MAXIMIZING THE PROBABILITY THAT ADJACENT ORDER STATISTICS OF SAMPLES FROM SEVERAL POPULATIONS FORM OVERLAPPING INTERVALS[1]

RICHARD COHN, FREDERICK MOSTELLER, JOHN W. PRATT AND MAURICE TATSUOKA

Rutgers University, Harvard University, Harvard University and University of Hawaii

1. Summary. Let samples of size n be drawn from each of k univariate continuous cumulative distribution functions on the same real line, and consider the intersection of the k intervals between the rth and $(r + 1)$st order statistics in the several samples. Then, to maximize the probability that that intersection be nonempty the distributions should be identical. Furthermore, for each sample, consider two intervals—that between the rth and $(r + 1)$st and that between the sth and $(s + 1)$st order statistics—then to maximize the probability that both the intersection of the "r" intervals and the intersection of the "s" intervals be nonempty, the distributions again should be identical and the value of the maximum probability is

$$\frac{\binom{n}{r}^k \binom{n-r}{s-r}^k}{\binom{kn}{kr}\binom{k[n-r]}{k[s-r]}}, \qquad r \leq s.$$

Some possible directions for generalization are discussed. The problem arose in connection with a sociological study of interaction behavior in small groups. The results make it possible to provide a test of the hypothesis that several samples of the same size are randomly drawn from possibly different populations, against the alternative that the samples are not independently and randomly drawn from distributions.

For example, suppose we observe the frequency of a particular sort of interaction for each member of five groups of size six. Suppose the five men with the highest frequencies each belong to a different group. Then we can say (ignoring discreteness) that an event has occurred whose probability under random sampling is at most 144/2639 or about 0.055. (The statistic would have but two values, either the five highest belong to different groups, or they do not. Such a test would be especially appropriate if group structure were thought to develop

Received October 20, 1958; revised February 23, 1960.

[1] This work was facilitated by a grant from The Ford Foundation and by the Laboratory of Social Relations of Harvard University; it was supported in part by the Rutgers Research Council, in part by a research grant M-926 from the National Institute of Mental Health of the National Institutes of Health, Public Health Service, and in part by grant G-5554 from the National Science Foundation.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Mathematical Statistics.* www.jstor.org

automatically certain specialized functions in members.) However, the main interest in this paper is in the problem in probability.

2. Introduction. Suppose that we have a sample of n observations from each of k one-dimensional populations with arbitrary continuous distributions. Let a_i denote the largest, b_i the second largest observation in the ith sample, $1 \leqq i \leqq k$. We consider the probability $P(\min a_i \geqq \max b_i)$. In particular we wish to know what relation must hold among the distributions of the k populations for this probability to be maximized. Alternative ways of describing this probability are these: (a) Let the k largest observations in the entire collection of kn observations be chosen. What is the probability that each of these is the largest in its own sample (i.e., is an a_i)? (b) Consider the k intervals formed by the largest and second largest observations in each sample. What is the probability that no two of these intervals fail to overlap?

It seems intuitively reasonable that the maximization is achieved when the k distributions are identical, and so it turns out to be, as we prove later. The property concerned is one which states a similarity among the k samples. It is reasonable to expect the samples to have the highest probability of being similar when the populations from which they come are identical. While it is possible that there is a corresponding theorem for multivariate distributions, dealing with overlaps of statistically equivalent blocks [6], the examples of Section 6 suggest that the generalization is not immediate.

This general probability question grew out of a study of sociological data. Prof. Matilda Riley, of the Department of Sociology of Rutgers University, was interested in ways of formulating mathematically the notion of social structure. One specific problem [5] was to show that the interaction scores of members of several groups were not random collections of scores drawn from single populations (one for each group). The results of the present paper yield one way of studying such a question for groups of the same size.

3. Derivation of the probability expression. Let us denote the k cumulative distribution functions by $y_1(x), y_2(x), \cdots, y_k(x)$. We assume these functions are continuous; the reader may assume they are differentiable, but our demonstrations are valid even if they are not.

Let the least of the largest observations in the k samples be u. This may, of course, be the largest of the n observations from $y_1(x)$, *or* the largest of the n observations from $y_2(x)$, and so forth through $y_k(x)$. That is, $u = \min a_i$ where a_i denotes the largest observation in the sample from $y_i(x)$.

