SEQUENTIALLY DETERMINED STATISTICALLY EQUIVALENT BLOCKS

By D. A. S. Fraser

University of Toronto

1. Summary. In 1943 Wald [2] gave a method for constructing tolerance regions in the multivariate case. Tukey generalized Wald's procedure in [4] and the results were interpreted for discontinuous distributions in [5] and [6].

This paper presents a further generalization of the method so that statistically equivalent blocks can be determined sequentially; the particular function used to cut off a block may depend on the shape or structure of previously selected blocks. The results are also interpreted for the case of discontinuous distributions.

Possible advantages of applying the method are discussed.

I. Continuous Case

2. Introduction. The general consideration of statistically equivalent blocks has its origin in Wilks' method [1] of forming tolerance regions by using order statistics. For any interval formed from the order statistics, the proportion of the population "covered", referred to as the "coverage", was considered as the value of a random variable. Wilks showed that the distribution of this "coverage" was independent of the particular continuous population sampled; in fact, it has a Beta distribution depending only on the sample size and the particular order statistics chosen to form the interval.

The method was extended to multivariate populations by Wald and Tukey [2], [4], the latter being responsible for the term "statistically equivalent block"—the multivariate analogue of the interval between two adjacent order statistics. The coverages of these blocks, $n + 1$ of them for samples of n, have a very elementary distribution closely related to that of the n order statistics of a sample from a uniform distribution [0, 1], and the coverage of any sum of blocks has a marginal Beta distribution.

The method used in previous papers to form blocks is, essentially, to have a fixed sequence of functions which are used successively to cut off blocks from the space of the random variable being sampled. In this paper the fixed sequence is replaced by one having the choice of function at any point in the sequence depend on the observed values at the cuts of functions already used. More generally the choice of function can be made randomly from a class of functions where both the probabilities and the class can depend on the functions already used and on their observed values. All the previous results still hold but the proof for the discontinuous case requires special treatment. A precise definition of the blocks is given in Section 4 and the general theorem in Section 5.

Advantages of this generalization for the practitioner can be illustrated by the following examples. Consider a sample of 25 from a continuous bivariate

distribution with the values plotted in Figure 1. Suppose a tolerance region is to be formed by deleting 12 blocks and a further requirement is made that the remaining region should be roughtly of a given shape, say square or octagonal. Corresponding to the example in [4] we shall consider the latter.

The functions to be used to form the region will be the following:

$$y, x, -y, -x, x + y, x - y, -x - y, -x + y.$$

Using the function y a block is formed by the method of [4], that is, the sample point yielding the largest value of y is selected and the first region consists of all points in the two-dimensional space having a larger value of y. Similarly form the second block using the function x: the method is the same as for the first block except that we consider only the $n - 1$ points remaining after deleting

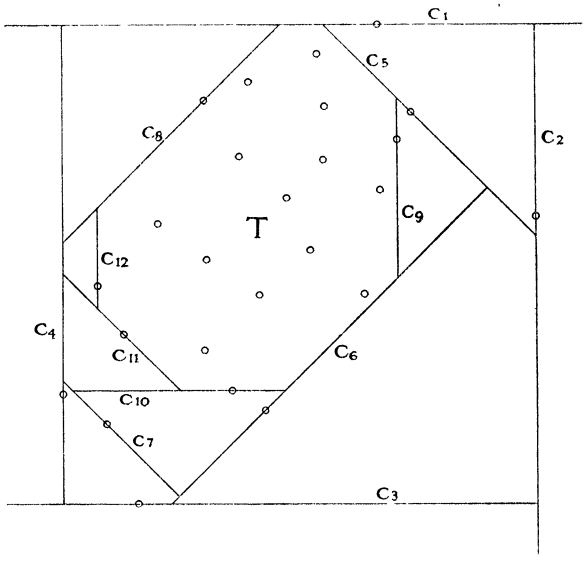

FIG. 1

the one determining the first block, and only that part of the plane after removing the first block. Form successively in this way blocks corresponding to the eight functions.

