A USEFUL GENERALIZATION OF THE STEIN TWO-SAMPLE PROCEDURE

By R. Wormleighton

University of Toronto

1. Introduction. An experimenter wishes to estimate the mean, μ, of a normal distribution using a sample mean. If the variance, σ^2, is known, the size of a single sample can be functionally related to the precision of the sample estimate, e.g.: $n^{\frac{1}{2}} = (1.96)\sigma/L$, where n is the sample size and $2L$ is the length of a 95% confidence interval for μ. The experimenter can choose in advance the point on the curve which provides a satisfactory balance between the cost of obtaining the sample and the precision of the final estimate.

If the variance is not known, and no reasonable estimate can be obtained, several simple procedures are presently available.

(i) The experimenter takes as large a sample as he can afford. The estimate of the mean has maximum precision, but it may be more precise than he requires.

(ii) He takes a preliminary sample to get an estimate of the variance. On the basis of this variance estimate, he decides on the size of a second sample. His estimate of μ is the mean of the second sample and its precision is determined from the second sample by the usual single-sample procedure. This method is wasteful of the information in the first sample.

(iii) He can use a Stein two-sample procedure. Here, he specifies the precision of his final estimate in advance, and the total number of observations becomes a random variable. This is often unattractive because the cost of the experiment is not pre-determined and may turn out to be excessive.

The experimenter would like to take a first sample to get a variance estimate, then decide on the total number of observations and the precision of the estimate of the mean, and finally use *all* his data in making the estimate. This can very nearly be accomplished by the generalized Stein procedure described below.

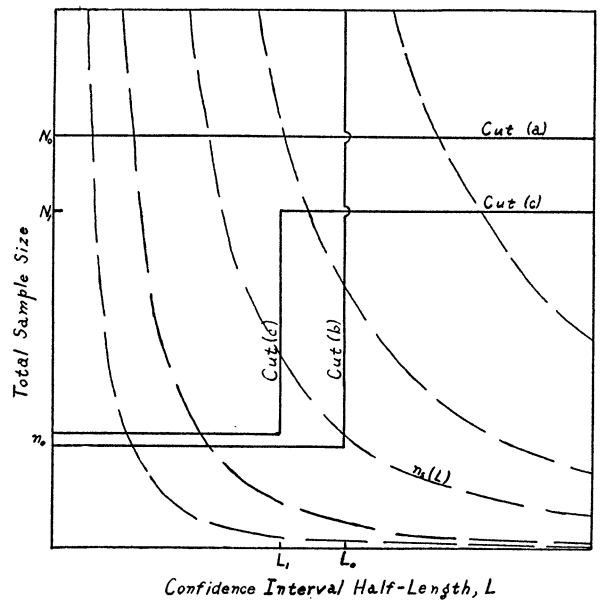
2. Procedure. We are given a normal population with unknown mean, μ, and unknown variance, σ^2. Consider a first sample of n_0 observations: $x_1, x_2, \cdots, x_{n_0}$. An estimate of σ^2, based on the first sample, is

$$s^2 = \frac{1}{n_0 - 1}\left\{\sum_1^{n_0} x_i^2 - \frac{1}{n_0}\left(\sum_1^{n_0} x_i\right)^2\right\}.$$

Corresponding to any particular value of s, we can plot the curve

$$n_s(L) = \left[\frac{t^{(\alpha)}s}{L}\right]^2$$

Received June 20, 1958.



Total sample size n vs. confidence interval half-length L
Cut (a): Fixed sample size
Cut (b): Modified Stein procedure
Cut (c): Bounded sample size

FIG. 1

where n = total sample size, $2L$ = length of a $(1 - \alpha)$ confidence interval, and $t^{(\alpha)}$ = $(1 - \alpha)$ point of a t-variate with $(n_0 - 1)$ degrees of freedom. We thus have a family of such curves which do not intersect. (See Fig. 1.)

(i) On each curve choose, in advance, a single point. This set of points constitute a "cut" across the family of curves.

(ii) Now actually take the first sample and calculate s. This determines a particular curve of the family and, because of the cut, a unique point, (n^*, L^*), say.

Remark: There appears to be no practical advantage and, in fact, a waste of information if $n^* < n_0$. We therefore exclude cuts which permit this situation to occur.

