LEHMANN, E. L.
Math. Annalen 150, 268—275 (1963)

A Class of Selection Procedures Based on Ranks*

For B. L. VAN DER WAERDEN on the occasion of his 60th birthday

By

E. L. LEHMANN in Berkeley/Calif.

Let X_{ij} ($j = 1, \ldots, n$; $i = 1, \ldots, s$) be independent samples from populations with cumulative distribution functions $F(x - \Theta_i)$. For selecting the population with the highest Θ-value, procedures based on the ranks of the observations are compared with the means procedure which selects the population with the largest mean $\overline{X}_i$. The asymptotic relative efficiency of two selection procedures is defined as the ratio of the sample sizes required to achieve the same minimum probability of selecting a "good" population. It is shown that the asymptotic relative efficiency of the procedures based on ranks relative to the means procedure is the same as that of the associated tests in the two — or c-sample problem. If the ratio of the sample sizes is equal to this efficiency, the two procedures being compared are shown to have the same asymptotic performance characteristic.

1. *The means procedure.* Let X_{ij} ($j = 1, \ldots, n$; $i = 1, \ldots, s$) be independent samples from populations Π_i, with distributions

$$P(X_{ij} \leqq x) = F(x - \Theta_i)\,, \tag{1}$$

and consider the problem of selecting the population with the largest Θ-value. When F is normal, the natural procedure is the *means procedure* M which selects Π_i if

$$\overline{X}_i = \max_k \overline{X}_k \tag{2}$$

where $\overline{X}_k = \sum_{j=1}^{n} X_{kj}/n$. Strong optimum properties of this procedure were proved by BAHADUR [1] and BAHADUR and GOODMAN [2].

The feature of the means procedure (2) on which attention has been focused in the literature (see for example BECHOFER [3] and HALL [5]), is the sample size n required to guarantee some desirable property, for example that

$$P\,(\text{selected distribution is good}) \geqq \gamma\,, \tag{3}$$

where the ith population is considered good if Θ_i is sufficiently close to the

---

* This paper was begun with the partial support of the Office of Naval Research (Nonr-222-43) and was completed while the author was Research Professor for the Adolph C. and Mary Sprague Miller Institute for Basic Research in Science, University of California, Berkeley.

largest Θ-value, say if

(4) $$\Theta_i \geqq \Theta_{\max} - \Delta^*$$

where Δ^* is a given constant.

If we do not wish to rely on the assumption of normality, we can find a large-sample solution, which depends only on the variance σ^2 of F. To this end, consider a sequence of situations for increasing n, and define the ith population as good if

(5) $$\Theta_i \geqq \Theta_{\max} - \Delta^{(n)},$$

where $\Delta^{(n)}$ will be defined below. This definition seems rather opportunistic since it appears that our idea of what Θ-values are acceptable is changing with the sample size. This is of course not the case: the sequence is only a mathematical device for approximating the actual situation. If in a concrete case, the definition (4) applies with a given value of Δ^*, then $\Delta^{(n)}$ will be identified with Δ^*.

Suppose now without loss of generality that $\Theta_s = \Theta_{\max}$. Then for all F the left hand side of (3) takes on its maximum value when

(6) $$\Theta_1 = \cdots = \Theta_{s-1} = \Theta_s - \Delta^{(n)},$$

and the sample size is therefore determined by the condition

(7) $$P(\overline{X}_s = \overline{X}_{\max}) = \gamma \text{ when } (\Theta_1, \ldots, \Theta_s) \text{ satisfies (6)}.$$

The large-sample solution of the sample size problem follows from the following lemma.