At this point we deviate from the procedure of [4]. For further functions we select from the given eight functional forms according to the values of the first eight functions at their respective cuts. To obtain a roughly octagonal region we shall make a ninth cut parallel to the shortest of the eight sides of the residual region. However, some of these sides may have vanished completely, in which case we take cuts parallel to the missing sides, commencing with the first when ordered according to the number of the function which produced the cut. This is carried out in the example in Figure 1 until twelve blocks altogether have been removed. The region T remaining after removal of the twelve blocks will be used

as the tolerance region whose minimum coverage with a given confidence level can be calculated by the theorem in Section 5.

Consider a second example in which we use the sample of 25 plotted again in Figure 2 but desire a circular tolerance region. The functions used are the following:

$$y,\ x,\ -y,\ -x,\ (x-\alpha)^2+(y-\beta)^2.$$

As before we remove four blocks using the first four functions, thus reducing the residual region to a rectangle. All further functions used will be identical to the fifth above where α, β are chosen to be the coordinates of the midpoint of the rectangle. Remove eight more blocks and use the residual T as a tolerance region with probabilities prescribed by the theorem in Section 5. Notice that tolerance

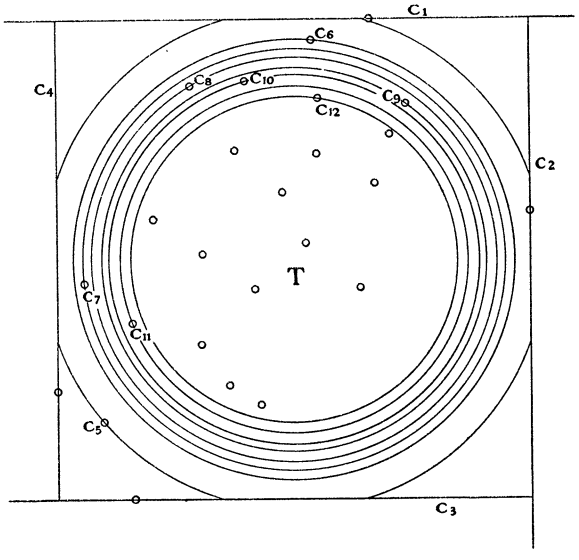

FIG. 2

regions so formed will be either circular or circular with two or four flat sides.

These simple procedures and many possible variations should permit the practitioner to impose quite general but approximate requirements on the shape of the final tolerance region.

3. Notation. Consider a probability distribution over a space S which could be Euclidean of one or higher dimension, or more general. By this we mean there exists a nonnegative additive set function over the space with measure one for the whole space. Denote by w an arbitrary point in the space and let W, called the chance quantity, symbolize the existence of the probability measure defined above. The symbolic operations on a chance quantity are the obvious operations with the probability measure. For example, corresponding to a real-valued Borel function $\varphi(w)$ over the space S, we can use the symbol $\varphi(W)$ for

a chance quantity whose probability measure is defined by $P_{\varphi(W)}(S) = P_W(\varphi^{-1}(S))$, where S is a Borel Set in R^1.

The expression "coverage of a set" is to be interpreted as the probability measure of the set. If the set is a chance quantity, then its coverage is a real-valued chance quantity or random variable.

4. Definition of the blocks. Consider n points in the space $S = \{w\}$ and a family of functions $\varphi_1(w)$, $\varphi_2(w \mid \varphi_1)$, $\cdots$, $\varphi_m(w \mid \varphi_1, \cdots, \varphi_{m-1})$, each of which yields a random variable having a continuous distribution for all values of $\varphi_1, \varphi_2, \cdots, \varphi_{i-1}$ except perhaps for a set having P-measure zero.

In the theorem that follows in the next section these n points will be considered as a sample of n for the chance quantity W.

