A stepwise Bayes justification for some Stringer type bounds in auditing problems

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

June 2003

*Research supported in part by NSF Grant DMS 9971331
SUMMARY

For auditors wishing to find an upper bound for the total amount of overstatement of assets in a set of accounts the Stringer bound has often been used despite the fact that in many cases it is known to be much too large. Here we will discuss a family of bounds that are closely related to the Stringer bound and have a stepwise Bayes interpretation. For many problems this should allow a practitioner to select a less conservative bound which reflects available prior information about the population.

AMS 1991 subject classifications Primary 62P05 and 62D05; secondary 62C10.

Key Words and phrases: Accounting, auditing, Stringer bound, stepwise Bayes models and dollar unit sampling.
1 Introduction

Consider a population of \( N \) accounts where the book values \( y_1, \ldots, y_N \) are known. If each account in the population were audited then we would learn the actual or true values \( x_1, \ldots, x_N \). Let \( Y \) denote the total of the book values and \( X \) the total of the audit values then

\[
D = Y - X
\]

denotes the total error in the book values. An auditor, who checks only a subset of the accounts, wishes to give an upper bound for \( D \) based on information contained in the sample. In the important special case where for each account \( y_i - x_i \geq 0 \) an upper confidence bound proposed in Stringer (1963) has been much discussed in the literature. It is easy to use but in many cases it is overly conservative. Moreover the 1989 National Research Council’s panel report on Statistical Models and Analysis in Auditing stated that ”... the formulation of the Stringer bound has never been satisfactorily explained.” This report is an excellent survey of this and related problems.

Fienberg, Neter and Leitch (1977) present confidence bounds based on the multinomial distribution with known confidence levels for all sample sizes. Their bounds however are quite difficult to compute as the number of errors increase. Two other instances where a multinomial model has been assumed are the approaches of Tsui, Matsumura and Tsui (1985) and McCray (1984). In both these cases one must specify a prior distribution to implement the analysis.

Here we introduce a family of stepwise Bayes models for this auditing problem. We show that one member of this family is closely related to the Stringer bound. This gives a new way to think about the Stringer bound. Also one can use prior information about the population to select a member of the family which gives a less conservative bound than the Stringer bound.

In section 2 we briefly discuss the Stringer bound and argue that it can be given a Bayes like interpretation. In section 3 we review the stepwise Bayes approach to multinomial problems and discuss a family of models that give new bounds for our auditing problem. In section 4 we give some simulation results comparing the bounds. We find that the stepwise Bayes models can have reasonable frequentist coverage probabilities and are less conservative than the Stringer bound. In section 5 we generalize the new bound to the situation where where \( y_i - x_i \) can be both positive and negative. In section 5 we conclude with some discussion.
2 A new view of the Stringer bound

2.1 The Stringer bound

We will begin by assuming for each unit that \( x_i \geq 0 \) and the difference \( d_i = y_i - x_i \geq 0 \). Then for \( y_i > 0 \)

\[
t_i = d_i / y_i
\]

is called the tainting or simply the taint of the \( i \)th item. Then the error of the book balance of the account is

\[
D = Y - X = \sum_{i=1}^{N} d_i = \sum_{i=1}^{N} t_i y_i
\]

An important feature of this problem is that a large proportion of the items of the population will have \( d_i = 0 \). In such cases it is not unusual for the sample to contain only a few accounts with \( d_i > 0 \). In fact if either this proportion is large enough or the sample size is small enough then there can be a nontrivial probability of the sample only containing units with \( d_i = 0 \). In such a case the standard methods of finite population sampling are of little use. Before presenting the Stringer bound we need some more notation.

Given an observation from a binomial(\( n, p \)) random variable which results in \( m \) successes let \( \hat{p}_u(m; 1 - \alpha) \) denote an \((1 - \alpha)\%\) upper confidence bound for \( p \).

Suppose we have a sample of \( n \) items where \( k \) of them have a positive error and the remaining \( n - k \) have an error of zero. For simplicity we will assume that the \( k \) taints of the items in error are all unique and strictly between zero and one. Let these \( k \) sampled taints be ordered so that \( 1 > t_1 > t_2 > \cdots > t_k > 0 \). Then the \((1 - \alpha)\%\) Stringer bound is defined by

\[
\hat{D}_{u, st} = Y \left\{ \hat{p}_u(0; 1 - \alpha) + \sum_{j=1}^{k} \left[ \hat{p}_u(j; 1 - \alpha) - \hat{p}_u(j - 1; 1 - \alpha) \right] t_j \right\}
\]

