The Annals of Statistics
1992, Vol. 20, No. 1, 510–522

THE ADMISSIBILITY OF THE LINEAR INTERPOLATION ESTIMATOR OF THE POPULATION TOTAL[1]

BY GLEN MEEDEN

University of Minnesota

Consider the problem of estimating the total of some finite population. Suppose the labels attached to the units of the population are such that members of the population whose labels are close together are more alike than units whose labels are far apart. For such a population a sensible estimator of the population total is one that interpolates linearly between successive members of the sample. The admissibility of this estimator will be demonstrated. The related interval estimators will be studied as well.

1. Introduction. Finite population sampling is one area where everyone agrees that prior information should be used. Frequentists or classical statisticians often use their prior information to stratify the population or choose their design. A Bayesian should choose a prior that reflects his or her prior information. However, because of the large dimension of the parameter space, this is often quite difficult to do. Recently the admissibility of a variety of estimators has been demonstrated using the stepwise Bayes technique [see, e.g., Meeden and Ghosh (1983) and Vardeman and Meeden (1984)]. In addition, this approach gives a conditional interpretation for an estimator which emphasizes the idea that point estimation in finite population sampling is really a prediction problem. That is, given the sample, a sensible estimator relates the observed units in the sample (the seen) to the remaining unobserved units in the population (the unseen) independent of how the sample was chosen. This line of reasoning gives a noninformative Bayesian justification to many of the classical sampling procedures when the statisticians prior beliefs about the population or particular strata are exchangeable in a certain weak sense [see, e.g., Meeden and Vardeman (1991)]. In all these cases it is possible to carry out an analysis without actually specifying a prior distribution. It would be of interest to extend this approach to cases where one's prior beliefs about the units cannot be characterized by some type of exchangeability.

In some instances it is reasonable to assume that units whose labels are close together are more alike than units whose labels are far apart. This may arise naturally from the geometry of the problem. In other cases the statistician may use all of his or her prior information to relabel the population so that this condition is satisfied. How should one make use of this state of affairs

---

Received January 1989; revised May 1991.

[1] Research supported in part by NSF Grant DMS-89-02580.

AMS 1980 *subject classifications*. Primary 62D05; secondary 62C15, 62C10.

Key words and phrases. Finite population sampling, admissibility, stepwise Bayes, linear interpolation, the trapezoid rule.

when choosing an estimator? One possibility is to stratify the population. The strata would consist of units whose labels are a set of successive integers. Even if it were not practical to make the strata small enough to be nearly homogeneous, it would still be foolish to ignore this information. For example, suppose in our sample we have observed that unit i_1 has the value a and unit i_2 has the value b where $i_1 < i_2$ and none of the units whose labels fall between i_1 and i_2 belong to the sample. Let j denote the label of a typical unsampled unit where $i_1 < j < i_2$. If we really believed that units whose labels are close together are more alike than those units whose labels are far apart, then our estimate for the value of the unit with label j should depend only on the numbers a and b, that is, on the values of the two units in the sample which are closest to it. Moreover a natural choice for the estimate is $[(j - i_1)b + (i_2 - j)a]/(i_2 - i_1)$, that is, just the value at j of the straight line passing through the points (i_1, a) and (i_2, b). That is, for every unsampled member of the population, we take as our estimate the value of the linear interpolation between its two closest points in the sample. This in a straightforward way leads to an estimate of the population total.

In Section 2, we will use the stepwise Bayes technique to prove that the above estimator of the population total is admissible under squared error loss. In Section 3, some examples will be considered and the related interval estimators studied as well.

2. Proving admissibility. Let U denote a finite population with N units labeled $1, \dots, N$. Let y_i be the value of a single characteristic attached to unit i. The vector $\mathbf{y} = (y_1, \dots, y_N)$ is the unknown state of nature and is assumed to belong to the parameter space θ. By convention, θ is often taken to be $\mathbb{R}^N$, that is, N-dimensional Euclidean space. Other sensible choices are also possible. Let $\boldsymbol{\lambda} = \{\lambda_1, \dots, \lambda_r\}$ be a set of known constants and let

$$\theta(\boldsymbol{\lambda}) = \{\mathbf{y} |\text{ for each } i = 1, \dots, N, y_i = \lambda_j \text{ for some } j = 1, \dots, r\}. \tag{2.1}$$

In what follows occasionally it will be convenient to assume that the parameter space $\theta = \theta(\boldsymbol{\lambda})$ for some $\boldsymbol{\lambda}$.