DEFINITION 4.1. *The set* $w_1, w_2, \cdots, w_n$ *and the functions* $\varphi_1, \varphi_2, \cdots, \varphi_m$ $(m \leq n)$ *define blocks as follows*:

$$S_1 = \{w \mid \varphi_1(w) > \alpha_1\},$$

where $\alpha_1 = \max \varphi_1(w_i) = \varphi_1(w_{i(1)})$, *which defines an* $i(1)$;

$$S_2 = \{w \mid \varphi_1(w) < \alpha_1, \varphi_2(w \mid \alpha_1) > \alpha_2\},$$

where $\alpha_2 = \max_{i \neq i(1)} \varphi_2(w_i \mid \alpha_1) = \varphi_2(w_{i(2)} \mid \alpha_1)$ *and* $i(2) \neq i(1)$, *which defines an* $i(2)$; *in general for* $1 < k \leq m$,

$$S_k = \{w \mid \varphi_1(w) < \alpha_1, \cdots, \varphi_{k-1}(w \mid \alpha_1, \cdots, \alpha_{k-2}) < \alpha_{k-1}, \varphi_k(w \mid \alpha_1, \cdots, \alpha_{k-1}) > \alpha_k\},$$

where $\alpha_k = \max_i \alpha_k(w_i \mid \alpha_1, \cdots, \alpha_{k-1}) = \varphi_k(w_{i(k)} \mid \alpha_1, \cdots, \alpha_{k-1})$, *the maximum being taken over all* i *except* $i(1), i(2), \cdots, i(k-1)$ *and* $i(k)$ *being chosen from the set over which the maximum is taken.*

If $m \leq n$, then

$$S_{m|n+1} = \{w \mid \varphi_1(w) < \alpha_1, \cdots, \varphi_m(w \mid \alpha_1, \cdots, \alpha_{m-1}) < \alpha_m\}.$$

The functions have thus defined $n + 1$ blocks if there are n functions, and if fewer functions, then m blocks and an associated region $S_{m|n+1}$. The definition of the blocks is unique unless $\varphi_i(w_j) = \varphi_i(w_k)$ for some i, j, k.

5. General results. Continuous case.

THEOREM 5.1. *If* $\varphi_i(W \mid \varphi_1, \cdots, \varphi_{i-1})$ *has a continuous distribution for all values of* $\varphi_1, \cdots, \varphi_{i-1}$ *(except perhaps for a set of P-measure zero) and for all* i, *and if for a sample of* n, $(W_1, \cdots, W_n)$, *from the distribution of* W *we define blocks* $S_1, \cdots, S_{m|n+1}$ *according to Definition* 4.1, *then*

(i) *the blocks are disjoint chance sets uniquely defined with probability one, and*

(ii) *the distribution of the coverages*

$$c_i = P(S_i),$$

$$c_{m|n+1} = P(S_{m|n+1})$$

is the same as that of $t_1, t_2, \cdots, t_m$ *and* $\sum_{i=m+1}^{n+1} t_i$, *where* t_i *are uniformly distributed on the barycentric simplex with* $n + 1$ *vertices.*[1]

(i) and (ii) could be replaced by the statement

(iii) $S_1, \cdots, S_{m|n+1}$ *are a partial family of statistically equivalent blocks of type* $n + 1$ *and an associated* $m \mid n + 1$ *tolerance region.*

PROOF. The proof using Wald's principle and induction on m follows closely that given in Section 8 of [4].

II. DISCONTINUOUS CASE

6. Introduction. Scheffé and Tukey [3] considered tolerance regions for discontinuous one dimensional distributions—previous results extended with inequalities replacing the equalities.

The multivariate discontinuous case was considered by Tukey [5]. As well as blocks, cuts must now be considered and this complicates the formation of tolerance regions. Some remarks on the main theorem in [5] are contained in [6].

The results of [5] and [6] carry over to the case where the functions used to form the blocks are decided upon "sequentially." The proof, although similar, requires special treatment and some new devices.

It is perhaps worth remarking that although the functions in [5] reduce all cuts to points, this is not necessary. A cut could be a line with perhaps two or more points on it. Select one by a chance procedure (each with the same probability) to represent the cut. The remaining points are then available to fix the cuts for other blocks.