Lemma 1. *For fixed γ, and with „goodness" of a population defined by* (5), *let n be determined so that* (6) *and* (7) *hold. Then as $n \to \infty$,*

(8) $$\Delta^{(n)} = \frac{\Delta\sigma}{\sqrt{n}} + o\left(\frac{1}{\sqrt{n}}\right).$$

Here σ^2 is the variance of F and Δ is determined by the condition

(9) $$Q\left(\frac{\Delta}{\sqrt{2}}, \ldots, \frac{\Delta}{\sqrt{2}}\right) = \gamma$$

where Q is the cumulative distribution function of a normally distributed vector $(U_1, \ldots, U_{s-1})$ with

(10) $$E(U_i) = 0, \quad \operatorname{Var}(U_i) = 1, \quad \operatorname{Cov}(U_i, U_j) = \frac{1}{2} \quad \text{for all} \quad i, j.$$

Proof. Let $Y_i = (\overline{X}_i - \overline{X}_s)\sqrt{\frac{n}{2\sigma^2}}$ and let $(\Theta_1^{(n)}, \ldots, \Theta_s^{(n)})$ be a sequence of parameter points satisfying (6). Then equation (7) is equivalent to

$$P(Y_i \leqq 0 \quad \text{for all} \quad i) = \gamma$$

and hence by the central limit theorem (and the fact that the convergence is uniform in the arguments of the cumulative distribution function) to

$$\lim_{n\to\infty} P\left(U_i \leqq (\Theta_s^{(n)} - \Theta_i^{(n)})\sqrt{\frac{n}{2\sigma^2}} \quad \text{for all} \quad i\right)$$
$$= \lim_{n\to\infty} P\left(U_i \leqq \Delta^{(n)}\sqrt{\frac{n}{2\sigma^2}} \quad \text{for all} \quad i\right) = \gamma.$$

If Δ is defined by (9), this equation will be satisfied if and only if (8) holds.

18*

Suppose now that we are given a value Δ^* and wish to find the smallest sample size n for which (3) holds with "goodness" defined by (4). It follows from the lemma that a large sample solution is obtained by putting $\Delta\sigma/\sqrt{n} = \Delta^*$ or

$$n = \left(\frac{\Delta\sigma}{\Delta^*}\right)^2 . \tag{11}$$

Since Δ is determined by γ, this defines n as a function of Δ^*, γ and σ.

2. *Procedures based on scores.* Suppose the observations X_{ij} are ranked and the rank of X_{ij} is denoted by R_{ij}. Let us consider the scores procedures which are obtained by replacing X_{ij} in (2) by a score $h(R_{ij})$. More particularly we shall assume that h is defined by

$$h(r) = E_{F_0}(Z^{(r)}) \tag{12}$$

where $Z^{(1)} < \cdots < Z^{(N)}$ is an ordered sample from a given distribution F_0. The resulting *F_0-scores procedure* $S(F_0)$ then selects the ith population if

$$V_i = \max_k V_k \tag{13}$$

where

$$V_i = \sum_j h(R_{ij}) . \tag{14}$$

Standard terminology suggests calling these procedures nonparametric or distribution-free. However, there is no significance level or similar quantity to be computed, which for procedure (2) would depend on the underlying distribution F but would be independent of F for (13). The terminology is therefore inappropriate in the present case, and so is the associated justification of (13). The relative merits of (13) and (2) must instead be decided on the basis of the sample sizes required to achieve (3), (or some similar criterion).

Let us therefore consider the sample size $m = g(n)$ required by the scores procedure (13) if it is to satisfy (3) with the definition of "good" still being given by (5) and (8). Since h is nondecreasing, it is easily seen that also for this procedure the left hand side of (3) takes on its minimum value when $(\Theta_1, \ldots, \Theta_s)$ satisfies (6), and hence that for large samples the sample size m is determined by the condition

$$\lim_{n\to\infty} P(V_s = V_{\max}) = \gamma \quad \text{when} \quad (\Theta_1, \ldots, \Theta_s) \quad \text{satisfies (6)} . \tag{15}$$

The following lemma is the analogue of Lemma 1 for the present case.

