A MINIMAX PROPERTY OF THE SAMPLE MEAN IN FINITE POPULATIONS[1]

By P. J. Bickel and E. L. Lehmann

University of California, Berkeley

Consider the problem of estimating the mean of a finite population on the basis of a simple random sample. It was proved by Aggarwal (1954) that the sample mean minimizes the maximum expected squared error divided by the population variance τ^2. Aggarwal also stated, but did not successfully prove, that the sample mean minimizes the maximum expected squared error over the populations satisfying $\tau^2 \le M$ for any fixed positive M. It is the purpose of this paper to give a proof of this second result, and to indicate some generalizations.

1. Introduction. Consider a population $\{a_1, \cdots, a_N\}$ of real numbers. In order to estimate the population average, $\bar{a}$, a simple random sample of size n is drawn, say $Y_1, \cdots, Y_n$. We are concerned with the problem of determining an estimator $\delta = \delta(Y_1, \cdots, Y_n\}$ which minimizes the maximum expected squared error

$$(1) \qquad \sup E(\delta - \bar{a})^2.$$

This problem is meaningless unless limitations are placed on the a's under which this minimax risk is finite. Two natural restrictions are
   (i) $A \le a_i \le B$ where A, B are given
   (ii) $\sum (a_i - \bar{a})^2 \le M$ where M is given.
A third possibility is
   (iii) to minimize instead of (1)

$$(2) \qquad \sup E\{(\delta - \bar{a})^2 / \sum (a_i - \bar{a})^2\},$$

in which case no restriction on the a's is required.

A minimax estimator for (i) was recently obtained by Hodges and Lehmann (1981). For problems (ii) and (iii), Aggarwal (1959) stated that the sample mean $\bar{Y}$ has the desired minimax properties. If

$$(3) \qquad a_i = \theta + d_i, \qquad \theta = \bar{a}, \qquad d_i = a_i - \bar{a}$$

Aggarwal obtained the Bayes estimator δ_σ of θ when θ is $N(0, \sigma^2)$ and the d's have a suitable singular multivariate normal distribution, and showed that the risk of δ_σ tends to the risk of $\bar{Y}$ as $\sigma \to \infty$. His argument does prove property (iii) but does not establish (ii) since the prior distribution of the a's violates (ii). The purpose of the present paper is to show that $\bar{Y}$ is in fact minimax when the a's are subject to the restriction

$$(4) \qquad \sum (a_i - \bar{a})^2 = 1.$$

It then immediately follows when 1 is replaced in (4) by an arbitrary M, and this establishes both (ii) and (iii).

---

2. Reduction of the minimax problem.   In order to minimize (1) subject to (4), consider the auxiliary problem in which the a's are given by (3) with θ unknown and the d's distributed according to a distribution Q over the intersection of the sphere $\sum d_i^2 = 1$ with the hyperplane $\sum d_i = 0$ to be specified below. We shall show that $\bar{Y}$ is a minimax estimator of θ for this auxiliary problem in which θ is the only unknown parameter, i.e., that $\bar{Y}$ minimizes

$$
(5) \qquad \sup_\theta \int E(\delta - \theta)^2 \, dQ(d_1, \cdots, d_N).
$$

Since for $\delta = \bar{Y}$, (5) is equal to

$$
(6) \qquad \sup_\theta \sup_{\sum(d_i - \bar{d})^2 = 1} E(\delta - \theta)^2,
$$

it will follow that $\bar{Y}$ is also minimax for the original problem under the restriction (4).

In the auxiliary problem if Q is invariant under permutation of the d_i, we can assume without loss of generality that

$$
(7) \qquad Y_i = \theta + D_i \qquad\qquad i = 1, \cdots, n,
$$

where the joint distribution of $(D_1, \cdots, D_N)$ is Q. The problem of estimating θ on the basis of the variables (7) is that of estimating a location parameter θ when the joint density of the variables is known except for θ. It was shown by Girshick and Savage (1951) that the Pitman estimator is then minimax provided an estimator with finite risk exists. It will therefore be enough to show that $\bar{Y}$ is the Pitman estimator on the basis of (7).

Let us next specify the distribution of the D's. Let $X_1, \cdots X_N$ be i.i.d. $N(0, 1)$ and let

$$
(8) \qquad U_i = \frac{X_i - \bar{X}}{S_N},
$$

where $\bar{X} = (1/N) \sum_{i=1}^N X_i$, $S_N^2 = \sum_{i=1}^N (X_i - \bar{X})^2$. Then the U's satisfy $\sum U_i^2 = 1$ and $\sum U_i = 0$ and we shall take as distribution of $(D_1, \cdots, D_N)$ that of $(U_1, \cdots, U_N)$ so that the joint distribution of the variables $Y_1, \cdots, Y_n$ is the same as that of $U_1 + \theta, \cdots, U_n + \theta$.