7. Definition of the cuts and blocks. The formation of the m-system of functions needs to be altered slightly to take care of the new procedure admitting a choice of function at any stage.

As in [5] we order finite sequences $(a_1, \cdots, a_m)$, $(b_1, \cdots, b_m)$ by means of the following rule. $(a_1, \cdots, a_m) > (b_1, \cdots, b_m)$ if any of the following hold:

$$a_1 > b_1,$$

$$a_1 = b_1 \quad \text{and} \quad a_2 > b_2,$$

$$a_i = b_i \ (i < m) \quad \text{and} \quad a_m > b_m.$$

We define $<$ similarly, and $=$ means identity.

DEFINITION 7.1. *An* m*-system of functions* $\Phi_1, \cdots, \Phi_m$ *is defined as follows:*

$$\Phi_k(w) = \{\varphi_{k,1}(w \mid \Phi_1, \cdots, \Phi_{k-1}), \cdots, \varphi_{k,p(k)}(w \mid \Phi_1, \cdots, \Phi_{k-1}),$$

where $\varphi_{k,i}(w \mid \Phi_1, \cdots, \Phi_{k-1})$ *is a real-valued Borel function of* w *in the space* $\mathcal{S}$ *and is also dependent on* $\Phi_1, \cdots, \Phi_{k-1}$, *where these are points in the Euclidean spaces* $R^{p(1)}, R^{p(2)}, \cdots, R^{p(k-1)}$ *and where* $p(k)$ *is a positive integer depending on* k.

[1] See Tukey [4].

We can order values of a function Φ_k by the lexicographical method, described above, for ordering sequences.

DEFINITION 7.2. *Given an m-system of functions and n points* $w_1, \cdots, w_n$ *in* $\mathfrak{S}$ $(m \leq n)$, *the corresponding blocks and cuts are defined by the following procedure: Select* $i(1)$ *to maximize* $\Phi_1(w_i)$. *If more than one value of* i *maximizes* Φ_1, *choose one at random (each taken to be equally likely). Let* α_1 *be this maximum value of* $\Phi_1(w_i)$.

$$S_1 = \{w \mid \Phi_1(w) > \Phi_1(w_{i(1)})\},$$

$$T_1 = \{w \mid \Phi_1(w) = \Phi_1(w_{i(1)})\}.$$

Next, $i(2)$ *is selected* $\neq i(1)$ *to maximize* $\Phi_2(w_i \mid \alpha_1)$, *using the chance procedure as for* $i(1)$ *in case of ties. Let the maximum value be* α_2.

$$S_2 = \{w \mid \Phi_1(w) < \alpha_1, \Phi_2(w \mid \alpha_1) > \Phi_2(w_{i(2)}) \mid \alpha_1)\},$$

$$T_2 = \{w \mid \Phi_1(w) < \alpha_1, \Phi_2(w \mid \alpha_1) = \Phi_2(w_{i(2)} \mid \alpha_1)\},$$

$$S_{m|n+1} = \{w \mid \Phi_k(w \mid \alpha_1, \cdots, \alpha_{k-1}) < \alpha_k ; k = 1, \cdots, m\}.$$

Also define $\bar{S}_1, \bar{S}_2, \cdots, \bar{S}_{m|n+1}$ *by the expressions above for* $S_1, S_2, \cdots, S_{m|n+1}$, *where* $<$ *is replaced by* $\leq$ *and* $>$ *is replaced by* $\geq$.

We denote by λ a subset of the indices $1, 2, \cdots, m$, $m \mid n + 1$.

DEFINITION 7.3. *The block group* B_λ *consists of the union of all* S_i *with* i *in* λ *and all* T_j *not continaed in* $\bar{S}_i$ *with* i *not in* λ. *The closed block group* $\bar{B}_\lambda$ *consists of the union of all* S_i *with* i *in* λ *and all* T_j *contained in any* $\bar{S}_i$ *with* i *in* λ.

