])$. This value is always less than one whenever $2 < n < N$. This shows, that in the special case when all the values in the population are distinct and the sample size is greater than two, there exists a weighted Polya such that credible intervals based on it will be similar to the standard intervals. Although this fact is of limited interest it does suggest that an approach based on the weighted Polya might be of some use in problems where the standard methods perform poorly.

But before discussing the implications of these formulas for interval estimation we will give a theoretical justification for the weighted posterior.

2.3 Theoretical Justification

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. This gives a theoretical justification for the 'Polya posterior'. See for example Meeden and Ghosh (1983), Vardeman and Meeden (1984) and Ghosh and Meeden (1997). A similar justification holds for the weighted Polya posterior.

In these stepwise Bayes arguments a finite sequence of disjoint subsets of the parameter space is selected, where the order of the specified subsets 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. Next the Bayes procedure is found for the sample points with positive probability under the third prior and which had not been considered in the first two stages. This process is continued over each subset of the sequence in the order given. If the sequences of subsets and priors are such that a procedure is defined at every sample point of the sample space then the resulting procedure is admissible. To prove the admissibility of a given procedure one must select the sequence of subsets, their order, and the sequence of priors appropriately. The argument for the weighted Polya posterior is the same as the one for the regular 'Polya posterior' except that at each step, but the first, the prior distribution is modified slightly.

As before, for a given data point (s, z_s), let $\mathbf{b} = (b_1, \ldots, b_k)^T$ be the vector of distinct values that appear in z_s and n_i be the number of times that b_i appears. Let $\mathbf{b}^o$ be any vector of distinct real numbers of length at least k whose values contain all the values of $\mathbf{b}$. Assume we are proving the admissibility of procedures based on the weighted Polya posterior when the parameter space is $\mathcal{Y}(\mathbf{b}^o)$. In the stepwise Bayes argument, the data point (s, z_s) is accounted for in the step when the subset of the parameter space is $\mathcal{Y}^*(\mathbf{b})$. This is the subset of $\mathcal{Y}(\mathbf{b})$ which consist of all vectors of length N where each b_i for $i = 1, \ldots, k$ appears at least once. For $\mathbf{y} \in \mathcal{Y}^*(\mathbf{b})$ let $c_y(i)$ be the number of b_i's that appear in $\mathbf{y}$. On this set the prior is proportional to

$$\int_0^1 \cdots \int_0^1 \{\prod_{i=1}^{k-1} \theta_i^{c_y(i)-\alpha-1}\}(1 - \sum_{i=1}^{k-1} \theta_i)^{c_y(k)-\alpha-1}\, d\theta_1 \cdots d\theta_{k-1} \tag{2.6}$$

When $\alpha = 0$ this prior becomes the prior used for proving admissibility for the 'Polya posterior' case.

In the 'Polya posterior' case the joint prior distribution is exchangeable since it is a mixture of independent and identically distributed random variables and it gives weight one to each unit in the sample, when constructing the urn. In the weighted Polya posterior the joint prior distribution is still exchangeable, however the weight assigned to a unit in the sample whose value is b_i is $1 - \alpha/n_i$ where n_i is the number of units in the sample whose value is b_i. Hence when constructing the urn for the weighted Polya posterior the weight associated with b_i is $n_i - \alpha$. So α can be thought of as an adjustment to each value that appears in the sample rather than to each unit that appears. The adjustment is the same for each value, no matter how many times it actually appeared in the sample. The idea of using weights, associated with units in the sample, to define estimators is an old one in finite population sampling. Meeden and Ghosh (1981) investigated when such estimators could be given a stepwise Bayes justification. Unfortunately this seems to be a difficult problem and few such estimators were found. The weighted Polya posterior is certainly in the spirit of such estimators even though the weights depend on the sample in a somewhat surprising way. In what follows we will argue that predictive distribution for the unseen given the seen leads to a useful interval estimator for the population mean when the population is skewed and the sample size is small.

2.4 Interval Estimation

As was noted in the introduction, the usual frequentist interval will under-cover for populations which are skewed. For simplicity we will always assumed that it is known a priori that the population is skewed to the right. Moreover when the standard interval fails to cover the true population mean it will more often be to the left of the mean than to the right. Hence an improved interval would need to be shifted to the right and perhaps lengthen as well. This in fact is what the interval due to Johnson (1978) does.

Following Barndorff-Nielsen and Cox (1989) let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables, with $E(X_1) = \mu$, $\mathrm{Var}(X_1) = \sigma^2$ and $E(X_1 - \mu)^3 = \mu_3$. Let $\rho_3 = \mu_3/\sigma^3$ be the standardized third cumulant. If

$$S_n^* = (\sum_{i=1}^{n} X_i - n\mu)/(\sigma\sqrt{n})$$

the Cornish-Fisher inversion of the the Edgeworth expansion yields

$$S_n^* = Z + \frac{1}{6\sqrt{n}}(Z^2 - 1)\rho_3 + O_p(n^{-1}) \tag{2.7}$$

where Z has the standard normal distribution. Using this approximation Johnson suggested as a $(1-\alpha)\%$ confidence interval for μ the interval

$$[\bar{x} + \hat{\mu}_3/6\hat{\sigma}^2 n] \pm t_{\alpha/2}\hat{\sigma}\sqrt{n} \tag{2.8}$$

This interval is just based on the linear term of the above approximation. Kleijnen et al. (1986) investigated what would happen if the quadratic term in the approximation was used as well. In some cases this led to a confidence set consisting of two disjoint intervals. One of which contained $\bar{x}$ and another which was further to the right. They argued that this did not make sense statistically and so they just eliminated the interval which did not contain $\bar{x}$. A further complication is that in some other cases the minimum value of the quadratic function of μ that needed to be considered was to large and no solution for the upper end point of the confidence interval could be found. For this problem they suggested two heuristic solutions. In their simulations they compared the usual t interval, Johnson's interval and three of their intervals. In our study will we use their last interval which they called procedure (5) and argued that it performed better than the others.