The Pitman estimator of θ in this last representation is given by

$$
(9) \qquad \bar{Y} - E_0[\bar{Y} \mid Y_1 - \bar{Y}, \cdots, Y_n - \bar{Y}].
$$

Since the second term in (9) is equal to

$$
(10) \qquad E_0[\bar{U}_n \mid U_1 - \bar{U}_n, \cdots, U_n - \bar{U}_n]
$$

where $\bar{U}_n = (U_1 + \cdots + U_n)/n$, it is enough to show that (10) is equal to zero.

3. A symmetry property.   To show that (10) is equal to zero, we shall now prove the stronger property that

$$
(11) \qquad \mathscr{L}(\bar{U}_n \mid U_1 - \bar{U}_n, \cdots, U_n - \bar{U}_n) =' \mathscr{L}(-\bar{U}_n \mid U_1 - \bar{U}_n, \cdots, U_n - \bar{U}_n).
$$

Condition (11) will follow if we show that

$$
(12) \qquad \mathscr{L}(\bar{X}_n - \bar{X} \mid X_1 - \bar{X}_n, \cdots, X_n - \bar{X}_n, S_N^2)
$$

is symmetric about zero since symmetry about zero is preserved under mixing with respect to a scale parameter.

Let $(X_1, \cdots, X_n) \to (Z_1, \cdots, Z_n)$ be an orthogonal transformation such that $Z_1 = \sqrt{n}\, \bar{X}_n$ and let $Z_{n+i} = X_{n+i}$ $(i = 1, \cdots, N - n)$. Then

$$
\bar{X}_n - \bar{X} = \frac{N - n}{n} \left(\frac{Z_1}{\sqrt{n}} - \bar{Z} \right)
$$

where $\tilde{Z} = (Z_{n+1} + \cdots + Z_N)/(N - n)$, and a straight-forward calculation shows that

$$S_N^2 = \sum_{i=2}^n Z_i^2 + \sum_{i=n+1}^N (Z_i - \tilde{Z})^2 + \frac{n(N - n)}{N}\left(\tilde{Z} - \frac{Z_1}{\sqrt{n}}\right)^2.$$

Consider now the transformation

(13) $(Z_1, \cdots, Z_N) \to (-Z_1, Z_2, \cdots, Z_n, -Z_{n+1}, \cdots, -Z_N)$

and note that it has the following three properties:
  (i) Under (12)

$$\frac{N - n}{N}\left(\frac{Z_1}{\sqrt{n}} - \tilde{Z}\right) \to -\frac{N - n}{N}\left(\frac{Z_1}{\sqrt{n}} - \tilde{Z}\right)$$

  (ii) the transformation (13) leaves S_N^2 invariant.
  (iii) The distribution of $(Z_1, \cdots, Z_N)$ is not changed by (13).
The distribution (12) in terms of the Z's becomes the conditional distribution

$$\mathscr{L}\left(\left(\frac{Z_1}{\sqrt{n}} - \tilde{Z}\right)\frac{N - n}{n} \mid Z_2, \cdots, Z_n, S_N^2\right)$$

which by (i) $-$ (iii) is seen to be equal to

$$\mathscr{L}\left(-\left(\frac{Z_1}{\sqrt{n}} - \tilde{Z}\right)\frac{N - n}{n} \mid Z_2, \cdots Z_n, S_N^2\right).$$

Re-expressing this last distribution in terms of the X's completes the proof.

4. Some extensions.

(i) *Labels.* The minimax result proved in the preceding sections did not take into account the labels of the population elements, which potentially provide additional information. It is however easy to see that the result remains valid even when the labels of the sampled elements are considered as part of the data. The population then consists of the N pairs (i, a_i), $i = 1, \cdots, N$ and the data of the n sampled pairs. The problem of estimating $\bar{a}$ remains invariant under permutation of the labels. Since the estimand $\bar{a}$ remains invariant under these transformations, an estimand $\delta((i_1, a_{i_1}), \cdots, (i_n, a_{i_n}))$ (where $i_1, \cdots, i_n$ are the labels of the n sample elements) is invariant if it remains unchanged under all permutations of the first arguments, i.e., if it does not depend on the labels. Since the group of all permutations is finite there exists an invariant minimax estimator, and this completes the proof.