The above definition covers all cases where the functions are sufficient to reduce all cuts to points. However, if such is not the case we need the more general definition:

DEFINITION 7.4. *The closed block group* $\bar{B}_\lambda$ *consists of the union of all* $\bar{S}_i$ *with* i *in* λ. *The block group* B_λ *consists of the complement of* $\bar{B}_{C(\lambda)}$ *where* $C(\lambda)$ *is the complement of* λ *with respect to the indices* $1, 2, \cdots, m$, $m \mid n + 1$.

The definition of block groups is unique for a given set of points if all cuts are points; otherwise the chance procedure determines the block groups for a given set of points.

DEFINITION 7.5. *As in* [5] *we let* $c(\lambda)$, $\bar{c}(\lambda)$ *stand for the coverages of the block groups* B_λ, $\bar{B}_\lambda$.

8. General results. Discontinuous case.

THEOREM 8.1. *Let* $\Phi_1, \Phi_2, \cdots, \Phi_m$ *be any m-system of functions (Definition* 7.1), $W_1, \cdots, W_n$ *be a sample of n from an arbitrary distribution designated by* $W(m \leq n)$, *and let the blocks, cuts, block groups, and coverages be formed according to Definitions* 7.2, 7.3, *and* 7.4. *Then if* $\gamma_1, \gamma_2, \cdots, \gamma_p$ *are any set of disjoint* λ*'s,*

$$Pr\{c(\gamma_1) < x_1, \cdots ; \bar{c}(\gamma_k) > x_k, \cdots, \bar{c}(\gamma_p) > x_p\}$$

$$\geq Pr\{t(\gamma_1) < x_1, \ldots ; t(\gamma_k) > x_k, \ldots, t(\gamma_p) > x_p\},$$

where $t_\lambda = \sum_i \varepsilon_\lambda t_i$, $t_{m|n+1} = \sum_{m+1}^{n+1} t_i$, *and* $t_1, t_2, \cdots, t_{n+1}$ *have a uniform distribution on the barycentric simplex.*[2] *In particular we have*

$$Pr\{c(\lambda) < x\} \geq I_x(s, n + 1 - s) \geq Pr\{\bar{c}(\lambda) < x\},$$

where s *is the number of indices in* λ $(m = n)$ *and* $I_x(s, n + 1 - s)$ *is the incomplete Beta function.*

9. The functions ψ. In a manner similar to that of Section 6 in [5], we replace the m-system of functions by real functions ψ.

LEMMA 9.1. *Given an* m*-system of functions* $\Phi_1, \cdots, \Phi_m$, *there exist real functions* $\psi_1(w)$, $\psi_2(w \mid \psi_1), \cdots, \psi_m(w \mid \psi_1, \cdots, \psi_{m-1})$ *such that if* $W_1, \cdots, W_n$ *form a sample of* n *from* W, *then*

(i) $\psi_i(w \mid \psi_1, \cdots, \psi_{i-1})$ *is defined except for values of* $\psi_1, \cdots, \psi_{i-1}$ *having* P*-measure zero*;

(ii) $Pr\{\Phi_i(W_j)$ *has a different relation* $(<, =,$ *or* $>)$ *to* $\Phi_i(W_k)$ *than that of* $\psi_i(W_j)$ *to* $\psi_i(W_k) \mid \Phi_1, \cdots, \Phi_{i-1}\} = 0$.

The functions ψ_i depend on the underlying distribution as seen from their definition below, but are only used as tools in proving the general theorem.

To prove Lemma 9.1 we need the following lemma:

LEMMA 9.2. *Let* $\Phi(w)$ *be a finite sequence of real functions ordered lexicographically (Definition* 7.1*) and let* W *be a chance quantity. Define*

$$\psi(w) = Pr\{\Phi(W) < \Phi(w)\}.$$

(i) *For each value of* $\psi(w)$ *we are able to associate at most one value of* $\Phi(w)$ *with unassociated values of* P*-measure zero.*

(ii) *If* $W_1, \cdots, W_n$ *is a sample of the chance quantity* W, *then with probability one the relation* $(<, =,$ *or* $>)$ *between* $\Phi(W_j)$ *and* $\Phi(W_k)$ *is the same as that between* $\psi(W_j)$ *and* $\psi(W_k)$.