In this note we are interested in finding an interval estimator of the population mean for skewed distributions when the sample size is small. As before let $\mu(\mathbf{y})$ denote the parameter of interest, the mean of $\mathbf{y}$. Let $\alpha < 1$ be fixed and suppose we wish to find .95 Bayesian credible interval for $\mu(\mathbf{y})$ using the weighted Polya posterior based on α. For a given sample such intervals can only be found approximately by simulating values from the induced distribution of $\mu(\mathbf{y})$ under the weighted Polya distribution defined by the observed sample data and α. This is done as follows. For a given (s, z_s) and α we simulate an entire copy of the population using the weighted Polya posterior. We then calculate the value of $\mu(\cdot)$ for this simulated copy of the entire population. This process is then repeated R times, usually $R = 1,000$ is sufficient, resulting in R simulated values of $\mu(\cdot)$. For this simulated population of values we let $q(.025)$ and $q(.975)$ be the .025 quantile and .975 quantile respectively. Then $(q(.025), q(.975))$ will be our announced set estimate and it will always have approximate "posterior probability" .95 under the weighted Polya posterior. This interval is easy to compute and in keeping with standard Bayesian practice, see Berger (1985). For the most part we will limit ourselves to sets with posterior coverage probability of .95, since this seems to be the popular value of a nominal confidence level in practice.

For a given value of α and a fixed population the relative frequency that the above .95 credible interval will contain the true population mean can also be investigated by simulation. There is no reason however to expected that this relative frequency will be .95 and in fact it will almost always be something else. However for fixed population it is possible by simulation to find a value of α which gives .95 credible intervals which contain the true mean approximately 95% of the time.

In what follows we will only consider populations whose values are all distinct. Hence we will restrict attention to samples with distinct values as well. In this case, we saw in equation 2.2 that the "posterior expectation" under the weighted Polya posterior was just the sample mean. In equation 2.5 we saw that the "posterior variance" was just the estimated variance of the sample mean, associated with the normal theory for simple random sampling, multiplied by the factor

$$\frac{n-1}{1+n(1-\alpha)}\frac{N-n\alpha}{N} \tag{2.9}$$

Note that on the interval $(-\infty, 1)$, this factor, is an increasing function of α for fixed n and N. In addition, as will be seen in the simulations, the intervals based on the weighted Polya also tend to be naturally shifted to the right. These two facts suggest that if α can be selected correctly a 95%

Bayesian credible interval should have good frequency coverage properties as well. This should become clear in the next section where some examples are considered.

2.5 Some Example Populations

We will compare the interval estimators using five different populations. The first are two artificial populations and the last three are actual populations. The first population, *ppgamma1.5* was a random sample of size 300 from a gamma distribution with shape parameter 1.5 and scale parameter 1. The second, *ppln* was a random sample of size 500 from a log-normal distribution with mean and standard deviation (of the log) 4.9 and .596 respectively. The third, *ppcounties*, is the 1960 population of a group of 304 American counties. The fourth, *ppsales*, is the total 1970 sales, in billions of dollars, for a group of 331 large corporations. The fifth, *ppcities*, is the the 1970 population, in millions, for a group of 125 American cities. These last three populations were discussed in Royall and Cumberland (1981). These populations were selected because they represent examples of skewed populations that one could expect to see in practice.

In what follows we will need to have a measure of skewness for a population. Various measures have been proposed and they all have the problem of being difficult to interpret. We considered three possible measures. The first is just ρ_3 which has a long history as a measure of skewness. Let F denote the distribution function of a random variable X with m and μ its median and mean respectively. The second measure is actually a family of measures depending on the parameter $\lambda \in (0, 1/2)$ and is given by

$$b_\lambda = [F^{-1}(1-\lambda) + F^{-1}(\lambda) - 2m]/[F^{-1}(1-\lambda) - F^{-1}(\lambda)]$$

For the case $\lambda = 1/4$ it was introduced by Bowley. The last measure is just an integrated version of b_λ where both the numerator and denominator are integrated on $(0, 1/2)$ and is given by

$$\gamma = (\mu - m)/E|X - m|$$

For populations which are skewed to the right all three measures are strictly positive and the last two bounded above by one. For a further discussion of skewness measures see Groenveld and Meeden (1984). The basic information about the populations is summarized in Table 1.

Unfortunately none of these skewness measures are satisfactroy for our problem. The difficulty is that for small sample sizes the natural sample estimators of these quantities are all biased. Furthermore the bias increases with the skewness of the population. Recently Arnold and Groeneveld (1995) have introduced a new measure which is based on the mode of the distribution. For a lognormal distribution whose standard deviation of its log is *sdlog* the measure is given by

$$\gamma_M = 2\Phi(sdlog) - 1$$

where $\Phi(\cdot)$ is the distribution function of the standard normal distribution. For sampling from a finite population the estimator $2\Phi(\hat{sdlog}) - 1$ is nearly unbiased for the analogous finite population quantity where $\hat{sdlog}$ is the standard deviation of the log of the observations. For this reason we will use γ_M as our measure of skewness in what follows.

Place Table 1 about here

To see how the weighted Polya posterior works in practice we report some simulation results for population *ppgamma1.5* in Table 2. For two sample sizes, ten and twenty we took 500 simple random samples and computed the usual 95% normal theory confidence interval for the mean, denoted by tI, and compared it to a weighted Polya posterior .95 credible interval for the mean denoted by wtppI. For each sample size we repeated this experiment three times with three different weights for the weighted Polya posterior. First of all it is clear from Table 2 that even though the population *ppgamma1.5* is only mildly skewed the usual normal theory interval under covers, especially when the sample size is ten. We also see, for both sample sizes, that the weighted Polya posterior interval is longer than the usual interval. This is to be expected from the variance formula in equation 2.5. Moreover the average length increases as α increases. Finally, the average centers of the weighted Polya posterior intervals are, as α increases, 1.60, 1.61 and 1.66 when the sample size is ten and 1.50, 1.52 and 1.55 when the sample size is twenty. In short, in terms of frequency of coverage, in all six cases the weighted Polya interval does at least as well as the usual interval. This happens because the weighted Polya posterior intervals are longer than the usual intervals and shifted to the right.