Consider as typical the case in which the sample from $y_1(x)$ contains the least of the largest observations; that is, the largest observation a_1 in the sample from $y_1(x)$ has the value u. Then, the probability that this occurs and that no second-largest observation b_i in any sample exceeds u may be found by computing this probability for fixed u and then integrating. We use differential language for convenience and first find:

Prob {1st sample has $n - 1$ numbers not exceeding u, 1 number in the interval $(u, u + du)$, and no number greater than $u + du$, and that the 2nd, 3rd, $\cdots$, kth samples each have $n - 1$ numbers not exceeding u and 1 number greater than u}.

Since the samples are independently drawn from the k populations, this probability can be written as the product of the following probabilities of the k events described above in the curly brackets:

$$\frac{n!}{(n-1)!1!0!}[y_1(u)]^{n-1}[dy_1(u)][1 - y_1(u)]^0 = ny_1^{n-1}\,dy_1,$$

$$\frac{n!}{(n-1)!1!}[y_2(u)]^{n-1}[1 - y_2(u)] = ny_2^{n-1}(1 - y_2),$$

$$\cdots$$

$$\frac{n!}{(n-1)!1!}[y_k(u)]^{n-1}[1 - y_k(u)] = ny_k^{n-1}(1 - y_k),$$

where $dy_1(u) = y_1(u + du) - y_1(u)$ and, in the simplified expression for each probability, the argument u has been omitted as understood.

Thus, the probability that the largest observation of the first sample be the smallest of the largest, and that none of the second-largest observations exceeds u, is given by

$$(1) \qquad n^k \left[\prod_{j=2}^{k} y_j^{n-1}(1 - y_j)\right] y_1^{n-1}\,dy_1,$$

for *any particular* value u. Hence, the probability that the above event should happen with *some* value of u is the integral of the expression (1) over u from $-\infty$ to $+\infty$; that is,

$$(2) \qquad n^k \int_{-\infty}^{+\infty} \left[\prod_{j=2}^{k} y_j^{n-1}(1 - y_j)\right] y_1^{n-1}\,dy_1,$$

where it is understood that each y_i has the argument u and the limits $-\infty$ and $+\infty$ are those for u.[2] If y_1 is differentiable, the integral $\int_{-\infty}^{+\infty} \cdots dy_1$ is $\int_{-\infty}^{+\infty} \cdots (dy_1/du)du$; if y_1 is not differentiable, the integral is construed in the Stieltjes sense.

Exactly similar considerations hold for the cases in which the 2nd, 3rd, $\cdots$,

[2] The argument of this paragraph may be made rigorous as follows. The sum of (1) over a set of mutually exclusive and exhaustive intervals $(u, u + du]$ approaches (2) as the lengths of the intervals approach 0. Now this sum is the probability that, for some one of the intervals $(u, u + du]$, each sample has exactly one number greater than u and the first sample has no number greater than $u + du$; hence (2) is the limit of this probability, which is the probability that $a_1 = \min a_i > \max b_i$, since the latter event is the limit of the former.

kth samples each in turn contains the smallest of the largest observations, and they yield the probability expressions,

$$(3)\qquad \begin{array}{c} n^k \int_{-\infty}^{\infty} \left[\prod_{j \neq 2}^{k} y_j^{n-1}(1 - y_j) \right] y_2^{n-1}\, dy_2, \\ \cdots \\ n^k \int_{-\infty}^{\infty} \left[\prod_{j=1}^{k-1} y_j^{n-1}(1 - y_j) \right] y_k^{n-1}\, dy_k, \end{array}$$

respectively. The $\prod y_j^{n-1}(1 - y_j)$ in each case lacks the factor $y_i^{n-1}(1 - y_i)$ for the particular value i that coincides with the subscript of the differential, dy_i, being integrated.

Now the event whose probability we want is the union of the k events described in the foregoing, since we do not care which sample happens to contain the smallest of the largest observations. And since these k events are mutually exclusive, the desired probability is the sum of the k separate probabilities. We denote this by P_k ; thus

$$(4)\qquad P_k = n^k \int_{-\infty}^{\infty} \left[\prod_{m=1}^{k} y_m^{n-1} \right] \sum_{i=1}^{k} \left[\prod_{j \neq i}^{k} (1 - y_j) \right] dy_i,$$

where again the integration is over u.

