ON MINIMUM VARIANCE AMONG CERTAIN LINEAR FUNCTIONS OF ORDER STATISTICS

By K. C. Seal

Calcutta University

1. Summary. Suppose there are n normal populations $N(\mu_i, 1)$, $i = 1, \cdots, n$ and that one random observation from each of these n populations is given. Let $x_1 \leqq x_2 \leqq \cdots \leqq x_n$ be the observations when arranged in order of magnitude and let the corresponding n random variables be denoted by X_i, $i = 1, \cdots, n$.

The following theorem is proved:

THEOREM.

$$\operatorname{Var}\left(\sum_{i=1}^{n} c_i X_i\right), \quad \textit{where}$$

$$\sum_{i=1}^{n} c_i = 1, \tag{1}$$

is minimum when $c_i = 1/n$, $i = 1, \cdots, n$.

The above theorem may be applied to provide a direct proof of the result that $\sum_{i=1}^{n} X_i$ is the best unbiased linear function of order statistics for estimating the sum $\sum_{i=1}^{n} \mu_i$.

2. Proof. Let (σ_{ij}) be the variance-covariance matrix of X_i and X_j, $i = 1, \cdots, n; j = 1, \cdots, n$. The above theorem will follow from the following lemma.

LEMMA 1.

$$\sum_{i=1}^{n} \sigma_{ij} = 1, \qquad j = 1, \cdots, n. \tag{2}$$

PROOF. The joint probability density function (pdf) of $X_1, \cdots, X_n$ can be easily shown (see [2], pp. 12–17) to be given by

$$(2\pi)^{-n/2} \sum_{\tau} \exp\left\{-\tfrac{1}{2} \sum_{i=1}^{n} (x_i - \mu_{t_i})^2\right\} d\xi, \tag{3}$$

$$x_1 \leqq x_2 \leqq \cdots \leqq x_n,$$

where $\tau = (t_1, \cdots, t_n)$ is a permutation of $(1, 2, \cdots, n)$, Σ_τ denotes the summation over $n!$ such permutations and ξ represents the row vector $(x_1, \cdots, x_n)$.

Let g be any differentiable function such that the integrals involved exist and we have identically in u,

$$\begin{aligned} Eg(X_j + u) &= \int_{x_1 \leqq \cdots \leqq x_n} \cdots \int g(x_j + u)(2\pi)^{-n/2} \sum_{\tau} \exp\left\{-\tfrac{1}{2} \sum_{i=1}^{n} (x_i - \mu_{t_i})^2\right\} d\xi \\ &= \int_{x_1 \leqq \cdots \leqq x_n} \cdots \int g(x_j)(2\pi)^{-n/2} \sum_{\tau} \exp\left\{-\tfrac{1}{2} \sum_{i=1}^{n} (x_i - u - \mu_{t_i})^2\right\} d\xi. \end{aligned} \tag{4}$$

Received August 25, 1955.

Differentiating both sides of (4) with respect to u and setting $u = 0$, we obtain

$$
\begin{aligned}
&Eg'(X_j) \\
&= \int_{x_1 \leqq \cdots \leqq x_n} \cdots \int g(x_j)(2\pi)^{-n/2} \sum_{\tau} \left[\sum_{i=1}^{n} (x_i - \mu_{t_i}) \exp\left\{ -\tfrac{1}{2} \sum_{i=1}^{n} (x_i - \mu_{t_i})^2 \right\} \right] d\xi \\
(5) \quad &= \int_{x_1 \leqq \cdots \leqq x_n} \cdots \int g(x_j) \sum_{i=1}^{n} (x_i - \mu_i)(2\pi)^{-n/2} \sum_{\tau} \exp\left\{ -\tfrac{1}{2} \sum_{i=1}^{n} (x_i - \mu_{t_i})^2 \right\} d\xi \\
&= E\left[g(x_j) \sum_{i=1}^{n} (x_i - \mu_i) \right].
\end{aligned}
$$

With $g(x) = x$, equation (5) gives the required lemma

$$1 = E\left[X_j \sum_{\gamma}^{n} (X_i - \mu_i) \right] = \sum_{i=1}^{n} \sigma_{ij} .$$

PROOF OF THE THEOREM.

$$\text{Var}\left(\sum_{i=1}^{n} c_i X_i \right) = \sum_{i=1}^{n} \sum_{j=1}^{n} c_i c_j \sigma_{ij} . \tag{6}$$

Hence, to minimize (6) subject to the condition (1), we get the following equations to be satisfied by c_i' s, $i = 1, \cdots, n$,

$$\sum_{i=1}^{n} c_i \sigma_{ij} = \lambda, \qquad j = 1, \cdots, n, \tag{7}$$

where 2 λ is used as Lagrangian undetermined multiplier.

From (2) and (7) it follows, on summing over the n equations, that $\lambda = 1/n$, so that the desired values of c_i' s, $i = 1, \cdots, n$, should satisfy

$$\sum_{i=1}^{n} c_i \sigma_{ij} = 1/n, \qquad j = 1, \cdots, n. \tag{8}$$

Comparing the equations (2)with (8) and noting that the matrix (σ_{ij}) is nonsingular, it follows that the solution of equation (8) is $c_i = 1/n, i = 1, \cdots, n$.

This proves the theorem.

In the above theorem, when

$$\mu_1 = \mu_2 = \cdots = \mu_n ,$$

Lemma 1 was derived by Lloyd [1]. Also we get in this special case the known result that Var ($\sum_{i=1}^{n} c_i U_i$), where $\sum_{i=1}^{n} c_i = 1$, and $u_1 \leqq u_2 \leqq \cdots \leqq u_n$ are n ordered values from $N(\mu, 1)$, is minimum when $c_i = 1/n, i = 1, \cdots, n$.

ACKNOWLEDGMENT. My thanks are due to Prof. Wassily Hoeffding for indicating the above proof of Lemma 1.

REFERENCES

[1] E. H. LLOYD, "Least squares estimation of location and scale parameters using order statistics," *Biometrika*, Vol. 39 (1952), pp. 88-95.

[2] K. C. SEAL, "On a class of decision procedures for ranking means," *Unpublished Ph.D. Thesis* (1954), University of North Carolina, Chapel Hill.