A subset s of $\{1, \dots, N\}$ is called a sample. Let $n(s)$ denote the number of elements belonging to s. Let S denote the set of all possible samples. A design is a function Δ defined on S such that $\Delta(s) \in [0, 1]$ and $\Sigma_{s \in S} \Delta(s) = 1$. Given $\mathbf{y} \in \theta$ and $s = \{i_1, \dots, i_n\}$, that is, $n(s) = n$, where $1 \le i_1 < \cdots < i_n \le N$, let $\mathbf{y}(s) = (y_{i_1}, \dots, y_{i_n})$. Suppose we wish to estimate $\gamma(\mathbf{y}) = \Sigma_{i=1}^N y_i$, the population total with squared error loss. Let $e(s, \mathbf{y})$ denote an estimator of $\gamma(y)$, where $e(s, \mathbf{y})$ depends on $\mathbf{y}$ only through $\mathbf{y}(s)$. If the design Δ is used in conjunction with the estimator e, then the risk function is

$$r(\mathbf{y}; \Delta, e) = \sum_s [e(s, \mathbf{y}) - \gamma(\mathbf{y})]^2 \Delta(s). \tag{2.2}$$

The goal of this note is to prove the admissibility of the estimator described in the Introduction for a given design when the parameter space $\theta = \mathbb{R}^N$. This result will follow easily once we have demonstrated its admissibility when the parameter space $\theta = \theta(\boldsymbol{\lambda})$ for an arbitrary vector $\boldsymbol{\lambda}$. For this choice of the parameter space, both the parameter space and sample space contain only a finite number of points. For such a problem to prove the admissibility of a given estimator, it is enough to show that it is a unique stepwise Bayes procedure against some finite sequence of priors. For our purpose a convenient statement of this result is Theorem 1 of Meeden and Ghosh (1981). The stepwise Bayes technique was first used in Johnson (1971) and named in Hsuan (1979). Brown (1981) gives a more useful general version when the parameter space is no longer finite.

In the Introduction we suggested an estimator of the population that, intuitively, would be appropriate when the statistician believes that units with labels close together are more alike than units with labels which are far apart. For an unsampled unit the estimate of its value is just the linear interpolation of the observed values of the two units which are closest to it in the sample. For units with labels less than the smallest label appearing in the sample, their estimate is just the observed value of the unit with the smallest label in the sample. Units with labels larger than the largest label appearing in the sample are handled in a similar manner. If $s = \{i_1, \dots, i_n\}$ with $n(s) = n$ and where $1 \le i_1 < \cdots < i_N \le N$ is a sample, then the estimator described above is given by

$$(2.3)\quad e^*(s,y) = \begin{cases} Ny_{i_1}, & \text{if } n(s) = n = 1, \\ \frac{1}{2}\{y_{i_1}(i_1 + i_2 - 1) + y_{i_2}(2N - i_1 - i_2 + 1)\}, & \\ & \text{if } n(s) = n = 2, \\ \frac{1}{2}\left\{y_{i_1}(i_1 + i_2 - 1) + \sum_{j=2}^{n-1} y_{i_j}(i_{j+1} - i_{j-1})\right. & \\ \quad \left. + y_{i_n}(2N - i_{n-1} - i_n + 1)\right\}, & \text{if } 2 < n(s) = n < N, \\ \sum_{i=1}^{N} y_i, & \text{if } n(s) = n = N. \end{cases}$$

Note that in the special case when all the members of the sample are observed to be the same value, the value of e^* for such a sample is just N times the common value.

THEOREM. *Consider the problem of estimating the population total with squared error loss and with the parameter space* $\theta = \theta(\boldsymbol{\lambda})$ *as defined in* (2.1) *for a given set* $\boldsymbol{\lambda}$. *Then for a given design* Δ, *the estimator* e^*, *defined in* (2.3), *is admissible.*

PROOF. Recall that if π is a prior over $\theta(\boldsymbol{\lambda})$, then given the sample $(s, \mathbf{y}(s))$, the Bayes estimate of the population total is

$$(2.4)\qquad \sum_{i \in s} y_i + \sum_{j \notin s} E\big(y_j \big| (s, \mathbf{y}(s))\big)$$

[see Basu (1969) for details].

Let $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_{r_0})$ be a set of known constants when r_0 is a known positive integer. If $r_0 = 1$, then the parameter space contains just one point and it is obvious that e^* is admissible. Suppose now $r_0 \geq 2$.

Our first prior puts mass $1/r_0$ on the r_0 points of the form $\mathbf{y}_j = (\lambda_j, \dots, \lambda_j)$ for $j = 1, \dots, r_0$. Under this prior the only points in the sample space are those where the observed units in the sample are all equal to one of the λ_j's. It is easy to check that under this prior e^* is the unique Bayes estimator for all those outcomes with positive probability.

Thus, we have shown that e^* is Bayes for the observed samples where the sampled units are all equal. We will next consider samples where the units take on only two possible values. Hence, from now on we only need to consider samples which contain at least two units. For the rest of the proof, Δ will be a fixed design which assigns positive probability to at least one sample s with $n(s) \geq 2$.