Place Table 2 about here

Recall that for populations where all the units have distinct values the mean of the weighted Polya posterior distribution is just the sample mean, for all choices of α. Hence in these cases the posterior mean will be an unbiased estimator of the population mean. Therefore, the weighted Polya posterior is using the shape of the sample to automatically yield an interval which is shifted to the right of the standard interval and is longer than the standard interval. This results in an improved frequency of coverage for the weighted Polya posterior. From Table 2 we see for population *ppgamma1.5* we should use a weighted Polya posterior with $\alpha = .7$ when the sample size is ten and $\alpha = .5$ when the sample size is twenty to get .95 credible intervals which are also approximate 95% confidence intervals.

Unfortunately, the choice of α which makes a .95 credible interval based on the weighted Polya posterior also be an approximate 95% confidence interval depends on both the population and the sample size. For example, for population *ppcounties* similar simulation studies show that $\alpha = .85$ and $\alpha = .7$ are reasonable choices when the sample sizes are ten and twenty respectively. The values of α are closer to one for this population because it is more skewed than the first population. From a Bayesian perspective one could argue that the choice of a value for α should be subjective, reflecting one's prior beliefs about how skewed the population under study is and one's knowledge about how a proper choice of α is related to the skewness of the population and the sample sized being used. This seems to be quite difficult to do theoretically since it is not clear how to formally relate the skewness of the population, and the sample size to a sensible choice of α. This will be discussed further in the next section.

3 Approximate Solution for the Mean

As we have seen, when estimating the population mean, the value of α for the weighted Polya posterior such that an approximate .95 credible interval is also an approximate 95% confidence interval depends both on the population and the sample size. One way to attempt to overcome this difficulty is as follows.

Select a family of skewed distributions which represent populations which occur in practice. Select a measure of skewness and select a few members of the family which take on typical values of the skewness measure. For these few distributions and a few sample sizes find values of α for which approximate .95 credible intervals under the weighted Polya posterior are also approximate 95% confidence intervals. Using these α's, sample sizes and skewness measures, find a function of sample size and skewness measure which yields a proper choice of α. If the population under study can be approximated reasonably well by a member of our selected family then this function can be used to make a subjective selection of α if prior information about the skewness of the population is at hand. If such information is not available then one can apply the function to the sample. That is, given a sample from the population, one first calculates the skewness measure of the sample. Assuming this estimate is the true value of the skewness measure of the population, one uses this estimate, the sample size and the above function to find the proper choice of α. Finally, one uses this value of α in the weighted Polya posterior to find, hopefully, an approximate 95% confidence interval for the mean. This program has been carried out and generally improves on the other methods however it still has some problems. The rest of this section describes what was done, points out some of the difficulties and shows how a modification of the above program leads to reasonable solutions.

The family of distributions selected was the lognormal distribution. It has long been known that many skewed populations that arise in practice are approximately lognormal. As noted before we will use the skewness measure γ_M. Clearly any measure of skewness will work better for some populations than others. But since we know of no analytic relationship between any measure of skewness and the behavior of the weighted Polya posterior we selected this measure because it worked best in practice.

The next step is to select some typical members of the family. The family is indexed by two parameters $mlog$ and $sdlog$ the mean and the standard deviation of the log of the distribution. For a fixed value of $sdlog$ the values of γ_M does not depend on the choice of $mlog$. So we assumed $mlog = 0$. We selected five values for $sdlog$. They were 0.3, 0.6, 0.9, 1.2 and 1.5. The corresponding values of γ_M are 0.24, 0.45, 0.63, 0.77 and 0.87. Populations with larger values of γ_M are quite skewed and as we shall see our approach will not work for such skewed populations and the small sample sizes considered here.

For these five distributions we are now ready to find, for various sample sizes, the proper value of α, i.e. one that yields an approximate 95% confidence interval. For a given population and sample size once can find through preliminary study an interval where the solution should lie. For a given α in the interval one could then select 1,000 random samples of the given size and find the frequency of coverage for the weighted Polya intervals. If the frequency was significantly less than .95 one should repeat the process with a value of α closer to one. On the other hand if the frequency was greater than .95 one should select a smaller value of α to repeat the process. By repeating this several times we can find an approximate solution for α. This is inefficient however and the sampling variability makes it difficult to know when the solution has been found approximately. A better way is to use some ideas of Geyer (1994).

Suppose the solution is known to be in the interval $[.21, .29]$. Then we select four possible values of α say $\alpha_1 = .22$, $\alpha_2 = .24$, $\alpha_3 = .26$ and $\alpha_4 = .28$. For a fixed observed sample (s, z_s) and fixed α let $f_\alpha(x)$ denote the probability function under the weighted Polya posterior, where x denotes a realization of the entire population from this distribution. We then observe 1,000 realizations of x from the distribution

$$f_{mix}(x) = \frac{1}{4}\sum_{i=1}^{4} f_{\alpha_i}(x)$$

Let $\alpha_0 \in (.21, .29)$ be fixed. Then by the usual importance sampling formula

$$E_{\alpha_0} g(X) \doteq \frac{1}{1000}\sum_x g(x)\frac{f_{\alpha_0}(x)}{f_{mix}(x)}$$

for any real valued function $g(\cdot)$ of x. In the same way we can find any quantile of the distribution of $g(X)$ under $f_{\alpha_0}(\cdot)$ approximately by using the distribution which is proportional to the weights $f_{\alpha_0}(x)/f_{mix}(x)$. In this way, for each observed sample (s, z_s) we can find approximately the .95 credible interval under the weighted Polya posterior defined by α_0. If this is done for 1,000 samples (s, z_s) we can find, approximately, the frequency of coverage of these intervals. The frequency of coverage calculated for the credible intervals found in this way is approximately the same as if it were found from 1,000 samples when α_0 was used in the weighted Polya posterior for the predictive distribution of the unseen given the seen. As α_0 ranges over the interval $[.21, .29]$ the frequency of coverage of the approximate .95 credible interval found in this fashion will be a smooth function of α_0. This will make it much easier to find an approximate solution. Note that the solution will still depend on the 1,000 actual samples drawn but it should be more accurate since it avoids the problems associated with repeating the whole process at several different values of α. For further discussion see Geyer (1994).