To help understand why the Stringer bound works let \( p \) be the proportion of the items with error in the population and assume that the number of items with error in the sample follows a binomial(\( n, p \)) distribution. We first consider the case where \( k = 0 \), that is all the items in the sample have no error or all the taints are zero. Now \( \hat{p}_u(0; 1 - \alpha) \) is a sensible upper bound
for $p$ and if we assume for each item with a positive error that $x_i = 0$, that is for each item in error the actual error is as large as possible, then $Y \hat{p}_u(0; 1 - \alpha)$, the Stringer bound in this case, is a sensible upper bound for $D$. Next we consider the case when $k = 1$. Now assume that for each item in the population their taint is either one or $t_1$. Then $\hat{p}_u(1; 1 - \alpha) - \hat{p}_u(0; 1 - \alpha)$ is a sensible upper bound for the proportion of items in the population whose taint equals $t_1$. So $Y[\hat{p}_u(1; 1 - \alpha) - \hat{p}_u(0; 1 - \alpha)]t_1$ is a sensible upper bound for the total error associated with such items. Now if we add this to our previous upper bound for all the items in the population with a taint of one we obtain the Stringer bound for this case. Continuing on in this way for samples which contain more than one error we see that the Stringer method adjusts in a marginal fashion bounds which assume that the errors in the population are as large as possible consistent with the observed errors.

### 2.2 A Bayesian interpretation

Here we will discuss how the logic underlying the Stringer bound can be given a Bayesian slant. The $(1 - \alpha)$% Stringer bound given in equation 1 uses the $(1 - \alpha)$% upper confidence bounds for the probability of success based on a binomial sample. Remembering that for the auditing problem a success is getting an item with a positive error it follows that

$$\alpha = \sum_{j=0}^{m} \left( \begin{array}{c} n \\ j \end{array} \right) [\hat{p}_u(m; 1 - \alpha)]^j [1 - \hat{p}_u(m; 1 - \alpha)]^{n-j}$$

$$= 1 - \sum_{j=m+1}^{n} \left( \begin{array}{c} n \\ j \end{array} \right) [\hat{p}_u(m; 1 - \alpha)]^j [1 - \hat{p}_u(m; 1 - \alpha)]^{n-j}$$

$$= 1 - \int_0^{\hat{p}_u(m; 1 - \alpha)} \frac{n!}{m!(n - m - 1)!} u^{m+1-1}(1 - u)^{n-m-1} du$$

where the first equation is just a restatement of the fact that $\hat{p}_u(m; 1 - \alpha)$ is a $(1 - \alpha)$% upper confidence bound and the third uses the well known relationship between the cumulative distribution function of a binomial random variable and the incomplete beta function. Hence $\hat{p}_u(m; 1 - \alpha)$ is the $(1 - \alpha)$th quantile of a beta($m + 1, n - m$) distribution. This is the crucial observation for what follows.

Consider now a sample where all $n$ items had no error. This is not an uncommon occurrence in auditing problems. In such cases it is reasonable to
assume that the population contains items in error even items with a taint equal to one. Let \( p_0 \) be the proportion of items in the population with taint equal to one. If we believe that there is at least one member of the population with such a taint then a sensible posterior given a sample with no items in error is

\[
\propto p_1^{n-1}(1 - p_0)^{n-1}
\]

In this case we are assuming that all the items in the population are of two types; their taints were either zero or one. Now under this posterior the expected value of the total error, \( D \), is \( Y_s/(n + 1) \) where \( Y_s \) is the total book value of the items not in the sample. Why is this? We can imagining selecting a value \( u \) from a beta(1, \( n \)) distribution and then selecting a random sample of size \( u \times (N - n) \) from the items not observed in the first sample. Then assuming that each of these \( u \times (N - n) \) items had a taint equal to one the sum of their book values would be a simulated value for \( D \) under this posterior. If we did this repeatedly for many choices of \( u \) then the mean of the set of simulated values for \( D \) will converge to \( Y_s/(n + 1) \). On the other hand if we wanted a sensible \((1 - \alpha)\)-upper bound for \( D \) we could use the \((1 - \alpha)\)th quantile of our simulated values. Recall for this case the Stringer bound is just \( Y \hat{p}_u(0; 1 - \alpha) = Y_s \hat{p}_u(0; 1 - \alpha) \), since none of the sampled units had any error, and where \( \hat{p}_u(0; 1 - \alpha) \) is the \((1 - \alpha)\)th quantile of the beta(1, \( n \)) distribution.

Note both bounds assume that any actual taint must be one. In addition they can both be interpreted as arising from the same posterior for the number of possible taints in the population. The difference is that Stringer use the product of the \((1 - \alpha)\)th quantile of this distribution and \( Y_s \) to get his bound while the stepwise Bayes approach uses the \((1 - \alpha)\)th quantile of the distribution of \( D \) induced from the posterior.