For notational convenience, let $a = \lambda_1$ and $b = \lambda_2$. Our proof will begin by defining a sequence of priors defined on disjoint subsets of $\theta(a, b)$. We will continue until we have shown that e^* is a stepwise Bayes estimator for all possible samples where the observed values are either a or b. Let

$$\begin{aligned}\theta^1(a,b) = \{\mathbf{y} | y_i = a \text{ for } i = 1, \dots, j \text{ and } y_i = b \text{ for } i = j+1, \dots, N \\ \text{or } y_i = b \text{ for } i = 1, \dots, j \text{ and } y_i = a \text{ for } i = j+1, \dots, N \\ \text{for some } j = 1, \dots, N-1\}\end{aligned}$$

and

$$\begin{aligned}\mathscr{X}^1(a,b) = \{(s, \mathbf{y}(s)) | \Delta(s) > 0, \mathbf{y} \in \theta^1(a,b) \text{ and} \\ \mathbf{y}(s) \text{ contains at least one } a \text{ and one } b\}.\end{aligned}$$

Hence, $\mathscr{X}^1(a,b)$ are all those samples where the units are first all a's and then followed by all b's or vice versa.

Now we consider the prior π which assigns equal probability to the members of $\theta^1(a,b)$. Let $s = \{i_1, \dots, i_n\}$ with $n(s) = n$ and suppose $(s, \mathbf{y}(s)) \in \mathscr{X}^1(a,b)$.

We first assume that there exists an integer j^* such that $1 \leq j^* \leq n-1$, $i_{j^*} + 1 = i_{j^*+1}$ and

$$y_{i_j} = \begin{cases} a, & \text{for } j = 1, \dots, j^*, \\ b, & \text{for } j = j^* + 1, \dots, n. \end{cases} \tag{2.5}$$

For such a sample the posterior probability distribution under π concentrates all its mass at one point of $\theta^1(a,b)$ and e^* agrees with the Bayes estimator against π. Note that a similar result holds if a and b are interchanged in (2.5).

Next we consider the case where j^* is as above except now $i_{j^*} + 1 < i_{j^*+1}$. For notational convenience, set $j_0 = i_{j^*}$ and $\gamma = i_{j^*+1} - i_{j^*}$. For this sample and prior π, the posterior distribution assigns equal probability to the γ points of the form

$$\mathbf{y}^r = \{y_i | y_i = a \text{ for } i = 1, \dots, j_0 + r \text{ and } y_i = b \text{ for } i = j_0 + r + 1, \dots, N\}$$

for $r = 0, 1, \dots, \gamma - 1$. Hence for $r = 1, 2, \dots, \gamma - 1$, the posterior probability that $y_{j_0+r} = b$ is r/γ and so

$$(2.6)\quad \sum_{r=1}^{\gamma-1} E(y_{j_0+r}|\text{sample}) = \sum_{r=1}^{\gamma-1}\left(\frac{\gamma-r}{\gamma}a + \frac{r}{\gamma}b\right) = \frac{(\gamma-1)(a+b)}{2}.$$

Note that the posterior expectation of any other unsampled unit is either a if its label is less than i_{j^*} or b if its label is greater than i_{j^*+1}. It is easy to check, using this along with (2.6) that (2.4) will yield (2.3). Note that a similar argument holds if a and b are interchanged in (2.5).

If the design Δ only assigns positive probability to samples of size two or less, the proof of the theorem would essentially be completed at this point. Hence, we will assume that this is not so. In this case it might be possible to observe a sample of size three or larger where $y_{i_1} = y_{i_3} = a$ and $y_{i_2} = b$. The next stage of the argument will consider points of this type.

Let

$$\theta^2(a,b) = \{\mathbf{y}|\text{ there exists integers } 1 \le v_1 < v_2 < N \text{ such that } y_i = a \text{ for } 1 \le i \le v_1 \text{ and } v_2 < i \le N \text{ and } y_i = b \text{ for } v_1 < i \le v_2 \text{ or vice versa}\}$$

and

$$\mathscr{X}^2(a,b) = \{(s,\mathbf{y}(s))|\Delta(s) > 0, \mathbf{y} \in \theta^2(a,b) \text{ and } \mathbf{y}(s) \text{ contains at least one } a \text{ and } b \text{ and } \mathbf{y}(s) \notin \mathscr{X}^1(a,b)\}.$$

Just as before we will take as our prior π, the distribution which assigns equal mass to the points in $\theta^2(a,b)$. The next step is to compute the Bayes estimate against π for all sample points in $\mathscr{X}^2(a,b)$ and see that it agrees with e^*.