4. Maximizing the probability. We now address ourselves to our main question: What relationship should hold among the k distributions y_i in order that the probability P_k shall be as large as possible? If, as suggested in the introduction, the maximum P_k is attained when all k distributions are identical: $y_1 = y_2 = \cdots = y_k = y$, then P_k reduces to

$$(5)\qquad kn^k \int (1 - y)^{k-1} y^{k(n-1)}\, dy.$$

Here the integration may be actually performed over y (instead of the underlying variable u). The limits of y are 0 and 1 and the integral is a complete beta-function. The value of expression (5) is $n^k \Big/ \binom{kn}{k}$ (This result can alternatively be obtained by a combinatorial argument.)

We now prove that P_k is maximized by the relation $y_1 = y_2 = \cdots = y_k$. We define

$$(6)\qquad H(u) = \prod_{i=1}^{k} [1 - y_i(u)].$$

Motivation for the introduction of H comes from noticing that dH (see below) is, except for sign, the last part of the expression under the integral sign in equation (4). $H(u)$ may be interpreted as the probability that k observations, one from each distribution, exceed u, or $1 - H(u)$ may be interpreted as the cumulative distribution function of the minimum of k such observations. By two applications of the well-known inequality between the arithmetic and the geometric mean we find[3]

$$\prod_{i=1}^{k} y_i = \prod_{i=1}^{k} [1 - (1 - y_i)] \leqq \left[1 - \frac{1}{k}\sum_{i=1}^{k} (1 - y_i)\right]^k \tag{7}$$
$$\leqq \left[1 - \prod_{i=1}^{k} (1 - y_i)^{1/k}\right]^k = (1 - H^{1/k})^k.$$

From definition (6), we see that the total differential of $H(u)$ is

$$dH = \sum_{i=1}^{k}\left[\frac{\partial}{\partial y_i}\prod_{j=1}^{k} (1 - y_j)\right] dy_i = -\sum_{i=1}^{k}\left[\prod_{j \neq i}^{k} (1 - y_j)\right] dy_i .$$

Substituting in (4) we find

$$P_k \leqq n^k \int_{-\infty}^{\infty} (1 - H^{1/k})^{k(n-1)}(-1)\, dH, \tag{8}$$

where again the integration is over u. Putting $y(u) = 1 - H^{1/k}$, so that $H = (1 - y)^k$, the right-hand side of (8) reduces to (5). This completes the proof.

The proof just given was suggested by the referee. It replaces a longer proof that routinely applied the method of Lagrange multipliers. Our original proofs of Theorems 1 and 2 below also used Lagrange multipliers, but, at the suggestion of the referee we developed proofs that are more in the spirit of the above method.

We now prove that P_k is maximized only where all y_i are equal. Note that, for any given set of y_k, (6) defines u as a function of H. The y_i are single-valued, continuous, monotonically decreasing functions of H except possibly at $H = 0$, even though u is not. For $H = 0$, let y_i equal its limit as H decreases to 0. Since there is no contribution to (4) when $H = 0$, for then at least one y_i is 1, it may be rewritten as

$$P_k = n^k \int_0^1 \prod_{j=i}^{k} y_j^{n-1}\, dH.$$

The right-hand side of (8) becomes

$$n^k \int_0^1 (1 - H^{1/k})^{k(n-1)}\, dH.$$

Both integrations are now over H.

[3] At the request of the editor we note that the inequality between the first and fourth member could have been rewritten as an inequality expressing the convexity of the function $[\Pi\, y_i]^{1/k}$:

$$[\prod y_i]^{1/k} + [\prod (1 - y_i)]^{1/k} \leqq 1.$$

That inequality is a special case of what Hardy, Littlewood, and Pólya [2] call the Hölder inequality, and they attribute the quite special form $[\prod \alpha_i]^{1/k} + [\prod \beta_i]^{1/k} \leqq [\prod (\alpha_i + \beta_i)]^{1/k}$ to Minkowski in 1896. We refer to the current reprint [4] of that work.

The difference of the two integrands is continuous and, by (7), nonnegative. Since [1] the integral of a continuous nonnegative function vanishes only if the function is identically 0, the two integrals can be equal only if the integrands are. But the inequalities in (7) are strict unless all y_i are equal. This shows that if P_k is maximized, the y_i must be all equal except possibly at $H = 0$. But then they must all approach 1 as H decreases to 0, and hence they must be equal to 1 when $H = 0$.