PROOF OF (i). Considering the function $\Phi(w)$, we ask when could two values of it, say Φ', Φ'', correspond to a single value of $\psi(w)$? Since the values of $\Phi(w)$ are ordered, this would mean

$$Pr\{\Phi' \leq \Phi(w) < \Phi''\} = 0,$$

and also any Φ between Φ' and Φ'' would have the same value of ψ. These points would form an interval for φ_1, or, if not, for φ_2, etc., and the P-measure for the interval is zero. To the corresponding value of ψ we associate the $\Phi(w)$ which is the upper limit point of the interval. Since there can be at most a countable number of disjoint intervals on the finite number of real lines and each with P-measure zero, then the P-measure is zero for the values of $\Phi(w)$ which are not associated with values of $\psi(w)$.

The proof of (ii) is given on p. 36 of [5].

The proof of Lemma 9.1 follows easily by using Lemma 9.2. The first part of Lemma 9.2 shows that the definition of $\psi_i(w)$ for values of $\psi_1, \cdots, \psi_{i-1}$ is

[2] See Tukey [4].

unique except for values of $\psi_1, \cdots, \psi_{i-1}$ which have P-measure zero. This establishes the first part of Lemma 9.1. The second part follows directly from Lemma 9.2.

10. The representation theorem. In the proof of the general theorem we can no longer consider the joint distribution of $\{\psi_i(W)\}$ as in [5]. The new proof does not need the general representation theorem of Section 8 in [5] but only the one-dimensional representation theorem in [3].

11. Proof of the general theorem. According to Lemma 9.1, the indices $i(1), \cdots, i(m)$ used to determine the blocks are with probability one the same whether we use the ψ_i or Φ_i. Also, if we consider the blocks themselves, for example

$$S_1' = \{w \mid \Phi_1(w) > \Phi_1(w_{i(1)})\},$$
$$S_1'' = \{w \mid \psi_1(w) > \psi_1(w_{i(1)})\},$$

Lemma 9.1 shows that these differ by a set of P-measure zero and hence have identical coverages. Similarly for the other blocks. Hence it is sufficient to prove our theorem using the real functions $\psi_1, \cdots, \psi_m$.

As in [5] we set up a continuous distribution which can produce by a mapping a distribution equivalent to that of the ψ_i. It happens that for this continuous distribution the functions used to form blocks can be preassigned.

Corresponding to $\psi_1(W)$ we define a function $g_1(U_1)$ of a uniform variate such that the distributions are identical (See [3]). As in Lemma 9.2 there is at most one value of $\psi_1(W)$ for each value of U_1 (if we neglect appropriate points of P-measure zero) and at least one value of U_1 for each value of $\psi_1(W)$. Thus a function depending on the value of $\psi_1(W)$ can just as well be determined by the value of U_1.

For $\psi_2(W \mid \psi_1)$ consider its conditional distribution for values of W restricted as follows: $\psi_1(w) < g_1(u_1)$ or $\psi_1(w) = g_1(u_1)$ with the probability measure of $\psi_1^{-1}(g_1(u_1))$ reduced by the factor α, where

$$\alpha = \frac{u_1 - \inf g_1^{-1}(g_1(u_1))}{\sup g_1^{-1}(g_1(u_1)) - \inf g_1^{-1}(g_1(u_1))}.$$

Define the function $g_2(U_2 \mid u_1)$ of the uniformly distributed random variable U_2 such that its distribution is identical to the above described distribution of $\psi_2(w \mid g_1(u_1))$.

Similarly further functions $g_3(U_3 \mid u_1, u_2), \cdots$ of uniformly distributed random variables $U_1, \cdots$ can be defined.