The argument now proceeds in a similar fashion where at any stage, given a sample the posterior distribution gives equal weight to all those populations which are the simplest possible step functions consistent with the sample. To aid in determining the proper order, we introduce a definition for $\mathbf{y} \in \theta(\boldsymbol{\lambda})$. We say that $\mathbf{y} = (y_1, \dots, y_N)$ has order k if the number of labels i such that $y_{i+1} \ne y_i$ is k. Note that the order of $y(s)$ is well-defined as well where we take the order of $y(s)$ to be 0 if s contains just one point. Hence, $\theta^2(a,b)$ is just all those vectors of order 2 whose individual components are either a or b.

The next step is to consider

$$\theta^3(a,b) = \{\mathbf{y}|\mathbf{y} \in \theta(a,b) \text{ and } \mathbf{y} \text{ is of order } 3\}$$

and

$$\mathscr{X}^3(a,b) = \{(s,\mathbf{y}(s))|\Delta(s) > 0, \mathbf{y} \in \theta^3(a,b) \text{ and } \mathbf{y}(s) \text{ is of order } 3\}.$$

The next step is to consider $\theta^4(a,b)$ and continue in this way until we have allowed for all possible samples in which just a and b appear. The value of the highest order will depend on Δ. It will be

$$(2.7)\qquad k_0 = \max\{\#(s)|\Delta(s) > 0\} - 1,$$

where $\#(s)$ is the number of elements belonging to s.

The next step is to repeat this process for $\theta(\lambda_1, \lambda_3)$ and then for $\theta(\lambda_1, \lambda_4)$ and so on to $\theta(\lambda_{r_0-1}, \lambda_{r_0})$. In this way we have taken care of all sample points where just two different values appear. (Recall in the first step we took care of all sample points where just one value appears.) The next stage is to consider all sample points where just three values appear. The following stage is to take care of all samples where just four different values appear and so on. If $r_0 \le k_0$, the final stage begins by defining $\theta^{r_0-1}(\lambda_1, \dots, \lambda_{r_0})$ and proceeding as before. If $r_0 > k_0$, the process stops after all samples in which k_0 different values appear have been considered. In either case this completes the proof since we have demonstrated that e^* is a unique stepwise Bayes estimator. □

Because the theorem holds for every possible choice of $\theta(\boldsymbol{\lambda})$, we have the following corollary using the reasoning of Section 3 of Meeden and Ghosh (1983).

COROLLARY. *For estimating the population total with squared error loss, with the parameter space* $\theta = \mathbb{R}^N$ *and for a given design* Δ, *the estimator* e^* *defined in* (2.3) *is admissible.*

As with every admissibility proof in finite population sampling, the result is essentially independent of the design, see, for example, Scott (1975).

Given that one will use the estimator e^*, how should one choose an appropriate design? For example, we suppose one can use any design in Γ, the family of designs of fixed sample size n, where n is some fixed positive integer. In this case, one would want the estimator e^* and the design chosen to be uniformly admissible relative to Γ. Mazloum and Meeden (1987) gives a method for finding such designs when the estimator being considered is a stepwise Bayes estimator against some sequence of priors, say, $\lambda^1, \dots, \lambda^w$. First one must find all those members Δ' of Γ with the property that $R(\Delta', e^*, \lambda^1) = \inf\{R(\Delta, e^*, \lambda^1) | \Delta \in \Gamma\}$, where $R(\Delta, e^*, \lambda^1)$ is the Bayes risk of the pair (Δ, e^*) against the prior λ^1. We denote this class by Γ^1 and repeat this minimizing step with Γ^1 and λ^2. In fact we repeat this process w times, once for each prior in our sequence, to find Γ^w. Then (Δ, e^*) is uniformly admissible relative to Γ if and only if the pair is uniformly admissible relative to Γ^w.

We will now implement the above procedure for the problem at hand. For the parameter space $\theta = \theta(\lambda)$, it is easy to check that for the first prior in the proof of the theorem, Γ^1 is just Γ. For the next step, our prior is concentrated on the set $\theta^1(a, b)$, where $a \ne b$. For $\mathbf{y} \in \theta^1(a, b)$, let $u(\mathbf{y})$ be the unique integer contained in the set $\{1, \dots, N-1\}$ such that $y_i \ne y_j$ if and only if $i \le u(\mathbf{y}) < j$. Let $s = \{i_1, \dots, i_n\}$ denote a sample of size n, where $1 \le i_1 < \cdots < i_n \le N$. For $\mathbf{y} \in \theta^1(a, b)$ with $u(\mathbf{y}) = u$, we have that