(ii) *Stratified sampling.* Suppose the population has been divided into s strata of sizes $N_1, \cdots, N_s$ from which simple random samples of sizes $n_1, \cdots, n_s$ are drawn. Let Y_i and $\theta_i = a_i$ be the mean of the sample and of the population of the ith stratum respectively, and let $a_{\cdot\cdot} = \sum N_i a_i / \sum N_i$ be the average of all $N = \sum N_i$ population elements a_{ij} $(j = 1, \cdots, N_i; i = 1, \cdots, s)$. The natural estimator for estimating $a_{\cdot\cdot}$ is $\delta = \sum (N_i/N) Y_i$. and its variance is

$$\bar{\tau}^2 = \sum \frac{N_i^2}{N^2} \frac{N_i - n_i}{N_i - 1} \frac{1}{n_i} \tau_i^2,$$

where τ_i^2 is the population variance of the ith stratum. Then in generalization of the result for simple random sampling one can conclude that, subject to $\bar{\tau}^2 \le M$, δ is minimax for estimating $a_{\cdot\cdot}$ with squared error loss.

The result follows immediately from the following two facts, for both of which the loss is squared error.

(a) Let $(\mathbf{Y}_1, \theta_1), \cdots, (\mathbf{Y}_s, \theta_s)$ be s independent location families. Then the uniformly minimum risk translation equivariant estimator of $\theta = \sum c_i \theta_i$ is $\delta = \sum c_i \delta_i$ where δ_i is the Pitman estimator of θ_i based on $(\mathbf{Y}_i, \theta_i)$

(b) The Girshick-Savage argument extends to the situation described in (a), see for example, Kiefer (1957), and the uniformly minimum risk equivariant of θ is therefore also minimax.

(iii) *Design.* We can use an invariance argument as in Blackwell and Girshick (1954) to get a stronger minimax property.

In the situation of Section (1) let $\mathscr{P}$ be the class of all sampling designs in which a sample of size n is selected from a population of size N, i.e., members of $\mathscr{P}$ are probability measures on the finite set S whose members are the $\binom{N}{n}$ subsets of size n of the integers $\{1, \cdots, N\}$. A procedure for estimating $\bar{a}$ now consists of a pair $P \in \mathscr{P}$ and estimator δ and the risk is, with obvious notation,

$$R(a_1, \cdots, a_N, P, \delta) = E_P[\delta(Y_1, \cdots, Y_n) - \bar{a}]^2$$

$$= \sum P(s)[\delta(\{a_i\} : i \in s) - \bar{a}]^2.$$

It turns out that if P_0 is simple random sampling i.e., the uniform distribution on S, then $(P_0, \bar{Y})$ again minimizes $\max\{R(a_1, \cdots, a_N, P, \delta) : \sum(a_i - \bar{a})^2 \leq M\}$ among all pairs (P, δ) as above. We need only invoke the results of Section 8.7 of Blackwell and Girshick to ensure that, for any fixed δ, (P_0, δ) has minimax risk within the class $\{(P, \delta) : P \in \mathscr{P}\}$ and then apply our main theorem.

We have formulated this argument for populations without labels and for sampling plans where the order in which elements are drawn is disregarded, but it is clear that the minimax property continues to hold under these weaker conditions.

We can also generalize the result to stratified sampling in the obvious way obtaining minimaxity among all sampling plans taking n_i observations from stratum i, $i = 1, \cdots, s$. By a further invariance argument we can get an even more general result (but under more special assumptions). Suppose we are given a total sample size n and we are permitted to allocate $n_1, \cdots, n_s$ ($\sum n_i = n$) observations to the various strata as we please and then choose a sampling plan and estimate $\bar{a}$. If the class of populations considered consists of all those with $\tau_i^2 \leq M$, $i = 1, \cdots, s$, and s divides n, then the optimal (minimax) plan and estimate is to sample $\dfrac{n}{s}$ observations from each stratum without replacement and then estimate $\bar{a}$ by the grand mean of the s samples.

REFERENCES

AGGARWAL, OM P. (1959). Bayes and minimax procedures in sampling from finite and infinite populations. *Ann. Math. Statist.* **30** 206–218.

BLACKWELL, DAVID and GIRSHICK, M. A. (1954). *Theory of Games and Statistical Decisions.* Wiley, New York.

GIRSHICK, M. A. and SAVAGE, L. J. (1951). Bayes and minimax estimates for quadratic loss functions. *Proc. Second Berkeley Symp Math. Statist. Prob.* 53–75.

HODGES, J. L., JR. and LEHMANN, E. L. (1981). Minimax estimation in simple random sampling (To appear in *Essays in Statistics and Probability* in honor of C. R. Rao (P. R. Krishnaiah, ed.). North-Holland.

KIEFER, J. (1957). Invariant minimax sequential estimation and continuous time processes. *Ann. Math Statist.* **28** 573–601.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA, BERKELEY
BERKELEY, CALIFORNIA 94720