Next consider a sample where \( n - 1 \) of the items had no error and one had an error with a taint \( t_1 \in (0, 1) \). If we let \( p_0 \) be the proportion of the population with taint equal to one, \( p_1 \) be the proportion of the population with taint equal to \( t_1 \) and \( p_2 \) the proportion of the population with no error then arguing as before a sensible posterior for \((p_0, p_1, p_2)\) is Dirichlet(1, 1, \( n - 1 \)). From this we see that the marginal posterior of \( p_0 + p_1 \) given the sample is beta(2, \( n - 1 \)). Recall that Stringer’s bound involves the \((1 - \alpha)\)th quantile of this distribution and of the beta(1, \( n \)) distribution and the value \( t_1 \).

Now we will give another bound based directly on the Dirichlet(1, 1, \( n - 1 \)) distribution. We begin by generating an observation \( p' = (p'_0, p'_1, p'_2) \) from
this distribution. Then we randomly divide the unsampled items into three groups whose sizes are proportional to \( p' \). In the first group we assume all the taints are one, in the second all the taints are \( t_1 \) and in the third all the taints are zero. Using these assigned taint values for all the unsampled items we find the corresponding value of \( D \) by adding all these simulated errors to the one observed error. This gives one possible simulated value for \( D \) from its posterior distribution induced by the Dirichlet(1, 1, \( n \) − 1) distribution. Now by doing this repeatedly for many choices of \( p' \) and finding the \((1 - \alpha)\)th quantile of this set of simulated values for \( D \) we would have a sensible upper bound for \( D \).

More generally assume we have observed \( k \) distinct taints \( 1 > t_1 > t_2 > \cdots > t_k > 0 \) and \( n - k \) taints equal to zero in the sample. Let \( p_0 \) be the proportion of the population with a taint of one and \( p_{k+1} \) the proportion with a taint of zero. For \( i = 1, \ldots, k \) let \( p_i \) be the proportion with taint equal to \( t_i \). In this case our posterior will be Dirichlet(1, 1, . . . , 1, \( n - k \)) where there are \( k + 1 \) parameters equal to one in this distribution. Now just as in the simpler cases this posterior induces a distribution for \( D \). It is easy to generate simulated values of \( D \) and take the \((1 - \alpha)\)th quantile of a large set of such simulated values as an upper bound for \( D \). One would expect that this bound would be somewhat smaller than the Stringer bound and in fact we will see that this is so.

3 A family of stepwise bounds

3.1 The multinomial model

Consider a random sample of size \( n \) from a multinomial distribution with \( k+1 \) categories labeled 0, 1, . . . , \( k \). Let \( p_j \) denote the probability of observing category \( j \) and \( W_j \) be the number of times category \( j \) appears in the sample. Then \( W = (W_0, \ldots, W_k) \) has a multinomial distribution with parameters \( n \) and \( p = (p_0, \ldots, p_k) \). Now \( W/n \) is the maximum likelihood estimator (mle) of \( p \) and is admissible when the loss function is the sum of the individual squared error losses and the parameter space is the \( k \)-dimensional simplex

\[
\Lambda = \{ p : p_i \geq 0 \text{ for } i = 0, 1, \ldots, k \text{ and } \sum_{j=0}^{k} p_j = 1 \}
\]

For many problems both Bayesian and non Bayesian statisticians would
be pleased to be presented with an “objective” posterior distribution which yielded inferences with good frequentist properties. Then give the data they could generate sensible answers without actually having to worry about selecting a prior distribution. For the multinomial model one such posterior distribution is well known. Given the data \( w = (w_0, \ldots, w_k) \) let \( I_w \) be the set of categories for which \( w_i > 0 \). We assume \( I_w \) contains at least two elements and let

\[
\Lambda_w = \{ p : \sum_{i \in I_w} p_i = 1 \text{ and } p_i > 0 \text{ for } i \in I_w \}
\]

Then we take as our “objective posterior”

\[
g(p|w) \propto \prod_{i \in I_w} p_i^{w_i-1} \quad \text{for } p \in \Lambda_w
\]  

(3)

Note the posterior expectation of the above is just the mle. This fact can be used to prove the admissibility of the mle. Although not formally Bayesian against a single prior these posteriors can led to a noninformative or objective Bayesian analysis for a variety of problems. It is the stepwise Bayes nature of these posteriors that explains their somewhat paradoxical properties. Given a sample each behaves just like a proper Bayesian posterior but one never had to explicitly specify a prior distribution. These posteriors are essentially the Bayesian bootstrap of Rubin (1981).

Cohen and Kuo (1985) showed that admissibility results for the multinomial model can be extend to similar results for finite population sampling. For more details on these matters see Ghosh and Meeden (1997). When the sample size is small compared to the population size the finite population setup is asymptotically equivalent to assuming a multinomial model. Since the later is conceptually a bit simpler, in what follows we will restrict attention to the multinomial setup. We next discuss a slight modification of the stepwise Bayes posteriors which gives the mle. This new family of posteriors will then be applied to the auditing problem.