$$\frac{r(\mathbf{y}; s, e^*)}{(b-a)^2} = \begin{cases} u, & \text{if } u < i_1, \\ N - u, & \text{if } u \ge i_n, \\ [u - (i_k + i_{k-1} - 1)/2]^2, & \text{if } i_k \le u < i_{k+1} \end{cases} \tag{2.8}$$

for $k = 1, \dots, n-1$. Furthermore, if λ^2 denotes the second prior used in the proof of the theorem, then one easily finds that

$$\begin{aligned}&(N-1)R(s,e^*,\lambda^2)/(b-a)^2\\ (2.9)\qquad &= (i_1-1)i_1/2 + (N-i_n)(N-i_n+1)/2 + \sum_{k=1}^{n-1} d_k h(d_k),\end{aligned}$$

where $d_k = i_{k+1} - i_k$ and $h(x) = (x^2 + 6x + 2)/12$. Hence for a fixed n and N, one must find the set of samples s which minimize (2.9). For example, when $n = 4$ and $N = 21$, there are two such sets, $\{6, 9, 12, 15\}$ and $\{7, 10, 13, 16\}$, when $n = 4$ and $N = 22$ just one set, $\{7, 10, 13, 16\}$, and when $n = 4$ and $N = 23$, there are three such sets, $\{7, 10, 13, 17\}$, $\{7, 11, 14, 17\}$ and $\{7, 10, 14, 17\}$.

These sets are essentially symmetric and are what one might expect except that the smallest and largest members seem to be quite far from their respective endpoints in the population. The reason for this can be seen in equation (2.8), where we note that the risk function, as a function of u, is linear at each end but a sequence of quadratic pieces in the middle.

The search for the minimizing samples can be simplified by assuming that i_1, i_n and the d_k's need not be integer valued. Then for a fixed i_1 and i_k, we should choose the d_k's to be all equal since the summation in (2.9) is Shur convex [see Arnold (1986)]. If we now let $i_1 = x$ and $i_n = N + 1 - x$ and $d_k = (N + 1 - 2x)/(n - 1)$, then equation (2.9) becomes a cubic in x whose minimum can be found approximately. For example, when $N = 100$ and $n = 11$, the minimum value of this cubic is just less 18. Then a local search shows that any s with $i_1 = 18$, $i_{11} = 83$ and five differences equal to six and five differences equal to seven minimizes (2.9). Since these samples are essentially the same, we will not continue the process any further.

The above suggests that when using the estimator e^*, one need only consider nonrandom samples whose labels are approximately equally spaced throughout the population. In some cases, however, it may be desirable to choose a sample whose labels are not equally spaced. A statistician could choose fewer units in regions where it was believed the units were quite similar and more units in regions with more variability. For example, it may be the case that y_i is approximately an increasing function of i, however, in some regions it is believed to be nearly constant while in other regions it increases quite rapidly. Formally, one could divide the population into several regions and use the estimator e^* within each region separately. More informally, one could choose the sampled units to reflect prior beliefs about variability within various regions of the population. In particular, one may not wish to choose the first and last units of the sample as far from the population endpoints as the theory suggests. This would be the case when one's prior knowledge about the units close to the two ends is not as reliable as units closer to the center of the population.

Presumably either the estimator given in (2.3) or others very similar to it have been considered in the past. However, the author does not know of any explicit references to this estimator. If the labels appearing in the sample are nearly equally spaced throughout the population, then the estimator in (2.3) will be approximately $(N/n(s))\Sigma_{i \in s} y_i$, the usual estimate of the population total. However, when the labels appearing in the sample are not equally spaced, the estimator can be quite different from the usual estimator.

REMARK. One can formulate a decision problem such that the admissibility argument given above can be modified to demonstrate the admissibility of the trapezoid rule as an estimate of an integral of some function f. This proof is both simpler and more complicated than the one given above. It is simpler because the grid of points where f is to be observable is fixed and hence we need not worry about a sampling design. It is more complicated because the parameter space contains an uncountable number of points.

3. Some examples with interval estimates.

In this section, we will consider some examples to see how the linear interpolator works in practice. Most practitioners prefer an interval estimate to a point estimate, however. We will show how admissible set estimators can be constructed from the posteriors which yield the linear interpolator and compare them to standard frequentist methods.