Lemma 2. *For fixed γ, let m be determined so that* (15) *holds, and suppose that F and $J = F_0^{-1}$ satisfy the regularity conditions of Theorem 6.2 and Lemma 7.2* of Puri [8]. *Then as $m \to \infty$*

$$\Delta^{(m)} = \frac{\Delta}{\sqrt{m}} \, \frac{A}{\int \frac{d}{dx}\{J[F(x)]\}\, dF(x)} + o\left(\frac{1}{\sqrt{m}}\right) \tag{16}$$

where

$$A^2 = \int J^2(u)\, du - \left(\int J(u)\right)^2 . \tag{17}$$

The proof of this lemma will be given at the end of section 4.

If m and n are determined by (3) and if the same definition of good is applied in both cases, the quantity $\Delta^{(m)}$ defined by (16) must agree with $\Delta^{(n)}$ defined by (8), and hence

$$\lim_{n\to\infty} \frac{n}{m} = \frac{\sigma^2}{A^2}\left(\int \frac{d}{dx}\{J[F(x)]\}\,dF(x)\right)^2. \tag{18}$$

This relative efficiency $e_{s(F_0),M}$ of the F_0-scores procedure $s(F_0)$ to the means procedure M is the same as that found by Chernoff and Savage [4] for the corresponding tests in the two-sample problem and shown by Puri [8] to be valid also for the c-sample problem.

Let us now consider two special cases:

(i) If F_0 is a rectangular distribution, (13) reduces to the rank-sum procedure R, which selects the ith population when $\sum_j R_{kj}$ takes on its maximum value for $j = i$. The efficiency (18), then is known to satisfy $e_{R,M}(F) \geqq .864$ for all F; $e_{R,M}(F) = 3/\pi \sim .955$ when F is normal; and $e_{R,M}(F) > 1$ for many non-normal distributions (Hodges and Lehmann [6]).

(ii) If F_0 is a normal distribution, the efficiency $e_{S(\Phi),M}$ of the resulting normal scores procedure is known to satisfy $e_{S(\Phi),M}(F) \geqq 1$ for all F, and $e_{S(\Phi),M}(F) = 1$ if and only if F is normal.

Thus from the efficiency point of view both the rank procedure and the normal scores procedure (or the asymptotically equivalent procedure based on van der Waerdens X-test [9]) appear to be advantageous compared with the means procedure, unless one can be reasonably sure of the absence of gross errors and other departures of normality. (For an efficiency comparison of the rank sum to the normal scores procedure see Hodges and Lehmann [7].)

Let us next consider the problem of sample size determination with a scores procedure. Suppose we are given a value Δ^* and wish to find the smallest sample size m for which (3) holds when the definition of "good" is given by (4). This is approximately achieved by putting

$$m = \left(\frac{\Delta}{\Delta^*}\right)^2 \frac{A^2}{\left(\int \frac{d}{dx}\{F_0^{-1}[F(x)]\}\,dF(x)\right)^2}.$$

To be specific, let us consider the rank-sum procedure for which this equation reduces to

$$m = \left(\frac{\Delta}{\Delta^*}\right)^2 \frac{1}{12(\int f^2(x)\,dx)^2}$$

where f denotes the probability density of F.

Suppose we expect F to be approximately normal with variance τ^2. For this distribution the sample size m becomes

$$m = \left(\frac{\Delta}{\Delta^*}\right)^2 \cdot \frac{\pi\tau^2}{3}.$$

This sample size determination is considerably more robust against gross errors or other deviations in the tail behaviour of F than that for the means procedure since $\int f^2(x)\,dx$ is much less sensitive to such deviations than is the variance of the X's.

3. *Comparison of performance characteristics.* If $m = g(n)$ is determined so that (18) holds, then for large n the procedures $S(F_0)$ and M have approximately the same minimum probability γ of selecting a good population. This, however, still leaves open the relation of the performance of the two procedures (as characterized by the probability of selecting a good population) for parameter points not satisfying (6).