3.2 Some new stepwise Bayes models

The posteriors of the previous section were selected because they lead to the mle. Here we present a slight modification of them to get posteriors which will be sensible for our auditing problem.

To help motive the argument imagine one believes that \( p_0 > 0 \) but there is a possibility that it could be quite small. In such a case it can be quite
likely that the category zero would not appear in a sample. But even when \( w_0 = 0 \) one would still want the posterior probability of the event that \( p_0 > 0 \) to be positive. Note this would not be true for the posteriors of the previous sections.

Keeping in mind that category zero will play a special role one extra bit of notation will be helpful. If \( w \) is any sample for which \( w_i > 0 \) for at least one \( i \neq 0 \) then we let \( I_{0,w} \) be category zero along with the set of categories for which \( w_i > 0 \) and
\[
\Lambda_{0,w} = \{ p : \sum_{i \in I_{0,w}} p_i = 1 \text{ and } p_i > 0 \text{ for } i \in I_{0,w} \}
\]

For such a \( w \) our posterior has the form
\[
g(p|w) \propto p_0^{w_0+1-1} \prod_{i \in I_{0,w} \text{ and } i \neq 0} p_i^{w_i-1} \text{ for } p \in \Lambda_{0,w}
\]

It useful to compare equations 3 and 4. The posterior in equation 3 gives positive probability only to categories that were observed in the sample and essential treats them the same. That is the posterior depends only on the observed sample frequencies. This is why it can be thought of as an objective or noninformative posterior. On the other hand the posterior in equation 4 always assigns positive probability to the event \( p_0 > 0 \) whether or not the zero category has appeared in the sample. But the observed categories are treated similarly.

We should also point out that in the first posterior the actual values of the categories play no role in the argument. The categories could be given any set of labels and the same argument applies. This is true for the second posterior as well with the exception that one must be able to identify the exceptional category, zero, before the sample is observed. With no loss of generality the actual values of the other categories can depend on the sample.

The family of posteriors in equation 4 is exactly the one described in section 2.2 when giving the Bayesian interpretation of the Stringer bound. There the special category corresponding to \( p_0 \) was units with a taint equal to one. We will denote this case as the “objective” stepwise Bayes bound. Note it is based on exactly the same assumptions that underlie the Stringer bound.

In principle there is no reason to assume that the unobserved category has taint equal to one. One could choose any value of \( t \in (0,1] \) for the
unobserved taint. Selecting a $t < 1$ will give a smaller bound than the “objective” choice of $t = 1$. Moreover it allows one to make use of their prior information about the population in a simple and sensible way. The theory describe in section 3.1 for the family of posteriors given in equation 3 extends in a straightforward way to the family of posteriors given in equation 4. Again one can check Ghosh and Meeden (1997) for details.

4 Some simulation results

In this section we present some simulation results which compare our proposed bounds to the Stringer bound. We also computed a bound proposed by Pap and Zuijlen (1996). Bickel (1992) studied the asymptotic behavior of the Stringer bound and they extended his work demonstrating the asymptotic conservatism of the bound. They proposed a modified Stringer bound which asymptotically has the right confidence level. Although their bound is easy to compute we will not say more about it here. You should check their work for further details.

Although not stated explicitly in the previous section the new bounds are based on the assumption that the size of an item’s taint is independent of its book value. This should be clear when one recalls how the simulated values for $D$ are generated. We should not expect the new method to work well for populations where this assumption is violated.

To see how the new bounds could work we constructed nine populations of 500 items. In each case the book values are the same, a random sample of 500 from a gamma distribution with shape parameter ten and scale parameter one. The sum of the book values for this population is 4942.3. The nine populations are divided into three groups, each with three populations. Within each group there was a population with 2%, 5% and 10% of the items in error. To construct the true values for the items we did the following. In the first group we selected a random sample of 2%, 5% and 10% of the items and set their $x$ value equal to the product of their $y$ value and a uniform(0,1) random variable. All the uniform(0,1) random variables were independent. For the remaining items their $x$ values were set equal to their $y$ values. The second group was constructed in exactly the same manner except that a uniform(0.01,0.1) distribution was used to generated the items in error. Note in both these groups the size of the taint should be independent of the book value. Perhaps it should be more difficult in the second group to find an up-
per bound for $D$ because the taints will be much larger on the average and so $D$ will be larger. The third group was constructed to violate the assumption that the size of the taint is independent of the book value. Here rather than selecting the items in error at random we selected the items with the largest 2%, 5% and 10% book values and then multiplied each of those in the set by a uniform(0.0,0.1) random variable.