In Meeden and Vardeman (1985), it was shown how to find admissible Bayesian credible sets for a given prior. One must choose a positive constant, say c, and then given the data, one's announced set is all the points in the parameter space whose posterior probability density function exceeds c. Note that this does not yield the usual highest posterior density region; see, for example, Berger (1985), since the posterior probability assigned to the announced set can vary with the data. However, in some cases, there exists a choice of c such that the announced probability is approximately 0.95 for all data points. Furthermore this admissible set estimation procedure is well approximated by removing the upper and lower 0.025 tails from the posterior and taking the remaining interval as your announced set. In Meeden and Vardeman (1991), the above argument was extended to the stepwise Bayes setup as well. To summarize, the sequence of priors defined in Section 2 can be used to construct admissible credible sets for estimating the population total or the population mean. One such collection of sets can be found approximately by lopping off the upper and lower 0.025 tails from the posterior. (Since the posterior cannot be found explicitly it must be simulated and it is this simulated posterior which has its tails lopped off.) Hence it is of interest to determine the frequentist converage probabilities of this 95% procedure. One would expect that for parameter points which are sufficiently smooth, that is, for populations whose members, with labels close together, are more alike than members with labels far apart; its coverage probability should be about 0.95. For less smooth parameter points, the coverage probability could be considerably less. In what follows this approximate 95% stepwise Bayes procedure will be compared to standard frequentist procedures for various populations.

Jessen [(1978), pages 18 and 19] discussed the following experiment due to F. Yates. The population was a collection of 126 stones of various sizes displayed on a table. Selectors were asked to purposely pick representative samples of various sizes. Then the estimates of the population mean based on their samples were compared to estimates based on random samples of the same size. It was found that the selectors did better than random sampling if the sample size was less than eight while random sampling did better for sample sizes greater than eight. In some sense this is quite a surprising result and suggests how difficult it can be to attempt to incorporate prior information in a naive fashion.

One approach to this problem is to use the techniques of this paper. For example, one could attempt to arrange the stones in increasing order from the smallest to the largest and then use the linear interpolator as one's estimate. If the ordering is reasonably accurate, then the linear interpolator should perform better than the usual estimate based on random sampling.

A frequentist sampler, faced with a population which was known to be smooth, would not use simple random sampling but would stratify. In the examples that follow, all the populations will be stratified. A sample will be drawn by choosing one unit at random from each stratum. A 95% confidence interval will be constructed using the method of collapsed strata; see Cochran [(1977, page 139]. This is done by assuming the usual estimate is approximately normally distributed and getting an estimate of its variance by pairing successive strata. (Cochran notes that in some cases this can lead to an overestimate of the sampling variance of the estimator.) This will be called the ST interval and will be compared to the approximate 95% stepwise Bayes procedure denoted by LI (linear interpolator) interval. This will be done for six different populations.

The first population, POP 1, was constructed from the set $A = \{a(1), \dots, a(100)\}$ of real numbers. $a(\cdot)$ was a piecewise linear function consisting of seven different pieces. The six change points were after the points 6, 16, 37, 66, 86 and 96. Then, y_i was set equal to $a(i)$ plus a realization of a normal random variable with mean 0 and variance $(i + 3)/4$. In Figure 1, POP 1 is plotted against its labels. Note that even though POP 1 is reasonably smooth, one should be able to order 100 stones on a table as well as this or perhaps even better if given sufficient time. POP 1 was divided into eleven strata. The first stratum contained the first five units while the last contained the last five units and the remaining nine strata each contained ten successive units. Five hundred stratified samples were taken where one unit was taken at random from each stratum. The results are given in the first row of Table 1. We see that on the average, the ST interval is almost twice as long as the LI interval. Furthermore, it covers almost all the time while the LI interval's frequency of coverage is exactly 0.95 in this case. In retrospect, this is not too surprising since the ST interval makes no use of the smooth nature of POP 1, beyond the stratification, while the LI interval is based on the smoothness assumption.

A second population, POP 2, was constructed from the set A. This time y_i was taken to $a(i)$ plus a realization of a normal random variable with mean 0 and variance $(i + 3)$. This population is plotted in Figure 2 and is not nearly as