For the five lognormal distributions selected above and various sample sizes values of α for which the .95 credible interval were also approximate 95% confidence intervals were found. The results of were not surprising. For a fixed population the the value of α decreases as n, the sample size, increases. For a fixed sample size the value of α increases as the population becomes more skewed. One problem however is that as the population becomes increasingly skewed there are more and more choices of n for which there are no values of α for which the weighted Polya posterior will yield an approximate 95% confidence interval. For example for the population with $sdlog = 1.5$ no value of α works for sample sizes of eleven or smaller. A more important difficulty is that credible intervals based on a weighted Polya with a value of α close to one are too wide because the lower boundary is to small. These two phenomena are related.

To see why this occurs note that for a small sample size n and a value of α very close to one that the .975 quantile of many weighted Polya simulations of population mean will be very close to the maximum value of the sample. This is because there is a very high probability that in any simulated copy of the entire population the vast majority of unseen units will take on the same value, the one that was assigned to the first unseen unit at the start of the simulation. Hence in about $1/n$ of the weighted Polya simulations of the entire population most of the unseen units will be given the maximum value in the sample. For a sample from a population which is strongly skewed to the right the maximum value of the sample will often be considerably larger than the next largest member of the sample. Then for such a sample about $1/n$ of the simulated copies of the entire population will have a mean very close to the maximum of the sample. Now if n is small and the distribution highly skewed the probability that maximum value in a sample exceeds the mean of the distribution can be less than .975. For such problems the program outlined here cannot work. Moreover for distributions and sample sizes where the program does work but needs a value of α very close to one the .025 quantile of many weighted Polya simulations of the population mean will be quite close to the minimum value of the sample. However when such a interval fails to cover it is almost always because it lies below the true mean rather than above it. So although the program outlined above could work for moderately skewed populations it cannot work sensibly for more highly skewed populations.

We can overcome this difficulty to some extent by treating the upper and lower tails separately. Let q_l be a fixed quantile close to 0. For a fixed population and sample size let $\alpha(q_l)$ be found so that under repeated sampling the true population mean is less than the q_lth quantile of the weighted Polya based on $\alpha(q_l)$ approximately q_l percent of the time. If q_u is a fixed quantile close to 1 then $\alpha(q_u)$ can be found in the same way. Then under repeated sampling the interval formed by taking as its lower endpoint the q_l quantile of the weighted Polya based on $\alpha(q_l)$ and as its upper endpoint the q_u quantile of the weighted Polya based on $\alpha(q_u)$ will be an approximate $(1-q_l-(1-q_u))$ percent confidence interval for the population mean. Now even though for a fixed population and sample size $\alpha(q_l)$ will depend on q_l one should be able to find one value of α which works well for $q_l = .01, .025, .05$ and $.075$ and another value of α which works for $q_u = .925, .95, .975$ and $.99$. This was done for the five lognormal populations selected above and various sample sizes. The results are give in Table 3.

Place Table 3 about here

The next step is to use these values to get α for the lower tail and for the upper tail as functions, approximately, of the skewness measure γ_M and n. Since these functions are relatively smooth a straight forward approach is to use least squares to fit a general quadratic function of γ_M and n to the two sets of values of α in Table 3. This was done and the resulting functions were used in the following.

If the value of γ_M for the population under consideration were known then one could use these functions to get the two values of α needed to find the .95 credible interval based on the weighted Polya. However in most cases the value of γ_M will not be known and must be estimated from the sample. As we noted before a natural way to estimate γ_M is to use the standard deviation of the log of the sample to estimate the *sdlog* of the lognormal distribution which is assumed to be a good approximation of the population under consideration. Since this estimator is still a bit biased we modified the esitmator as follows.

When estimating γ_M to find α for the lower tail we devided the squared deviations from the mean for the log of a sample of size n by $n-2$ instead of $n-1$. For estimating γ_M to find α for the upper tail we made a more substantial modification. We are assuming that any population being considered is unimodal and skewed to the right and can be well approximated by some lognormal distribution. However it would be useful if the method was robust against these assumptions. In particular if the population is not close to some lognormal distribution then the *log* of a sample from it will tend not to be symmetric. Now for estimating the α for the lower tail this does not seem to matter much. However it does seem to matter when estimating the α for the upper tail. We see from Table 3 that for a fixed sample size n the values of α vary more for the upper tail than the lower tail. In general as we will see the upper tail problem is more difficult than the lower tail one. Since it is the behavior of the larger values in the sample that is most important in selecting a good α value for the upper tail these are the ones we use in estimating γ_M. In particular for the *log* of a sample we first find its mean. Then our estimate of *sdlog* is just the average of the squared deviations from the mean for those points in the *log* sample which are greater than the mean. For populations which are approximately lognormal this will on the average ignore about half the sample. But for other populations it will give an improved estimate of α for the upper tail.

Finally we always assumed that for any population being consider that $\gamma_M \geq .15$ and so for any sample where the estimated value of γ_M was less than 0.15 we replaced it with the value 0.15. This is consistent with the values of γ_M used in Table 3 and assumes that we have a prior knowledge that the population under consideration is skewed to the right. For a given sample the next step is to use the estimated value of γ_M and the sample size and the two functions found by least squares to find the two values of α to be used in the weighted Polya distributions. Then these two distributions are used to find through simulation the .025 quantile for the lower tail and the .975 quantile for the upper tail for the population mean which determine the interval estimate for the population mean. In what follows we will denote this interval estimator by wtppI.