To compare the performance of the bounds over the nine populations we took 500 samples of sizes 40 and 80 where the items were selected at random without replacement with probability proportional to the book values. In auditing problems this is commonly referred to as *Dollar Unit Sampling*. For each sample we computed the 95% Stringer bound. We also computed a bound proposed by Pap and Zuijlen (1996). Bickel (1992) studied the asymptotic behavior of the Stringer bound and they extended his work demonstrating the asymptotic conservatism of the bound. They proposed a modified Stringer bound which asymptotically has the right confidence level. Finally we computed the objective stepwise Bayes bound where we assume the unobserved taint $t = 1.0$. The new bound was found approximately for each sample by simulate 500 values of $D$ using its posterior. We then use the 0.95 quantile of these simulated values as our bound.

The results are given in Table 1 when the sample size was 40. (The results when the sample size was 80 were very similar.) The first thing to note is that Stringer bound is indeed quite generous (column six of the table) and exceeds its nominal level. The objective stepwise bounds are indeed shorter than the Stringer bounds but in most cases still quite generous. The frequency of coverage of these bounds never falls below the nominal 0.95 level, a quite acceptable performance. The adjusted Stringer bound is clearly the best in the first seven populations but significantly under covers in the last two populations. Recall that in these populations the size of the taint does depend on the book value and were constructed as extreme cases. Perhaps it is not surprising that for these small sample sizes of 40 and 80 an asymptotic argument cannot improve on Stringer for all possible populations.

Let $t$ denote the value of the unobserved taint that our model assumes is in the population even when it is not observed in the sample. Instead of assigning it the value one we could let it represent our prior guess for the average value of all the positive taints in the population. This will yield shorter bounds and allows one to incorporate prior information into the problem. To see how this could work we ran simulations for the nine populations with $t = 0.5$. This is in fact approximately the average value of the taints in the
first three populations but much to small for for the last six. The results are
given in table two. We see that reasonable prior information can result in
significantly smaller bounds without sacrificing approximate correct coverage
probability.

When all the items in the sample have no error the frequency of coverage
depends entirely on the model underlying the procedure. This is true to a
lesser extent when the sample contains just one item with a positive taint.
This is true even for the Stringer bound, For example for the third population
in the third group where the proportion of items in error is 10% we took 500
samples of size 18 and compared the Stringer and objective stepwise Bayes
bound. The frequency of coverage for the Stringer bound was 0.970 and
for the stepwise Bayes bound 0.976. The ratio of the average length of the
stepwise bound and the average length of Stringer’s bound was 0.923. This
occurred because in the 15 cases where all the sampled items had no error
the Stringer bound never covered while the stepwise Bayes bound covered
in three samples. The sample size was selected so that The Stringer bound
would be too small when the sample contained no items in error.

The stepwise Bayes bounds proposed here essentially assume a multino-
mial model where the number of categories used depends on the observed
sampled. Two other instances where a multinomial model has been assumed
in a nonparametric Bayesian setting are the approaches of Tsui, Matsumura
and Tsui (1985) and McCray (1984). In these cases however one must specify
a full prior distribution in contrast to just the one parameter needed here.

5 Some generalizations

Our stepwise Bayes posteriors identifies a category which may or may not
appear in our sample but which we always assume appears in the population.
In the objective stepwise Bayes posterior we have assumed that items in this
category have taint one. A more subjective approach lets the unobserved
taint \( t \) represent a guess for the average value of all the positive taints in the
population.

Suppose the population contains items with both positive and negative
taints. This is an important case for which the Stringer approach does not
apply. An upper bound of one for the positive taints is always available al-
though in some cases prior information might suggest a smaller upper bound.
In some cases a lower bound for the negative taints could also be at hand.
In this case let $t_1$ and $t_2$, denote these bounds. More generally we can think of $0 < t_1 \leq 1$ and $t_2 < 0$ as reflecting our beliefs about how much positive and negative error there is in the population. So instead of just one special category we now have two. By specifying the parameters $(t_1, t_2)$ we can select a stepwise Bayes model which reflects our prior information.

To see how this could work in practice we constructed four populations where the set of book values was the same random sample from a gamma distribution used for the nine earlier populations. In each of the four populations there were 50 items in error. To get the taint of a selected item we took twice the observation from a beta distribution and subtracted one to get a taint in $(-1, 0, 1)$. We then took one minus the taint value and multiplied it by the book value to get the true value for the item. In the first pair of populations the items in error were selected at random while in the second pair they were the items with the 50 largest book values. It remains to describe what beta distributions were used. For the first population in the first pair we took independent random samples of sizes 15, 20 and 15 from beta(1,7), beta(7,7) and beta(7,1) distributions respectively. This should produce a group of taints which are roughly symmetric about zero and with 40% of them close to zero. We will denote this population by $sym(rnd)$. To get a collection of taints that would be skewed to the right for the second population in the first pair we took the random sample sizes to be 10, 20 and 20 from the three distributions. We will denote this population by $skr(rnd)$. The taints for the true values for the second group were produced in exactly the same way and we denote these populations by $sym(ext)$ and $skr(ext)$.