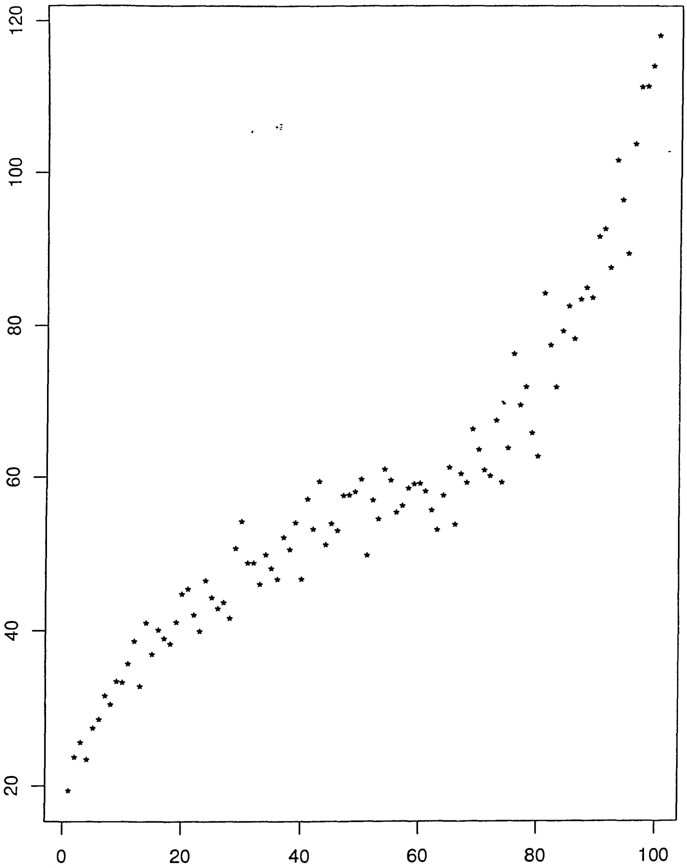

FIG. 1. *Plot of population* POP 1.

TABLE 1

	Number of samples	Average length of ST interval	Frequency of coverage	Average length of LI interval	Frequency of coverage
POP 1	500	9.12	0.992	4.61	0.95
POP 2	200	11.35	0.92	5.21	0.71
POP 3	500	21.64	0.986	8.74	0.718
POP 4	500	28.61	0.992	12.59	0.904
POP 5	500	15.95	1.0	7.08	0.916
POP 6	500	9.65	0.994	5.02	0.952

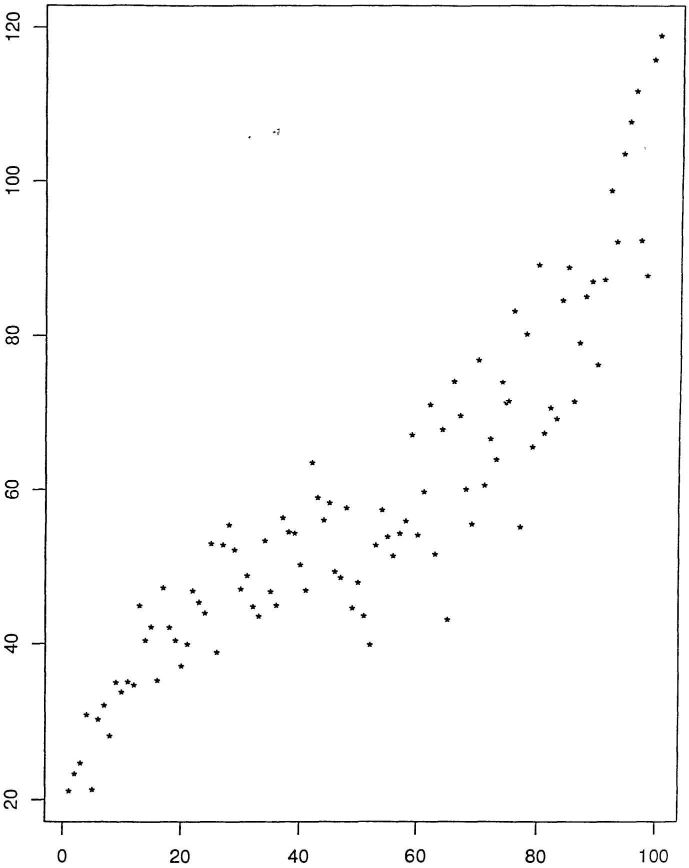

FIG. 2. *Plot of population* POP 2.

smooth as POP 1. It was stratified in the same way as POP 1 and 200 stratified samples were taken. The results are in row 2 of Table 1. As was to be expected the LI interval does poorly since POP 2 is not sufficiently smooth.

The final four populations come from sunspot data discussed in Andrews and Herzberg (1985). These data are successive monthly means of daily sunspot numbers beginning in the year 1749. Because of the cyclic nature of sunspot activity, we would expect this sequence of observations to exhibit some smoothness. POP 3 is the first 100 observations of this sequence, POP 4 the first 500, POP 5 the first 1,000 and POP 6 the first 2,400. POP 3 was stratified in the same manner as POP 1 and POP 2, while the last three populations were divided into 25, 50 and 120 strata, respectively, where each stratum was of length 20. For each of these four populations, 500 samples were taken and the two intervals computed. The results are given in the last four lines of Table 1. We see that the ST interval is too long and overcovers. On the other hand, the LI interval is considerably shorter but only yields the appropriate frequency of coverage when the population gets sufficiently large. This indicates that the degree of smoothness of a population depends in part on how many units are in the population.

In Table 1, we did not include the average value for the point estimates for the two different methods. This is because in each case the average differed by less than 0.1. If the strata were of varying sizes, this would no longer be the case.