(iii) Take $[n^* - n_0] + 1$ further observations, where $[q]$ denotes the largest integer strictly less than q.

(iv) Calculate $\bar{x}$, the mean of *all* the observations. Then, $\bar{x} \pm L^*$ is a $(1 - \alpha)$ confidence interval for μ.

Remark: If the cut is such that n^* is not an integer, the exact confidence coefficient is slightly higher than $(1 - \alpha)$. The approximation can be avoided either by excluding cuts which might yield non-integral values of n^*, or by giving the last observation a smaller weight in calculating the sample mean, $\bar{x}$. See Section 5.

3. Proof. The proof is that given by Stein. Let $U = (\bar{x} - \mu)\sqrt{n^*}/s$. For a given cut, the value of n^* depends only on s which is statistically independent of $\bar{x}$. Hence, the conditional distribution of U, given s, is $N(0, \sigma^2/s^2)$.

Consider a variate $T = y/s$, where Y is $N(0, \sigma^2)$ and statistically independent of s. The conditional distribution of T, given s, is also $N(0, \sigma^2/s^2)$. U and T are therefore identically distributed. But T, and hence U, has a t distribution with $(n_0 - 1)$ degrees of freedom.

$$\begin{aligned}(1 - \alpha) &= \Pr\{-t^{(\alpha)} \leqq U \leqq t^{(\alpha)}\} \\ &= \Pr\left\{\bar{x} - \frac{t^{(\alpha)}s}{\sqrt{n^*}} \leqq \mu \leqq \bar{x} + \frac{t^{(\alpha)}s}{\sqrt{n^*}}\right\} \\ &= \Pr\{\bar{x} - L^* \leqq \mu \leqq \bar{x} + L^*\}.\end{aligned}$$

4. Possible cuts.

(a) *Fixed sample size*: (Line (a) in Figure.) The cut is defined by $n = N_0$, a constant. ($N_0 \geqq n_0$). The length of the confidence interval is a random variable. If $N_0 = n_0$, we have the usual single-sample procedure. If $N_0 > n_0$, the procedure differs from the single-sample procedure in that the variance estimate is based on fewer degrees of freedom—$(n_0 - 1)$ instead of $(N_0 - 1)$. The length of the confidence interval will thus have larger expectation and larger variance than one calculated from a single-sample.

(b) *A modified Stein method*: (Line (b) in Figure.) A confidence interval length, $2L_0$, is preassigned. The cut is defined by

$$\begin{aligned} n &= n_0 & 0 < L \leqq L_0 \\ L &= L_0 & n_0 \leqq n < \infty. \end{aligned}$$

In Stein's exact, although not in his approximate procedure, a second sample of at least one observation is always required, and a weighted average is used to estimate μ. If s should turn out to be so small that only one additional observation is required, the last observation is given an excessive weight so that the precision of the estimate of μ is actually reduced. This device ensures that, even in this situation, the pre-assigned confidence interval length, $2L_0$, is obtained. See Section 5.

In our modification, defined by cut (b), we use an unweighted average. When s turns out to be so small that no further observations are required, we obtain a confidence interval shorter than we anticipated, and the method reduces to the usual single-sample procedure.

(c) *Bounded sample size*: (Cut (c) in Figure.) An experimenter would like to have a $(1 - \alpha)$ confidence interval of length $2L_1$, but is not willing to take more than N_1 observations. He may then use the cut

$$\begin{aligned} n &= n_0 & 0 < L \leqq L_1 \\ L &= L_1 & n_0 \leqq n \leqq N_1 \\ n &= N_1 & L \geqq L_1. \end{aligned}$$

He will then obtain the desired precision, or better, with the minimum number of observations, if this number is less than N_1 ; otherwise, he will take N_1 observations and settle for the precision he gets.

(d) *Delayed decision*: Conceivably, one could define a cut by considering, in advance of the first sample, each possible value of s, and choosing a point on each of the curves, $n_s(L)$. It seems superfluous, however, to make a large number of decisions when only one will be implemented. A possible procedure would be to take the first sample and calculate s, and only then make the decision on the basis of the one resulting curve. It is absolutely necessary that the decision not be influenced by the first sample mean; otherwise, n^* is not independent of $\bar{x}$ and the proof in Section 3 is not valid. This condition could be ensured when the decision-maker does not himself collect or analyze the data, or even see them, by informing him only of the observed value of s. It can then be argued that the decision on sample size is identical with that which would have resulted from a cut completely defined in advance of the first sample.