To discuss this relationship, consider any sequence of parameter points satisfying

$$\Theta_s^{(n)} - \Theta_i^{(n)} = \Delta_i^{(n)} = \frac{\Delta_i \sigma}{\sqrt{n}} + o\left(\frac{1}{\sqrt{n}}\right). \tag{19}$$

We shall further restrict attention to points for which $\Delta_i \neq \Delta$ for all i to avoid ambiguities as to whether or not a population is good. Without loss of generality suppose that

$$\Theta_1^{(n)}, \ldots, \Theta_r^{(n)} < \Theta_s^{(n)} - \Delta^{(n)} < \Theta_{r+1}^{(n)}, \ldots, \Theta_{s-1}^{(n)} \leqq \Theta_s^{(n)} \tag{20}$$

so that for the means procedure M

$$\lim_{n \to \infty} P \text{ (selected population is good)} = \sum_{i=r+1}^{s} P_i(\Delta_1, \ldots, \Delta_{s-1}) \tag{21}$$

where

$$P_i(\Delta_1, \ldots, \Delta_{s-1}) = \lim_{n \to \infty} P(\overline{X}_i = \overline{X}_{\max}) . \tag{22}$$

Since the joint limiting distribution of the variables

$$Y_i = (\overline{X}_i - \overline{X}_s)\sqrt{\frac{n}{2\sigma^2}}$$

is the same as that of an $(s-1)$-dimensional normal vector $(U_1, \ldots, U_{s-1})$ with

$$E(U_i) = -\frac{\Delta_i}{\sqrt{2}}, \quad \operatorname{Var}(U_i) = 1, \operatorname{Cov}(U_i, U_j) = \frac{1}{2} \quad \text{for all} \quad i, j \tag{23}$$

it follows that

$$P_s(\Delta_1, \ldots, \Delta_{s-1}) = P(U_j \leqq 0 \quad \text{for all} \quad j) = Q\left(\frac{\Delta_1}{\sqrt{2}}, \ldots, \frac{\Delta_{s-1}}{\sqrt{2}}\right) \tag{24}$$

and

$$P_i(\Delta_1, \ldots, \Delta_{s-1}) = P(U_i = U_{\max} \text{ and } U_i > 0) \quad \text{for all} \quad i = 1, \ldots, s-1. \tag{25}$$

Consider now the F_0-scores procedure based on samples of size $m = g(n)$ where m again satisfies (18). Then the limiting behaviour of the procedure can be seen from the following theorem, which will be proved in the next section.

Theorem 1. *For* $n = 1, 2, \ldots$ *let* X_{ij} $(j = 1, \ldots, m = g(n);\ i = 1, \ldots, s)$ *be independently distributed according to* $F(x - \Theta_i^{(n)})$. *Suppose that the sequence of parameter points* $\Theta^{(n)} = (\Theta_1^{(n)}, \ldots, \Theta_s^{(n)})$ *satisfies the assumptions of Theorem 6.2 and Lemma 7.2 of* [8]. *Let* $m = g(n)$ *be determined so that* (18) *holds. Then if*

$$T_i^{(n)} = (V_i - V_s)/\sqrt{2mA^2} \tag{26}$$

where V_i *is defined by* (14) *and* (12), *the joint limiting distribution of the random variables* $(T_1^{(n)}, \ldots, T_{s-1}^{(n)})$ *is the distribution of an* $(s-1)$*-dimensional normal vector* $(U_1, \ldots, U_{s-1})$ *satisfying* (23).

It follows from this theorem that the F_0-scores procedure based on samples of size m, with $m = g(n)$ determined in accordance with (18), satisfies (21) where the P_i are given by (24) and (25). For sequences $\Theta^{(n)}$ satisfying (19), this procedure therefore has the same asymptotic performance characteristic as the means procedure based on samples of size n. This statement holds not only for sequences $(\Theta_1^{(n)}, \ldots, \Theta_s^{(n)})$ for which all differences $\Theta_s^{(n)} - \Theta_i^{(n)}$ tend to zero at the rate $1/\sqrt{n}$ as assumed in (19). This follows from the fact that the behaviour for differences tending to zero faster than this rate is obtained by putting $\Delta_i = 0$ in (19), and sequences tending to zero more slowly than $1/\sqrt{n}$ (or tending to a positive limit) by putting $\Delta_i = \infty$.