This interval, wtppI, was compared to the usual t interval and the modification due to Johnson (1978) and the modification due to Kleijnen et al. (1986). They noted that in their simulations the behavior of Johnson's interval was very similar to that of the usual t interval. We found that to be true as well. Typically it was shifted a bit to the right and in 1,000 samples covered the true mean less than ten more times than the t interval did. For this reason no results for the Johnson interval will be included. We let tI denote the usual t interval and kkmI denote the interval of procedure (5) of Kleijnen et. al. (1986).

In the simulation study for each population 1,000 samples of sizes 8, 13 and 18 were taken. For each sample the tI and kkmI intervals were computed and the wtppI interval was found using $R = 1,000$ realizations of the weighted Polya posteriors. The weights were found from the sample using the best fitting quadratic functions. The results are given in Tables 4, 5 and 6.

$\boxed{\text{Place Tables 4, 5 and 6 about here}}$

As was expected the usual normal theory interval tI, based on Student's t distribution, has a frequency of coverage less than .95 in every case, sometimes significantly less when the sample size is small and the population is strongly skewed.

The behavior of the kkmI is more surprising. It does the best job for *ppsales* but gives intervals that are too large for *ppgamma1.5* and *ppcities*. On the other hand the intervals it gives for *ppln* and *ppcounties* are ridiculously short. For the populations *ppsales*, *ppgamma1.5* and *ppcities* the quadratic function of μ, the population mean, used in their approximation has a minimum value of -.50, -0.56 and -.01 when the true population parameter values are used. This means the quadratic function is quite flat and their heuristic solution in such cases yields quite long intervals. This works out well for *ppsales* which is extremely skewed but very poorly for other other populations. On the other hand the minimum values of the quadratic function for *ppln* and *ppcounties* are -47.3 and -10,009. Here the quadratics are quite steep and just keeping the interval which contains the sample mean results in an interval estimate which is way to short. In summary the kkmI intervals perform very poorly.

With the exception of *ppsales* the frequency of coverage for the weighted Polya intervals is reasonably close to .95. *ppsales* is the one population for which the wtppI intervals significantly under cover. The reason appears to be that this population can not be approximated as well by a lognormal distribution as the others. For example, in Figure 1 we compared the histogram of *ppsales* and *ppcounties* using frequency of occurrence for each bin rather than the raw counts, to an approximating lognormal distribution. The parameter values for the lognormal distribution were just the mean and standard deviation of the log of the populations. It is clear from the plots in Figure 1 that the approximation for *ppcounties* is better than the one for *ppsales*. The actual counts in the fifteen bins for *ppcounties* are 141, 77, 38, 18, 11, 2, 2, 4, 3, 1, 2, 2, 0, 1 and 2 and for *ppsales* are 269, 43, 7, 3, 3, 1, 1, 1, 1, 0 0, 1, 0, 0 and 1. More formally one can compare the the distribution function for each population, i.e. the one which puts mass $1/N$ at each unit, to the distribution function of the approximating lognormal distribution at the values of the units belonging to the population. For *ppsales* the maximum absolute difference of these two distribution functions is .0884 and the average absolute difference is .0441. For *ppcounties* the maximum absolute difference is .0430 and the average absolute difference is .0182.

As we noted before for a value of α close to one the 0.975 quantile of the distribution of the population mean under the weighted Polya posterior will almost always be just the maximum value in the sample. For 500 random samples of size eight from *ppsales* the maximum value of the sample was found to fall above the population mean 88.2% of the time. Hence it is not possible to find for this problem an interval estimator based on the weighted Polyas which will contain the population mean approximately 95% of the time. More generally for extremely skewed populations and small sample sizes it seems impossible to find approximate 95% interval estimates for the population mean unless additional prior information is used.

Place Figure 1 about here

4 Hypotheses Testing

Suppose now we have a random sample of n real valued random variables from some unknown distribution with mean μ and we wish to test $H_0 : \mu = \mu_0$ against $H_1 : \mu > \mu_0$ where μ_0 is some specified real number. If n is small and the distribution is skewed then the usual t test based on normal theory performs poorly. Johnson (1978) and Sutton (1993) suggested modifications of the t test for such problems. Chen (1995) considered another modification of the t test and showed that it worked better than the earlier ones. In this section we will propose a test based on the weighted Polya posterior and compare it to the method suggested by Chen. We will also see that the weighted Polya posterior can also be used in some problems of the form $H_0 : \mu = \mu_0$ against $H_1 : \mu < \mu_0$.

The unweighted Polya posterior is closely related to the Bayesian bootstrap of Rubin (1981). See also Lo (1985). One consequence of this fact is that a finite population inference procedure based on the Polya posterior often is closely related to nonparametric inference procedure based on the Bayesian bootstrap. For example the credible intervals for quantiles in finite population sampling based on the Polya posterior found in Meeden and

Vardeman (1991) are closely related to the credible intervals for quantiles in nonparametric problems based on the Bayesian bootstrap found in Meeden (1993). For a more theoretical discussion of this relationship see Ghosh et al. (1985). This fact and the fact that there is also a close relationship between tests of hypotheses and confidence intervals suggest that the weighted Polya distribution could be used to find tests for the problem stated just above.

Suppose for definiteness we want a test which has probability .05 of making the Type I error. We construct our test as follows. After a sample has been observed we estimate γ_M the skewness of the underlying unkown distribution just as before. Next we use this estimate, the sample size n and the function found earlier by least squares for the lower tail to get an estimate of α to use in the weighted Polya. We then find by simulation the .05 quantile of the population mean where we assume that the population size is large and then we reject $H_0 : \mu = \mu_0$ if and only if this simulated quantile is greater than μ_0. In the previous section we only considered credible intervals but we could have also considered upper and lower credible bounds. The test just described is just the test gotten by inverting the .95 lower credible bound in exactly the same way that is done to find the test based on lower confidence bounds. We will denote this test by wtpp and will compare it to the test recommended by Chen (1995) denoted by modt. Actually Chen considered several tests. We will use one that he recommends on the top of page 771. In his notation we used the critical region $t_2 > z_\alpha$ since we only considered sample sizes smaller than 20. Here only $\alpha = .01$ and .05 and denotes the size of the test being considered.