From the above construction of the mapping of $u_1, u_2, \cdots, u_m$ it is obvious that the mapping of a sample of n from the uniform distribution on the product space $[0, 1]^m$ yields n values of variates to be associated with the $\psi_1, \psi_2, \cdots, \psi_m$. The distribution of the largest value of U_1 is the distribution of $\psi_1(w_{i(1)})$,

and similarly for the others. The mapping has reproduced the part of the distribution of the $\psi_1, \cdots, \psi_m$ in which we are interested. Also we note that the largest value of U_1 yields the largest value of $g_1(U_1)$, etc.

Apply our previous theorem to a sample of n from the uniform distribution on the product space $[0, 1]^m$ with functions $u_1, u_2, \cdots, u_m$ and consider the following sets:

$$S_1' = \{(U_1, \cdots, U_m) \mid U_1 > u_1(i(1))\},$$

$$S_2' = \{(U_1, \cdots, U_m) \mid U_1 < u_1(i(1)),\ U_2 > u_2(i(2))\},$$

$$S_{m|n+1}' = \{(U_1, \cdots, U_m) \mid U_1 < u_1(i(1)), \cdots, U_m < u_m(i(m))\}.$$

Also define:

$$S_1^* = \{g_1(U_1), \cdots, g_m(U_m \mid U_1, \cdots, U_{m-1}) \mid g_1(U_1) > g_1(u_1(i(1)))\},$$

$$S_2^* = \{g_1(U_1), \cdots\cdots\cdots\cdots\cdots \mid g_1(U_1) < g_1(u_1(i(1))),\ g_2(U_2) > g_2(u_2(i(2)))\},$$

$$S_{m|n+1}^* = \{g_1(U_1), \cdots\cdots\cdots\cdots\cdots \mid g_1(U_1) < g_1(u_1(i(1))), \cdots\}.$$

$\bar{S}_1^*, \bar{S}_2^*, \cdots$ are defined as $S_1^*, S_2^*, \cdots$ except $<$ is replaced by $\leq$ and $>$ by $\geq$.

Consider now the inverse mapping of the sets $S_1^*, S_2^*, \cdots$ and $\bar{S}_1^*, \bar{S}_2^*, \cdots$ into the space of $(u_1, u_2, \cdots, u_m)$. We shall have

$$g^{-1}(S_i^*) \subset S_i' \subset g^{-1}(\bar{S}_i^*),$$

because

$$g_i(u_i) > g_i(a) \rightarrow u_i > a \rightarrow g_i(u_i) \geq g_i(a).$$

Thus we have the following inequality for the corresponding coverages:

$$\text{cov } S_i^* \leq \text{cov } S_i' \leq \text{cov } \bar{S}_i^*.$$

The theorem follows directly from these relations and the theorem for the continuous case.

12. Selection of the cutting function Φ_i by a random process. As indicated in the third paragraph of the introduction, the general Theorem 8.1 still remains valid if the functions are chosen by a random process from a class of such functions. The particular class from which Φ_i is chosen may depend on the functions previously selected (Φ_j with $j < i$) and on their values at the respective cuts. The point is that Theorem 8.1 is true for any sequence $\Phi_1, \cdots, \Phi_m$ and consequently is true when the sequence is chosen by a random process.

REFERENCES

[1] S. S. WILKS, "Determination of sample sizes for setting tolerance limits," *Annals of Math. Stat.*, Vol. 12 (1941), pp. 91–96.

[2] A. WALD, "An extension of Wilks' method for setting tolerance limits," *Annals of Math. Stat.*, Vol. 14 (1943), pp. 45–55.

[3] H. SCHEFFÉ AND J. W. TUKEY, "Non-parametric estimation. I. Validation of order statistics," *Annals of Math. Stat.*, Vol. 16 (1945), pp. 187–192.

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

[5] J. W. TUKEY, "Nonparametric estimation, III. Statistically equivalent blocks and multivariate tolerance regions—the discontinuous case," *Annals of Math. Stat.*, Vol. 19 (1948), pp. 30–39.

[6] D. A. S. FRASER AND R. WORMLEIGHTON, "Nonparametric estimation IV," *Annals of Math. Stat.*, Vol. 22 (1951), pp. 294–298.