5. Two generalizations. The result obtained in the preceding section is extended in two ways in this section. The first extension concerns the maximization of the probability of overlap of the k intervals formed by the rth and $(r + 1)$st order statistics of the k samples. The result, as before, is that all distributions should be identical, and we have removed the restriction $r = 1$, imposed in Section 4. We sketch the proof.

Denoting the rth observation (in descending order of magnitude) in the ith sample by x_{ri}, the probability to be maximized turns out to be given by

$$\text{Prob}\,\{\min x_{ri} > \max_i x_{r+1,i}\} \\ = r\binom{n}{r}^k \sum_{i=1}^{k} \int_{-\infty}^{\infty} \left[\prod_{j\neq i}^{k} y_j^{n-r}(1 - y_j)^r\right] y_i^{n-r}(1 - y_i)^{r-1}\, dy_i. \tag{9}$$

Defining $H(u)$ and $y(u)$ as in Section 4, we find that this probability is bounded by

$$r\binom{n}{r}^k \int_{-\infty}^{\infty} (1 - H^{1/k})^{k(n-r)} H^{r-1} \sum_{i=1}^{k} \frac{H}{1 - y_i}\, dy_i \\ = r\binom{n}{r}^k \int_0^1 (1 - H^{1/k})^{k(n-r)} H^{r-1}\, dH. \tag{10}$$

The last expression in (10) is the probability given in (9) when $y_1 = y_2 = \cdots = y_k = y$. Hence, this is again the maximizing condition. We shall not take space for a uniqueness proof similar to that at the close of Section 4.

In considering the intervals $[x_{r+1}, x_r]$, $r = 1, 2, \cdots, n - 1$, included between the adjacent order statistics x_{r+1} and x_r, it is convenient to extend the set of intervals to include $[x_{n+1}, x_n]$ and $[x_1, x_0]$, where x_{n+1} and x_0 are interpreted as $-\infty$ and $+\infty$ respectively. The whole set of intervals then forms the basic statistically equivalent blocks for a univariate distribution. If we substitute $r = 0$ or $r = n$ in expression (11) below for the minimum probability, we get unity in each case, as we should.

We summarize the foregoing in

THEOREM 1. *If samples of size n are drawn from each of k continuous distribution functions $y_i(x)$, $i = 1, 2, \cdots, k$, the probability that*

$$\min_i x_{ri} > \max_i x_{r+1,i}$$

for a given r $(= 1, 2, \cdots, n - 1)$ *is maximized if and only if* $y_1 = y_2 = \cdots = y_k$, *and the maximum value is*

$$\binom{n}{r}^k \Big/ \binom{kn}{kr}. \tag{11}$$

As our second extension, we consider a more general situation in which we pay attention to two corresponding intervals from each sample, say the interval formed by the rth and $(r + 1)$st order statistics and that formed by the sth and $(s + 1)$st order statistics. What is the condition for maximizing the probability that the "r" intervals overlap and also the "s" intervals overlap? The answer turns out to be the same as before: the k distribution functions should be identical. We sketch the demonstration.

It can be shown that the probability in question (letting $s > r$) is given by

$$\begin{aligned} &\text{Prob}\,\{\min_i x_{ri} > \max_i x_{r+1,i} \text{ and } \min_i x_{si} > \max_i x_{s+1,i}\} \\ &= A \sum_{i=1}^{k} \sum_{j=1}^{k} \int_{-\infty}^{\infty} dy_i(u) \int_{-\infty}^{u} \frac{\prod_{l=1}^{k} y_l(v)^{n-s}[y_l(u) - y_l(v)]^{s-r}[1 - y_l(u)]^r}{[y_j(u) - y_j(v)][1 - y_i(u)]}\, dy_j(v), \end{aligned} \tag{12}$$

where $u = \min_i x_{ri}$ and $v = \min_i x_{si}$ and

$$A = r(s - r)\left[\frac{n!}{r!(s - r)!(n - s)!}\right]^k.$$

We plan to integrate first with respect to v. To facilitate this we pull out from the integrand factors involving u alone, and we factor out enough powers of $y_l(u)$ so that every $y_l(v)$ and $dy_j(v)$ may have a matching denominator of $y_l(u)$. We then substitute $z_l(v) = y_l(v)/y_l(u)$, noting that $z_l(v)$ can itself be regarded as a cumulative distribution. After these substitutions we have as the integral with respect to v

$$\int_{-\infty}^{u} [\prod z_l(v)]^{n-s} [\prod (1 - z_l(v))]^{s-r} \sum dz_j(v)/(1 - z_j(v)). \tag{13}$$