To compare the tests wtpp and modt we used five of the populations studied by Chen. They are Chi-squared which is a chi-squared with 3 degrees of freedom. Exponential which is the standard exponential distribution with $\lambda = 1$. Gamma which is gamma distribution with scale parameter one and shape parameter four, Lognormal which is a lognormal distribution with mean and standard deviation of the log being 0 and 1 and Weibull which is a Weibull distribution with in his notation parameters $a = 1$ and $b = .5$. Note that two of these distributions, the Exponential and the Weibull, are not unimodal since their density functions are decreasing over their range but they are skewed to the right. Also the Chi-squared and the Gamma are not particularly close to a lognormal distribution. Hence these choices will be a good test of the robustness of the method suggested here. Let μ and σ denote the mean and standard deviation of the distribution under H_0 which will be one of the above five distributions.

To study the power of his test Chen considered for a given distribution five alternatives of the same form and same variance with means given by $\mu_0 + k\sigma/\sqrt(n)$ for $k = 0.5, 1, 1.5, 2$ and 2.5. This allows one to use the same samples generated to calculate the type I error rate to calculate the power as well. We did the same and for each of the five populations generated 1,000 samples of size 10 and 1,000 samples of size 15. In each case for both tests their error rates and power were calculated. To calculate the wtpp test 2,000 simulated copies of the weighted Polya were found. Some of the results are given in Tables 7 and 8.

Place Tables 7 and 8 about here

We see from the tables that the two procedures behave quite similarly when $n = 15$. The results for the $n = 10$ simulations are similar so we have not included them. Chen suggested that when n is small(i.e. around 10) or when $\alpha = 0.01$ the rejection region $t_2 > (z_\alpha + t_{n-1,\alpha})/2$ could be used as well. For the sample size $n = 10$ we repeated the simulations for the cases given in Table 7 using this new rejection region which will result in lower observed power for modt. In this case there were nine cases where the oberved power of the two procedures were the same. For the remaining 41 cases wtpp had on the average a 35% increase in its observed power over that of modt.

Chen noted that the modt procedure cannot at present be modified to test $H_0 : \mu = \mu_0$ against $H_1 : \mu < \mu_0$. Formally this is not a problem for the weighted Polya however. Just as before after a sample has been observed we estimate γ_M the skewness of the underlying unkown distribution. Next we use this estimate, the sample size n and the function found earlier by least squares for the upper tail to get an estimate of α to use in the weighted Polya. Assuming we want a level 0.05 test We then find by simulation the .95 quantile of the population mean where we assume that the population size is large and reject $H_0 : \mu = \mu_0$ if and only if this simulated quantile is less than μ_0. For the five distributions simulations similar to those done in Table 7 were done to study the behavior of this test, wtpp.

The results for the $n = 15$ case are given in Table 8. We see that the procedure gives reasonable results in terms of accuracy when $\alpha = .01$ although the power of the test is not high as for the other hypothesis. For the $\alpha = .05$ case the test seems to be less accurate. Recall that we tried to find α values that worked well for a range of quantiles. From here it is clear that our choice works better for tests of level .01 than it does for test of level

.05. For the simulations done for Tables 4, 5 and 6 we kept track of when we failed to cover if it was the fault of the lower limit or the upper limit. The frequency of coverage of the lower limit was always very close to .975 while the frequency of coverage of the upper limit was the one that caused the problem when the method preformed poorly. One way to improve the performance of the method would be to find a different set of α's for the upper tail for each quantile of interest.

The results for the two hypothesis testing problems indicate that the method seems to be fairly robust against departures from the lognormal assumption. Presumably its performance could be improved a bit if some additional values of α were found in Table 3 or for the upper tail a different set of α's were found for each quantile. One difficulty with this approach is that any skewness measure is only a crude summary of the shape of the population. It is possible that two populations have similar values of a skewness measure but are still different enough that the proper choices of α are different. In one sense the actual shape of the population is not important since the weighted Polya posterior automatically takes that into account no matter what value of α is used. However to make the .95 Bayesian credible interval into an approximate 95% confidence then the value of α becomes crucial. In any case the approach given here will improve on the performance on the usual frequentist intervals for a variety of populations that arise in practice.

References

[1] Arnold, B. C. and Groeneveld, R. A. (1995), Measuring skewness with respect to the mode, *The American Statistician*, 49, 34-38.

[2] Barndorff-Nielsen, O. E. and Cox, D. R. (1989), Asymptotic techniques for use in statistics, Chapman and Hall, London.

[3] Berger, J. O. (1985), Statistical decision and Bayesian analysis, Springer-Verlag, New York.

[4] Chen, L. (1995), Testing the mean of skewed distributions, *Journal of the American Statistical Association* 90, 767-772.

[5] Cochran, W. G. (1977), Sampling techniques, Wiley, New York.

[6] Feller, W. (1968), An introduction of probability theory and its applications, volume I. Wiley, New York, 1968.

[7] Ghosh, M. and Meeden, G. (1997), Bayesian methods for finite population sampling, Chapman & Hall, London.

[8] Ghosh, M., Meeden, G. and Vardeman, S. (1985), Some admissible nonparametric and related finite population sampling estimators, *Annals of Statistics* 13, 811-817.

[9] Groeneveld, R. and Meeden, G. (1984), Measures of skewness, *The Statistician*, 33, 391-399.

[10] Johnson, N. J. (1978), Modified t tests and confidence intervals for asymmetrical populations, *Journal of the American Statistical Association* , 73, 536- 544.

[11] Kleijnen, J. P. C., Kloppenburg, G. L. J. and Meeuwsen, F. L. (1986), Testing the mean of an asymetric population: Johnson's modified t test revisited. *Communications in Statistics, Part B– Simulation and Computation*, 15, 715-732.