4. *Proof of Theorem 1 and Lemma 2.* Under the assumptions of Theorem 1, let

$$F_j(x) = F\left((x + \frac{\Delta_j}{\sqrt{m}}\right) \tag{27}$$

and

$$\mu_j = \int J[H(x)]\, dF_j(x) \tag{28}$$

where $H(x) = \sum_{j=1}^{s} F_j(x)/s$. Then Puri [8] has shown that the variables

$$[V_i - m\,\mu_i(\Theta^{(n)})]/\sqrt{m} \tag{29}$$

asymptotically have a joint normal distribution with zero means and covariance matrix

$$\sigma_{ij} = \left(\delta_{ij} - \frac{1}{s}\right) A^2, \tag{30}$$

where the δ_{ij} are the Kronecker deltas and where A^2 is given by (17).

We require the following lemma, the proof of which is immediate, but which is the key for the simple comparability of the limit distributions of the variables $(T_1^{(n)}, \ldots, T_{s-1}^{(n)})$ and $(Y_1, \ldots, Y_{s-1})$.

Lemma 3. *Let the covariance matrix of the random vector* $(W_1, \ldots, W_s)$ *be given by*

$$\sigma_{ij} = \begin{matrix} \tau^2 & \text{if} & i = j \\ \varrho\tau^2 & \text{if} & i \neq j \end{matrix} \tag{31}$$

and let $Z_i = W_i - W_s$ $(i = 1, \ldots, s)$. *Then*

$$\operatorname{Cov}(Z_i, Z_j) = \begin{matrix} 2\tau^2(1-\varrho) & \text{if} & i = j \\ \tau^2(1-\varrho) & \text{if} & i \neq j \end{matrix}$$

and hence the correlation coefficient of Z_i *and* Z_j *is* $1/2$ *regardless of the values of* ϱ *and* τ^2.

It follows from this lemma that two different random vectors having the structure of the Z's assumed in the lemma and having normal limit distributions can, as in the univariate case, be compared by comparing their means and variances.

To prove the theorem, consider the random variables $T_i^{(n)}$ defined by (26). By Lemma 3, the variance of the limiting distribution of $T_i^{(n)}$ is $2\tau^2(1-\varrho)$ where by (30),

$$\tau^2 = (s-1)/2s \quad \text{and} \quad \varrho = -1/(s-1)\,. \tag{32}$$

The variance of the limiting distribution of $T_i^{(n)}$ is therefore seen to be equal to 1.

It only remains to show that the mean of $T_i^{(n)}$ in the limiting distribution is equal to $-\Delta_i/\sqrt{2}$, i.e. that for all $i = 1, \ldots, s-1$,

$$\lim_{n\to\infty} \sqrt{m}\,[\mu_i(\Theta^{(n)}) - \mu_s(\Theta^{(n)})]/\sqrt{2A^2} = -\Delta_i/\sqrt{2}\,. \tag{33}$$

Let $\Delta_s = 0$ and $\bar{\Delta} = (\Delta_1 + \cdots + \Delta_s)/s$. Then we shall show that for all $i = 1, \ldots, s$

$$\lim_{n\to\infty} \sqrt{n}\,[\mu_i(\Theta^{(n)}) - \mu_i(0)] = \sigma(\bar{\Delta} - \Delta_i) \int \frac{d}{dx}\{J[F(x)]\}\, dF(x)\,. \tag{34}$$

Equation (33) is then an immediate consequence of (34) and (18).