This procedure can be compared with one in which the experimenter has an independent estimate of the variance available to him when he is planning the experiment. If he intends to use the preliminary variance estimate for calculating the length of his confidence interval, and not the variance of the single sample he plans, then he can predetermine both sample size and confidence inteval length as in the case where the true variance is known. In our procedure, the preliminary variance estimate is obtained from part of the sample, but it is nevertheless independent of the sample mean.

Throughout the discussion, we have assumed that the confidence interval length, $2L$, is given in absolute units, the same as those of the observations. Scientists and engineers frequently prefer to specify the error as a percentage of the mean, and in order to convert absolute error to an approximate percentage error an estimate of the mean is required. There is a temptation in a delayed-decision procedure to use the first-sample mean for this purpose. This clearly is not permissible because the decision would be influenced by the first-sample mean.

(e) *Minimum cost*: (Suggested by referee.) If $c_1(n)$ is the cost of taking n observations and $c_2(L)$ is the cost of an interval of length $2L$, then a cut can be defined by the minimization of the total cost, $c_1(n) + c_2[t^{(\alpha)}s/\sqrt{n}]$, with respect to n, for given s. Since total cost is minimized for every s, the expected total cost is a minimum.

5. Weighted mean. In the exact Stein procedure, a weighted mean of the observations provides an estimate of μ. The weights depend on s, and must satisfy the conditions:

(i) Sum of all weights = 1.

(ii) First-sample observations are each given the same weight, a.

(iii) Sum of squares of the weights = z/s^2 where z is a pre-assigned constant. It is also possible and desirable to require the condition.

(iv) All weights are non-negative.

The total sample size, n, is determined by

$$n = \max\left\{\left[\frac{s^2}{z}\right] + 1, n_0 + 1\right\}.$$

Case 1. If s^2 turns out so small that $z/s^2 = (1/n_0) + \epsilon$, $\epsilon > 0$, then only one additional observation is made, which is given the weight $(1 - n_0 a)$ to satisfy condition (i). Condition (iii) requires that $n_0 a^2 + (1 - n_0 a)^2 = (1/n_0) + \epsilon$. The quadratic in a has two solutions, but one of them makes $(1 - n_0 a)$ negative and is inadmissible by (iv). The other solution is

$$a = \frac{1}{n_0 + 1}\left\{1 - \frac{1}{n_0}\sqrt{1 + \epsilon n_0(n_0 + 1)}\right\} < \frac{1}{n_0 + 1}$$

$$1 - n_0 a = \frac{1}{n_0 + 1}\left\{1 + \frac{1}{n_0}\sqrt{1 + \epsilon n_0(n_0 + 1)}\right\} > \frac{1}{n_0 + 1}$$

Hence, the weighting is uniquely determined. The last observation is given an excessive weight, thereby reducing the precision to the pre-assigned value.

Case 2. $s^2/z > n_0$. Then, $n = [s^2/z] + 1$. Let

$$\delta = \frac{s^2}{z} - \left[\frac{s^2}{z}\right].$$

There is considerable freedom in the choice of weights but we can require, for simplicity, that the first $(n - 1)$ observations be given the same weight, a. Then the last observation receives weight, $1 - (n - 1)a$.

Under this restriction there are still two admissible solutions, one of which is

$$a = \frac{1}{n} + \frac{1}{n}\sqrt{\frac{1}{(n - 1)(n - 1 + \delta)}}$$

$$1 - (n - 1)a = \frac{1}{n} + \frac{n - 1}{n}\sqrt{\frac{1}{(n - 1)(n - 1 + \delta)}}.$$

This gives the last observation a reduced weight. One can interpret this weighted average as a simple mean of all the observations except for a fraction of the last observation—in effect, making the total sample size a continuous variable.

REFERENCE

[1] CHARLES STEIN, "A two-sample test for a linear hypothesis whose power is independent of the variance," *Ann. Math. Stat.*, Vol. 16 (1945), pp. 243–258.