By the methods used earlier, the maximum of this expression is readily found to be $1 \Big/ \left[(s - r)\binom{k(n - r)}{k(s - r)}\right]$. Then aside from constants we are left with the expression

$$\int_{-\infty}^{\infty} \prod y_l(u)^{n-r} \prod (1 - y_l(u))^r \sum dy_i(u)/(1 - y_i(u)), \tag{14}$$

whose maximum is readily found from equations (10) and (11) to be

$$1 \Big/ \left[r\binom{kn}{kr}\right].$$

The constant A of equation (12) and the two numerical results just obtained, when multiplied together, give the bound shown below in Theorem 2. It is easy to verify that this upper bound is achieved by the right-hand side of equation (12) when all the $y_i(u)$ are identical. Again we omit a uniqueness argument like that at the close of Section 4. We state the conclusion as

THEOREM 2. *If samples of size n are drawn from each of k continuous distribution functions* $y_i(x)$, $i = 1, 2, \cdots, k$, *the probability that*

$$\min_i x_{ri} > \max_i x_{r+1,i}$$

and

$$\min_i x_{si} > \max_i x_{s+1,i}$$

for a given $r(= 1, 2, \cdots, n-1)$ *and a given* $s(=r, r+1, \cdots, n-1)$, *is maximized if and only if* $y_1 = y_2 = \cdots = y_k$, *and the maximum value is*

$$\frac{\binom{n}{r}^k \binom{n-r}{s-r}^k}{\binom{kn}{kr}\binom{k[n-r]}{k[s-r]}}. \tag{15}$$

A degenerate case occurs when $s = r$, because then only one interval is denoted, and then expression (15) should reduce to expression (11), as it does.

Theorem 2 extends to three or more corresponding intervals, and the probability in the case of identical distributions is the obvious extension of expressions (11) and (15). The following alternative proof generalizes directly. The probability in Theorem 2 is the probability P that $\min_i x_{ri} > \max_i x_{r+1,i}$ times the conditional probability Q that $\min_i x_{si} > \max_i x_{s+1,i}$ given $\min_i x_{ri} > \max_i x_{r+1,i}$. The probability P is maximized, by Theorem 1, if and only if $y_1 = y_2 = \cdots = y_k$. Theorem 2 follows if we show this condition also maximizes Q. Given that $u = \min_i x_{ri} > \max_i x_{r+1,i}$, there are exactly $n - r$ observations below u in each sample. Thus, conditionally, $x_{r+1,i}, \cdots, x_{n,i}$ are the order statistics of a sample of $n - r$ from the conditional distribution $y_i(x)/y_i(u)$, $x < u$. By Theorem 1, the conditional probability Q is maximized if $y_1 = y_2 = \cdots = y_k$, which is all that remained to be shown.

6. Examples in two dimensions. The generalization of our results to higher dimensions is not immediate. Examples show that, even though the samples are drawn from the same distribution and statistically equivalent blocks are constructed for each sample by the same program, the probability of overlap depends on both the common distribution and the common program.

Even in one dimension, the probability may depend on the program: if the first block in each of k samples is the interval between the rth and $(r + 1)$st order statistics, the probability that the intersection of the k first blocks is nonempty depends on r, by equation (11) of Theorem 1.

Here the dependence of the probability on the program can be removed by renumbering the blocks correctly. In the following example, this dependence is perhaps even clearer, since it cannot be removed by renumbering blocks.

EXAMPLE 1. From any continuous two-dimensional cumulative, take two samples of size two. *Program* 1. For each sample, construct statistically equivalent blocks by using as boundary functions the vertical lines through the two sample points, thus obtaining three regions in the plane: the right, the left, and the middle. The right-hand regions of the two samples always overlap, as do the left-hand regions, but the middle regions overlap with probability 2/3 because the construction program essentially reduces the problem to one dimension and equation (11) applies with $n = 2$, $k = 2$, $r = 1$.

Program 2. In each sample, use the vertical line through the right-hand sample point for the first boundary function, and, for the second, the horizontal half-line running from $-\infty$ through the left-hand sample point to the first vertical line. This program produces three regions for each sample: the right-hand region, the upper-left, and the lower-left. Each region always overlaps its mate in the other sample.