[12] Lo, A. Y. (1988), A Bayesian Bootstrap for a finite population, *Annals of Statistics* 16, 1684-1695.

[13] Meeden, G. and Ghosh, M. (1981), On the admissibility and uniform admissibility of ratio type estimators. In Statistics: applications and new directions, pages 378-390. Indian Statistical Institute Golden Jubilee International Conference.

[14] Meeden, G. (1993), Noninformative nonparametric Bayesian estimation of quantiles, *Statistics and Probability Letters*, 16, 103-109.

[15] Meeden, G. and Ghosh, M. (1983), Choosing between experiments: applications to finite population sampling, *Annals of Statistics* , 11, 296-305.

[16] Meeden, G. and Vardeman, S. (1991), A noninformative Bayesian approach to interval estimation in finite population sampling, *Journal of the American Statistical Association* , 86, 972-980.

[17] Rubin, D. B., (1981), The Bayesian bootstrap, *Annals of Statistics* , 9, 130-134.

[18] Royall, R. M. and Cumberland, W. D. (1981), An empirical study of the ratio estimator and estimators of its variance, *Journal of the American Statistical Association* , 76, 66-88.

[19] Sutton, C. D. (1993), Computer-intensive methods for tests about the mean of an asymmetrical population, *Journal of the American Statistical Association* , 88, 802-810.

[20] 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.

Table 1: Summary information for the five populations

Population	Mean	Median	Variance	μ_3/σ^3	$b_{1/4}$	γ	γ_M
ppgamma1.5	1.44	1.10	1.50	1.62	0.26	0.38	0.70
ppln	158.63	133.71	9,779.7	1.56	0.22	0.34	0.45
ppcounties	32,916.8	18,328.8	166×10^7	3.01	0.41	0.64	0.66
ppsales	2.41	1.24	17.73	6.10	0.42	0.73	0.60
ppcities	0.285	0.19	0.0448	1.61	0.52	0.63	0.48

Table 2: For population *ppgamma1.5* a comparison, based on 500 samples, of the 95% normal theory confidence interval for the mean, denoted by tI, and three weighted Polya posterior .95 credible intervals, denoted by wtppI.

Sample size	Type	Value of α	Average lower endpoint	Average length	Frequency of coverage
	tI		.63	1.63	.914
	wtppI	.5	.75	1.70	.934
10	tI		.63	1.62	.910
	wtppI	.6	.69	1.84	.940
	tI		.62	1.63	.896
	wtppI	.7	.62	2.07	.954
	tI		. 89	1.08	.934
	wtppI	.3	.95	1.10	.934
20	tI		.90	1.07	.916
	wtppI	.4	.93	1.18	.936
	tI		.90	1.08	.926
	wtppI	.5	.90	1.30	.956

Table 3: Values of α for which the weighted Polya posterior yields quantiles in the lower tail and upper tail for the population mean approximately equal to the frequency which the sample mean is less than or equal to the quantile for five different lognormal populations and various choices of the sample size n.

	$\gamma_M(sdlog)$				
n	.19(.30)	.36(.60)	.53(.90)	.67(1.2)	.83(1.5)
	The Lower Tail				
9	0.24	0.16	-0.08	-0.10	
12	0.21	0.14	-0.02	-0.07	-0.14
15	0.19	0.09	0.01	-0.06	-0.29
18	0.09	0.07	-0.05	-0.19	-0.23
	The Upper Tail				
9	0.65	0.80	0.91		
12	0.47	0.68	0.88	0.93	
15	0.49	0.59	0.73	0.89	0.98
18	0.40	0.59	0.73	0.85	0.89

Table 4: Coverage properties of the 95% confidence intervals tI and kkmI and the .95 weighted Polya credible interval wtppI for 1,000 random samples of size eight.

Population	Type	Average lower endpoint	Average length	Frequency of coverage
	tI	0.50	1.83	0.891
ppgamma1.5	kkmI	-0.11	5.52	1
	wtppI	0.85	2.47	0.938
	tI	81.5	153.5	0.904
ppln	kkmI	158.3	4.84	0.048
	wtppI	108.2	189.3	0.923
	tI	5,391	54,091	0.828
ppcounties	kkmI	33,655	4.92	0.002
	wtppI	17,516	80,730	0.887
	tI	0.20	4.21	0.690
ppsales	kkmI	0.78	5.85	0.890
	wtppI	1.23	6.51	0.810
	tI	0.13	0.31	0.880
ppcities	kkmI	-0.56	3.47	1
	wtppI	0.18	0.42	0.926

Table 5: Coverage properties of the 95% confidence intervals tI and kkmI and the .95 weighted Polya credible interval wtppI for 1,000 random samples of size thirteen.

Population	Type	Average lower endpoint	Average length	Frequency of coverage
	tI	0.74	1.37	0.918
ppgamma1.5	kkmI	-0.03	5.32	1
	wtppI	0.93	1.94	0.952
	tI	102.2	114.8	0.927
ppln	kkmI	159.0	4.42	0.067
	wtppI	115.6	138.6	0.951
	tI	11,082	44,566	0.853
ppcounties	kkmI	34,422	4.53	0.0
	wtppI	19,228	83,377	0.925
	tI	0.59	3.74	0.742
ppsales	kkmI	1.01	5.67	0.889
	wtppI	1.37	8.09	0.860
	tI	0.17	0.23	0.910
ppcities	kkmI	-0.52	3.25	1
	wtppI	0.20	0.31	0.952

Table 6: Coverage properties of the 95% confidence intervals tI and kkmI and the .95 weighted Polya credible interval wtppI for 1,000 random samples of size eighteen.