A sensible choice of $(t_1, t_2)$ would be $(1, -1)$. This is certainly true for the two symmetric populations. But it could be more problematical for the two skewed populations where there are more positive taints then negative ones. In order to see how robust these choices are we also considered two other selections, $(3/4, -1)$ and $(1, -3/4)$. The results are given in Table 3 where as before we used Dollar Unit Sampling to select the samples. In all cases except one (population $skr(rd)$ with $n = 80$) the choice $(1, -1)$ yields the point estimator which is closest to being unbiased. Overall the intervals work well and fall significantly below the nominal level in just one case for population $skr(ext)$ with $n = 80$ and the choice $(1, -3/4)$. Although the results are not given in Table 3 the same set of simulations was done for the choice $(3/4, -3/4)$. For $n = 40$ the coverage frequencies for the 0.95 credible intervals were 0.992, 0.928, 0.95 and 0.932 and for $n = 80$ they were 0.928, 0.918, 0.886 and 0.906.
In Table 4 we repeated these simulations when \((t_1, t_2) = (0.5, -0.5)\). We also included the average absolute error of the point estimator of \(D\). We would expect that the method should work for the population \(sym(rnd)\) so it is somewhat surprising that it preforms poorly for the \(n = 80\) case for this population. These simulations suggest that although these bounds do behave sensibly their frequentist coverage probability can fall below their nominal level when the prior information used to select a stepwise Bayes model does not reflect the population at hand. However this approach should work well if reasonable bounds for the taints can be found.

In equation 4 the exponent of \(p_0\) is \(w_0 + 1 - 1\). More generally one could take this exponent to be \(w_0 + a - 1\) where \(a > 0\) is some constant. In this case \(E(p_0 \mid w) = a/(n + a)\) is an increasing function of \(a\). An alternative method of incorporation prior information would be keep its taint value equal to one but vary \(a\). Here we have chosen to keep \(a = 1\) and vary \(t\) since it seems to us to be an easier way to think about the problem.

More generally in some situations one might have a good prior guess for the distribution of the set of possible taints. This could be represented by vectors \(t\) and \(a\). The components of \(a\) should all be positive and their relative sizes would reflect how likely their corresponding \(t\) values appear in the population. The sum of the components of \(a\) would influence how much weight is give to the prior guess in the posterior. We will not present any simulations for these more complete prior guess for the taints. Clearly the performance of these procedures will depend on the quality of the prior assessments but they could prove useful when such information is present.

6 Discussion

We have presented here a new family of stepwise Bayes bounds one of which is closely related to the well known bound of Stringer. We have seen that using prior information about the population can lead to a significantly smaller bound than Stringer’s. These new bounds will have good coverage properties if they are based on sensible prior information but their performance can degrade when that information is faulty. The models can be extended to cover the case when errors can be both positive and negative. One advantage of the approach give here is that it is very easy to simulate values of \(D\) given a sample. For more details see the appendix.

In the cases where all the errors are positive the new bounds will have
the greatest improvement over the Stringer bounds for populations when the frequency of error is quite small and the size of the errors is very small. When the sample contains at least a couple of items with positive error then the subclass of bounds studied here all become quite similar and behave much like the Stringer bound except that they tend to be a bit shorter.

It seems problematical that for small sample sizes one will be able to find bounds which dramatically improve on the Stringer bounds across all possible populations, i.e. bounds which are significantly smaller but still have approximately the nominal coverage probability. Consideration of difficult populations like the third population in group three indicates why this is so. For larger sample sizes most samples will have several items with positive taints. For such samples the Stringer bounds are not much larger than those of the objective stepwise Bayes model. This objective stepwise Bayes model is closely related to the Polya posterior which is closely related to many of the standard procedures in finite population sampling. (For details see Ghosh and Meeden (1997).) Even though there is no known argument proving that the Stringer bound achieves at least its nominal level for all possible populations we believe that the material presented here indicates that the Stringer bound is using the data in a sensible although perhaps a slightly conservative way. This is consistent with the results of Fienberg, Neter and Leitch (1977) as well. The methods given here allows one to get less conservative bounds when prior information suggests that the Stringer bound will be too generous.
References


Appendix

The bounds presented here are easy to compute approximately. The key fact used in the program is that a Dirichlet random variable can be generated from a sum of independent gamma distributions. A program has been written in R which allows one to simulate values of $D$ once the vectors $t$ and $a$ have been specified. A interested reader can use this program in RWeb at the author’s web site

http://www.stat.umn.edu/~glen/

Once there click on the “Rweb functions” link and select “simulateD.html”. Then you can construct simple examples to see how the methods presented here work in practice.
Table 1: The behavior of the 95% Stringer bound, the 95% adjusted Stringer bound and the 0.95 objective stepwise Bayes bound. The results are based on 500 samples of size 40 for the nine populations. The average values of the Stringer, the adjusted Stringer and the stepwise Bayes bounds are respectively $avStgB$, $avAdjB$ and $avSwBB$.