Thus we have shown that, even when samples are drawn from the same distribution, different methods of constructing statistically equivalent blocks can produce different probabilities of overlap, here 2/3 for one region and 1 for two regions of Program 1 and 1 for all regions of Program 2.

In this example, for neither program did the probability of overlap depend on the distribution sampled. The following example shows this too is possible.

EXAMPLE 2. Again draw two samples of size 2.

Program. For each sample the first boundary is a circle about the origin through the sample point farther from the origin; the second boundary is the chord of this circle passing horizontally through the remaining sample point. Now there are three regions, the one exterior to the circle, and the two interior to the circle, say, the upper-interior and the lower-interior.

Let the samples be drawn from a uniform distribution over a circle of radius 1 with center at a point $(c, 0)$ on the horizontal axis. If c is large, then the upper region is practically the upper half of a large circle about the origin, and such regions are sure to overlap from sample to sample. (Actually, they are sure to overlap near $(0, 1)$ if $|c| > 2$.) On the other hand, if c is 0, then both points of one sample may fall within one half unit of the origin while both points in the other sample fall at least one half unit above the horizontal axis. Then the upper-interior regions do not overlap. Thus without further calculation we can say that the probability of non-overlap is strictly positive for $c = 0$ and zero for large c.

All told then we have shown that, for the two-dimensional problem, even if two samples are drawn from the same distribution and statistically equivalent blocks are constructed in each sample by the same program, the probability of overlap may vary from program to program for a fixed distribution, and from distribution to distribution for a fixed program. It is suggestive that in Example 1 the smaller probability of overlap is associated with the "one-dimensional" Program 1.

One generalization of Theorems 1 and 2 to several dimensions is readily presented, but it succeeds essentially by reducing several dimensions to one. In fact, these theorems apply immediately to statistically equivalent blocks defined by the order statistics of a real-valued function f of the observations. (That is, the first block is the region where $f(x)$ exceeds $f(X_j)$ for every member X_j of the sample, the second block is the region where $f(x)$ exceeds $f(X_j)$ for every X_j but one, etc.) The sample may be drawn from any probability distribution on any set whatever, as long as f is measurable and $f(X_j)$ has a continuous c.d.f. (The distribution must be non-atomic or no $f(X_j)$ has a continuous c.d.f.).

7. Remarks. In connection with another problem, Lehmann ([3], p. 172, Lemma 4.1) proves what amounts to the special case of Theorem 1 with $n = 2$, $k = 2$, $r = 1$.

Several possible generalizations besides those already considered suggest themselves. For example, Theorem 1 concerns a set of k corresponding intervals, one interval for each of k samples, and states that the probability that all k intervals overlap is maximized when the distributions sampled are identical. Is the same true of the probability that at least m of these k intervals overlap? Theorem 2 states that the probability that overlap occurs in both of two sets of corresponding intervals is also maximized when the distributions are identical. Is the same true of the probability that overlap occurs in at least one of the two sets of intervals?

The probability that non-corresponding intervals overlap is not necessarily maximized when the distributions are identical. For example, for one sample of one, take the interval from the observation to $+\infty$; for another sample of one, take the interval from the observation to $-\infty$. These intervals overlap if and only if the second observation exceeds the first. This has probability 1/2 if both observations have the same distribution, but may be larger (even 1) if the observations have different distributions.

Some work is being done on problems of samples of unequal sizes.

Finally, of course, all such questions can be raised about statistically equivalent blocks.

We wish to express our appreciation to William Kruskal for numerous suggestions.

REFERENCES

[1] G. H. Hardy, *A Course of Pure Mathematics*, 10th Edition, Cambridge University Press, Cambridge, 1955, p. 340.

[2] G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities*, Cambridge University Press, Cambridge, 1934, p. 21.

[3] E. L. Lehmann, "Consistency and unbiasedness of certain non-parametric tests," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 172–173.

[4] Herman Minkowski, *Geometrie der Zahlen*, Chelsea Publishing Co., New York, 1953, p. 117, footnote.

[5] M. W. Riley, J. W. Riley, and Jackson Toby, *Sociological Studies in Scale Analysis*, Rutgers University Press, 1954, pp. 25–26.

[6] J. W. Tukey, "Non-parametric estimation II. Statistically equivalent blocks and tolerance regions—the continuous case," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 529–539.