To see (34), note that its left hand side can be written as

$$\sqrt{n} \int \left\{ J\left[\frac{1}{s}\sum_{j=1}^{s} F(x + \Theta_i^{(n)} - \Theta_j^{(n)})\right] - J[F(x)]\right\} dF(x)$$

$$= \int A_n(x)\, B_n(x)\, dF(x)\,,$$

where

$$A_n(x) = \frac{J\left[\frac{1}{s}\sum_{j=1}^{s} F\left(x + \frac{(\Delta_j - \Delta_i)\sigma + R_n}{\sqrt{n}}\right)\right] - J[F(x)]}{\frac{1}{s}\sum_{j=1}^{s} F\left(x + \frac{(\Delta_j - \Delta_i)\sigma + R_n}{\sqrt{n}}\right) - F(x)}$$

and

$$B_n(x) = \sqrt{n}\left[\frac{1}{s}\sum_{j=1}^{s} F\left(x + \frac{(\Delta_j - \Delta_i)\sigma R_n}{\sqrt{n}}\right) - F(x)\right].$$

Here R_n is independent of x and tends to zero as $n \to \infty$. Under the assumed regularity conditions,

$$\lim_{n\to\infty} A_n(x) = \frac{d}{du} J(u)|_{u=F(x)}, \quad \lim_{n\to\infty} B_n(x) = \sigma(\bar{\Delta} - \Delta_i)\, F'(x)$$

and, since differentiation under the integral is permitted, this completes the proof of (34) and hence of Theorem 1.

To prove Lemma 2, consider a sequence $\Theta^{(m)}$ of parameter points satisfying

$$\Theta_1^{(m)} = \cdots = \Theta_{s-1}^{(m)} = \Theta_s^{(m)} - \Delta^{(m)}\,.$$

It follows as before from Puri's theorem that the random variables $[(V_i - V_s) - m(\mu_i(\Theta^{(m)}) - \mu_s(\Theta^{(m)}))]/\sqrt{2mA^2}$ have the limiting distribution of an $(s-1)$-dimensional normal vector $(U_1, \ldots, U_{s-1})$ satisfying (10). Condition (15) with n in (6) replaced by m is equivalent to

$$\lim_{m\to\infty} P\left(U_i \leqq (\mu_s(\Theta^m) - \mu_i(\Theta^{(m)}))\sqrt{m/2A^2} \quad \text{for all} \quad i\right) = \gamma\,.$$

Since μ_i is independent of i, this reduces to

$$\lim_{m \to \infty} [\mu_s(\Theta^{(m)}) - \mu_i(\Theta^{(m)})] \sqrt{m/2A^2} = \Delta/\sqrt{2} \quad \text{for all} \quad i = 1, \ldots, s-1 .$$

By the argument used to prove (34), this condition is seen to be equivalent to (16), which completes the proof.

References

[1] Bahadur, Raghu Raj: On a problem in the theory of k populations. Ann. Math. Stat. **21**, 362—375 (1950).

[2] —, and L. A. Goodman: Impartial decision rules and sufficient statistics. Ann. Math. Stat. **23**, 553—562 (1952).

[3] Bechofer, R. E.: A single sample multiple decision procedure for ranking means of normal populations with known variances. Ann. Math. Stat. **25**, 273—289 (1954).

[4] Chernoff, H., and I. R. Savage: Asymptotic normality and efficiency of certain nonparametric test statistics. Ann. Math. Stat. **29**, 972—994 (1958).

[5] Hall, W. J.: The most-economical character of some Bechofer and Sobel decision rules. Ann. Math. Stat. **30**, 964—969 (1959).

[6] Hodges, J. L., jr, and E. L. Lehmann: The efficiency of some nonparametric competitors of the t-test. Ann. Math. Stat. **27**, 324—335 (1956).

[7] — — Comparison of the normal scores and Wilcoxon tests. Proc. 4th Berk. Symposium on Math. Stat. and Prob. **1**, 307—317 (1961).

[8] Puri, M. L.: Asymptotic efficiency of a class of c-sample tests. Submitted to Ann. Math. Stat.

[9] van der Waerden, B. L.: Ein neuer Test für das Problem der zwei Stichproben. Math. Ann. **126**, 93—107 (1953).

(Received October 15, 1962)