<table>
<thead>
<tr>
<th>Group (pop)</th>
<th>D</th>
<th>Freq of Coverage</th>
<th>As a % of avStgB</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Stg</td>
<td>SwB</td>
</tr>
<tr>
<td>1(2%)</td>
<td>61.7</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>1(5%)</td>
<td>150.2</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>1(10%)</td>
<td>252.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>2(2%)</td>
<td>75.9</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>2(5%)</td>
<td>239.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>2(10%)</td>
<td>480.6</td>
<td>0.990</td>
<td>0.976</td>
</tr>
<tr>
<td>3(2%)</td>
<td>180.1</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>3(5%)</td>
<td>415.2</td>
<td>0.996</td>
<td>0.996</td>
</tr>
<tr>
<td>3(10%)</td>
<td>770.7</td>
<td>0.970</td>
<td>0.968</td>
</tr>
</tbody>
</table>

Table 2: The behavior of the 0.95 stepwise Bayes bound when the unsampled taint $t = 0.5$. The results are based on 500 samples of size 40. The last column gives the ratio of the average value of the $t = 0.5$ bounds to the average value of the $t = 1.0$ bounds.

<table>
<thead>
<tr>
<th>Group (pop)</th>
<th>D</th>
<th>ave Bnd</th>
<th>Freq of $t.5/t1$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Freq</td>
<td></td>
</tr>
<tr>
<td>1(2%)</td>
<td>61.7</td>
<td>269 1.0</td>
<td>0.66</td>
</tr>
<tr>
<td>1(5%)</td>
<td>150.2</td>
<td>397 1.0</td>
<td>0.75</td>
</tr>
<tr>
<td>1(10%)</td>
<td>252.0</td>
<td>542 0.904</td>
<td>0.85</td>
</tr>
<tr>
<td>2(2%)</td>
<td>75.9</td>
<td>301 1.0</td>
<td>0.69</td>
</tr>
<tr>
<td>2(5%)</td>
<td>239.0</td>
<td>569 0.884</td>
<td>0.88</td>
</tr>
<tr>
<td>2(10%)</td>
<td>480.6</td>
<td>909 0.942</td>
<td>0.98</td>
</tr>
<tr>
<td>3(2%)</td>
<td>180.1</td>
<td>490 0.824</td>
<td>0.85</td>
</tr>
<tr>
<td>3(5%)</td>
<td>415.2</td>
<td>841 0.914</td>
<td>0.96</td>
</tr>
<tr>
<td>3(10%)</td>
<td>770.7</td>
<td>1309 0.96</td>
<td>1.01</td>
</tr>
</tbody>
</table>
Table 3: Results for two-sided 0.95 credible intervals and point estimators for the stepwise Bayes bound for various choice of \((t_1, t_2)\) for 500 samples of sizes 40 and 80 for four populations.