Population	Type	Average lower endpoint	Average length	Frequency of coverage
	tI	0.86	1.14	0.920
ppgamma1.5	kkmI	0.02	5.01	1
	wtppI	1.00	1.51	0.958
	tI	112.1	92.4	0.923
ppln	kkmI	157.2	4.27	0.061
	wtppI	121.4	108.1	0.941
	tI	14,934	35,153	0.838
ppcounties	kkmI	33,300	4.36	0.0
	wtppI	20,632	55,980	0.927
	tI	0.85	3.13	0.743
ppsales	kkmI	1.00	5.46	0.909
	wtppI	1.45	5.91	0.857
	tI	0.19	0.19	0.918
ppcities	kkmI	-0.50	3.15	1
	wtppI	0.21	0.25	0.961

Table 7: Power comparison of two testing procedures with upper-tailed rejection regions of nominal size 0.05 for 1,000 samples of size $n = 15$ when $\mu = \mu_0 + k\sigma/\sqrt{n}$.

Distribution	Test	$k = 0$	$k = .5$	$k = 1.0$	$k = 1.5$	$k = 2.0$	$k = 2.5$
Chi-squared	modt	.040	.114	.274	.544	.781	.943
	wtpp	.048	.121	.286	.559	.797	.949
Exponential	modt	.032	.114	.302	.564	.852	.982
	wtpp	.045	.143	.333	.610	.874	.983
Gamma	modt	.042	.128	.294	.489	.741	.900
	wtpp	.036	.110	.271	.479	.732	.892
Lognormal	modt	.038	.137	.405	.799	.995	1
	wtpp	.044	.142	.412	.800	.993	1
Weibull	modt	.028	.147	.487	.980	1	1
	wtpp	.045	.194	.560	.985	1	1

Table 8: Power comparison of two testing procedures with upper-tailed rejection regions of nominal size 0.01 for 1,000 samples of size $n = 15$ when $\mu = \mu_0 + k\sigma/\sqrt{n}$.

Distribution	Test	$k = 0$	$k = .5$	$k = 1.0$	$k = 1.5$	$k = 2.0$	$k = 2.5$
Chi-squared	modt	.013	.040	.119	.307	.591	.858
	wtpp	.014	.047	.144	.347	.651	.892
Exponential	modt	.011	.037	.121	.331	.635	.928
	wtpp	.017	.049	.172	.423	.738	.956
Gamma	modt	.009	.028	.111	.266	.497	.726
	wtpp	.007	.028	.097	.254	.483	.703
Lognormal	modt	.008	.035	.173	.574	.955	.999
	wtpp	.020	.056	.242	.651	.975	1
Weibull	modt	.004	.035	.206	.839	.996	.999
	wtpp	.015	.095	.367	.942	1	1

Table 9: The frequnecy of rejecting the null hpothesis for lower-tailed tests based on the weighted Polya of nominal sizes 0.05 and 0.01 for 1,000 samples of size $n = 15$ when $\mu_x = \mu_0 - k\sigma/\sqrt{n}$.

Distribution	$k = 0$	$k = .5$	$k = 1.0$	$k = 1.5$	$k = 2.0$	$k = 2.5$
			nomial size = .05			
Chi-squared	.038	.083	.144	.238	.334	.451
Exponential	.032	.071	.130	.206	.270	.372
Gamma	.057	.103	.189	.340	.470	.592
Lognormal	.078	.150	.250	.342	.433	.518
Weibull	.035	.074	.123	.185	.240	.305
			nomial size = .01			
Chi-squared	.008	.015	.042	.078	.143	.212
Exponential	.010	.026	.052	.091	.141	.194
Gamma	.005	.019	.054	.115	.196	.295
Lognormal	.036	.071	.122	.177	.247	.305
Weibull	.012	.038	.068	.098	.129	.164

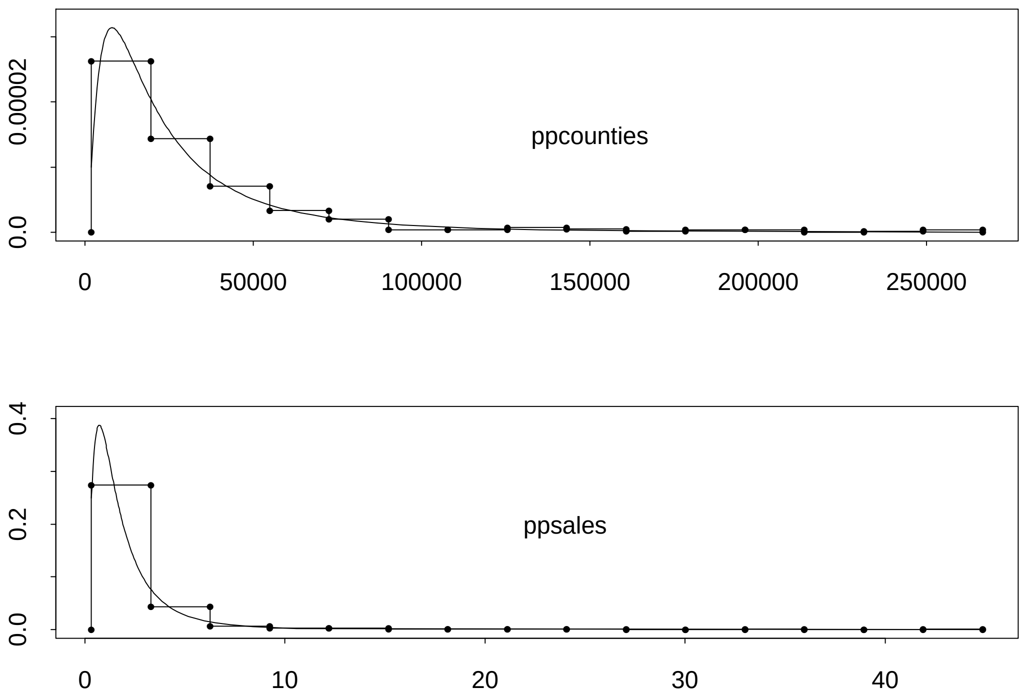

Figure 1: For *ppcounties* and *ppsales* the histogram of the population, based on the relative frequency of each bin and not the raw counts, and an approximating log-normal density.