In the chapter on systematic sampling, Cochran (1977) discussed approaches to estimation for some naturally occurring populations like the sunspot data. In such cases it is often very difficult to get a sensible estimate of variance and usually such estimates depend on an assumed model for the data. Note that to use the LI interval, one only needs to assume the population is smooth and have some idea about the length of the sunspot cycles. This is important since the sunspot cycles need to be much longer than the sampling strata, so that the population is approximately linear between observations.

We choose the ST interval as the one to compare to the LI interval because we thought it would be the strongest competitor. For another possibility, consider population POP 6 and divide it into 60 strata each of size 40 and choose two units at random from each stratum. Five hundred samples were taken and the usual interval computed. The average length was 7.98 and 0.928 was the frequency of coverage. This is not as good as the LI interval discussed above. This is not surprising since, on the average, choosing two observations out of every 40 units should not be as representative as choosing one out of every 20.

A referee noted that the sequence of priors used in the proof concentrates mass on a rather thin subset of the parameter space. So even though the linear interpolator (LI) is admissible, it may be of limited practical value because it will be nonrobust. Furthermore, it was suggested that a prior or sequence of priors concentrating on a larger subset of the parameter space could yield a better estimator. Recall that the LI estimator and the LI interval arise from a very intuitively appealing pseudo posterior distribution. If one believes that the population is smooth, then, given the sample, this pseudo posterior puts equal mass over a set of simple and essentially smooth parameter points consistent with the sample. Since this is true no matter what the actual values appearing in the sample are, the estimator will perform well over a wide class of smooth populations. We believe that this is demonstrated by the results of Table 1. This is not to say that a prior with larger support could not be more robust, but it would be surprising if it could also outperform the LI estimator for smooth populations. However, at this time it is not clear what a sensible choice for such a prior would be.

Dealing with smooth populations is difficult not only from the straightforward Bayesian point of view but from the frequentist point of view as well. Beyond stratifying and taking a random sample of size 1 or 2 from each stratum, it is not clear how a frequentist would proceed. Furthermore, it is just these situations where the frequentist theory is the most unsatisfying. Choosing the labels of a finite population to make it smooth can be thought of as a generalization of stratification, that is, a non-Bayesian method of incorporating prior information. However, once the sample is chosen, one can use the pseudo posterior distribution to find admissible point and set estimates. Finally, if the population is indeed smooth, then the set estimator will have good frequentist properties as well.

Acknowledgment. The author wishes to thank a referee for several helpful comments and suggestions.

REFERENCES

ANDREWS, D. F. and HERZBERG, A. M. (1985). *Data: A Collection of Problems From Many Fields for the Student and Research Worker*. Springer, New York.

ARNOLD, B. C. (1986). *Majorization and the Lorenz Order: A Brief Introduction*. Springer, New York.

BASU, D. (1969). Role of sufficiency and likelihood principles in sample survey theory. *Sankhyā Ser. A* **31** 441–454.

BERGER, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. Springer, New York.

BROWN, L. D. (1981). A complete class theorem for statistical problems with finite sample spaces. *Ann. Statist.* **9** 1289–1300.

COCHRAN, W. G. (1977). *Sampling Techniques*. Wiley, New York.

HSUAN, F. C. Y. (1979). A stepwise Bayesian procedure. *Ann. Statist.* **7** 860–868.

JESSEN, R. A. (1978). *Statistical Survey Techniques*. Wiley, New York.

JOHNSON, B. M. (1971). On the admissible estimators for certain fixed sample binomial problems. *Ann. Math. Statist.* **42** 1579–1587.

MAZLOUM, R. and MEEDEN, G. (1987). Using the stepwise Bayes technique to choose between experiments. *Ann. Statist.* **15** 269–277.

MEEDEN, G. and GHOSH, M. (1981). Admissibility in finite problems. *Ann. Statist.* **9** 846–852.

MEEDEN, G. and GHOSH, M. (1983). Choosing between experiments: Application to finite population sampling. *Ann. Statist.* **11** 296–305.

MEEDEN, G. and VARDEMAN, S. (1985). Bayes and admissible set estimation. *J. Amer. Statist. Assoc.* **80** 465–471.

MEEDEN, G. and VARDEMAN, S. (1991). A noninformative Bayesian approach to interval estimation in finite population sampling. *J. Amer. Statist. Assoc.* **86** 972–980.

SCOTT, A. J. (1975). On admissibility and uniform admissibility in finite population sampling. *Ann. Statist.* **3** 489–491.

VARDEMAN, S. and MEEDEN, G. (1984). Admissible estimators for the total of a stratified population that employ prior information. *Ann. Statist.* **12** 675–684.

SCHOOL OF STATISTICS
270 VINCENT HALL
UNIVERSITY OF MINNESOTA
206 CHURCH STREET
MINNEAPOLIS, MINNESOTA 55455