<table>
<thead>
<tr>
<th>Pop</th>
<th>(D)</th>
<th>((t_1, t_2))</th>
<th>(n)</th>
<th>Ave ptest</th>
<th>Ave lowbd</th>
<th>Ave length</th>
<th>Freq of cov</th>
</tr>
</thead>
<tbody>
<tr>
<td>(sym(rnd))</td>
<td>-23.3</td>
<td>(1, -1)</td>
<td>40</td>
<td>-13.4</td>
<td>-399.6</td>
<td>774.3</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3/4, -1)</td>
<td>40</td>
<td>-47.9</td>
<td>-419.0</td>
<td>711.1</td>
<td>0.992</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1, -3/4)</td>
<td>40</td>
<td>0.08</td>
<td>-340.2</td>
<td>712.9</td>
<td>0.990</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1, -1)</td>
<td>80</td>
<td>-28.5</td>
<td>-241.7</td>
<td>424.2</td>
<td>0.964</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3/4, -1)</td>
<td>80</td>
<td>-34.6</td>
<td>-239.9</td>
<td>399.1</td>
<td>0.952</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1, -3/4)</td>
<td>80</td>
<td>-10.1</td>
<td>-204.9</td>
<td>400.4</td>
<td>0.946</td>
</tr>
<tr>
<td>(skr(rnd))</td>
<td>74.1</td>
<td>(1, -1)</td>
<td>40</td>
<td>78.2</td>
<td>-311.8</td>
<td>802.3</td>
<td>0.994</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3/4, -1)</td>
<td>40</td>
<td>40.0</td>
<td>-338.6</td>
<td>744.8</td>
<td>0.980</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1, -3/4)</td>
<td>40</td>
<td>105.8</td>
<td>-236.5</td>
<td>740.2</td>
<td>0.984</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1, -1)</td>
<td>80</td>
<td>81.1</td>
<td>-137.7</td>
<td>452.3</td>
<td>0.962</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3/4, -1)</td>
<td>80</td>
<td>72.4</td>
<td>-139.4</td>
<td>430.1</td>
<td>0.936</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1, -3/4)</td>
<td>80</td>
<td>92.3</td>
<td>-106.2</td>
<td>424.9</td>
<td>0.942</td>
</tr>
<tr>
<td>(sym(ext))</td>
<td>31.3</td>
<td>(1, -1)</td>
<td>40</td>
<td>28.4</td>
<td>-428.2</td>
<td>918.0</td>
<td>0.976</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3/4, -1)</td>
<td>40</td>
<td>20.8</td>
<td>-429.3</td>
<td>884.1</td>
<td>0.968</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1, -3/4)</td>
<td>40</td>
<td>69.7</td>
<td>-350.7</td>
<td>871.8</td>
<td>0.964</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1, -1)</td>
<td>80</td>
<td>42.3</td>
<td>-223.3</td>
<td>539.9</td>
<td>0.912</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3/4, -1)</td>
<td>80</td>
<td>15.5</td>
<td>-244.4</td>
<td>516.4</td>
<td>0.912</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1, -3/4)</td>
<td>80</td>
<td>50.1</td>
<td>-202.7</td>
<td>518.0</td>
<td>0.916</td>
</tr>
<tr>
<td>(skr(ext))</td>
<td>168.7</td>
<td>(1, -1)</td>
<td>40</td>
<td>166.2</td>
<td>-274.4</td>
<td>906.9</td>
<td>0.974</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3/4, -1)</td>
<td>40</td>
<td>141.5</td>
<td>-287.1</td>
<td>866.6</td>
<td>0.962</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1, -3/4)</td>
<td>40</td>
<td>195.6</td>
<td>-207.5</td>
<td>854.7</td>
<td>0.974</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1, -1)</td>
<td>80</td>
<td>184.0</td>
<td>-68.8</td>
<td>524.2</td>
<td>0.928</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3/4, -1)</td>
<td>80</td>
<td>180.9</td>
<td>-67.3</td>
<td>509.5</td>
<td>0.920</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1, -3/4)</td>
<td>80</td>
<td>197.4</td>
<td>-42.3</td>
<td>505.1</td>
<td>0.886</td>
</tr>
</tbody>
</table>
Table 4: Results for two-sided 0.95 credible intervals and point estimators for the stepwise Bayes bound with \((t_1, t_2) = (0.5, -0.5)\) for 500 samples of sizes 40 and 80 for four populations.

<table>
<thead>
<tr>
<th>Pop</th>
<th>D</th>
<th>n</th>
<th>Ave ptest</th>
<th>Ave Abserr</th>
<th>Ave lowbd</th>
<th>Ave length</th>
<th>Freq of cov</th>
</tr>
</thead>
<tbody>
<tr>
<td>sym(rnd)</td>
<td>-23.3</td>
<td>40</td>
<td>-30.5</td>
<td>99.9</td>
<td>-306.8</td>
<td>548.3</td>
<td>0.952</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>-26.1</td>
<td>75.7</td>
<td>-198.5</td>
<td>342.7</td>
<td>0.896</td>
<td></td>
</tr>
<tr>
<td>skr(rnd)</td>
<td>74.1</td>
<td>40</td>
<td>73.5</td>
<td>115.5</td>
<td>-200.9</td>
<td>583.5</td>
<td>0.918</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>78.6</td>
<td>77.7</td>
<td>-94.0</td>
<td>367.3</td>
<td>0.902</td>
<td></td>
</tr>
<tr>
<td>sym(ext)</td>
<td>31.3</td>
<td>40</td>
<td>43.0</td>
<td>162.2</td>
<td>-330.4</td>
<td>756.3</td>
<td>0.876</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>45.7</td>
<td>114.6</td>
<td>-189.7</td>
<td>477.1</td>
<td>0.868</td>
<td></td>
</tr>
<tr>
<td>skr(ext)</td>
<td>168.7</td>
<td>40</td>
<td>175.9</td>
<td>154.9</td>
<td>-166.3</td>
<td>727.6</td>
<td>0.906</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>193.1</td>
<td>113.3</td>
<td>-28.9</td>
<td>468.4</td>
<td>0.880</td>
<td></td>
</tr>
</tbody>
</table>