NON-PARAMETRIC ESTIMATION II. STATISTICALLY EQUIVALENT BLOCKS AND TOLERANCE REGIONS—THE CONTINUOUS CASE

By John W. Tukey

Princeton University

1. Summary. Wald [2, 1943] extended the usefulness of tolerance limits to the simplest multi-dimensional cases. His principle is here used to provide many new ways of using a sample of n to divide the range of the population into $n + 1$ blocks of known behavior. The exact tolerance distribution for the proportions of the population covered by these blocks is extended from the case of a continuous probability density function to the case of a continuous cumulative distribution function. Such an extension is needed in dealing completely with multivariate cases *even* where the underlying distribution is as smooth as a multivariate normal distribution.

The devices used in Paper I [1] to extend the usefulness of tolerance limits to the case of a discontinuous underlying distribution will be applied in the next paper of this series, with some extension, to extend the usefulness of these general tolerance regions to the case of a discontinuous distribution. Some of these results specialize into new results for the univariate case, although they do not seem to have any immediate practical application.

The author wishes to acknowledge the stimulation given to his work on this problem by Henry Scheffé, whose modesty has kept this paper from the joint authorship of papers I [1, Scheffé and Tukey 1945] and IV (not yet written).

2. Introduction. Wald's great contribution to the theory of tolerance limits was his method of successive elimination. As originally presented for a bivariate situation it ran roughly as follows: Let $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$ be a sample of n from an arbitrary bivariate population. The type of tolerance region to be used is determined by four preassigned integers, k_1, k_2, k_3, and k_4. The procedure is as follows: Order the n observations according to their x values. Select the k_1 highest, and let the x coordinate of the lowest of these k_1 be x_u. Select the k_2 lowest, and let the x coordinate of the highest of these k_2 be x_l. Discard these $k_1 + k_2$ selected observations, and order the remaining $n - k_1 - k_2$ observations according to their y values. Select the k_3 highest of these remaining observations, and let the y coordinate of the lowest of these k_3 be y_u. Select the k_4 lowest of these remaining observations, and let the y coordinate of the highest of these k_4 be y_l. The tolerance region, consisting of all points (x, y), with $x_l < x < x_u$ and $y_l < y < y_u$ depends on the sample, and, hence, so does the fraction of the population falling in (= covered by) this region. Wald showed that the distribution of this fraction covered was independent of the underlying bivariate distribution, so long as this latter distribution had a continuous probability density function. He showed that the distribution was the same as that arising in the one-dimensional case when a tolerance region was set with the aid of $k_1 + k_2 + k_3 + k_4$ observations. (Numerical approximation to these distributions will be discussed in Paper IV of this series.

The important device in this process, and the one which makes the conclusion possible, is the discarding of the $k_1 + k_2$ observations after they have played their part by determining x_l and x_u.

We shall shortly be able to describe this procedure of Wald's as a special case of a more general procedure, but we shall first go back to the simplest one dimensional case to explain some of our notions and terminology.

Consider the uniform distribution from 0 to 1, draw a sample of n, and let the sample values, ordered according to size be $t_1, t_2, \cdots, t_n$. These n values divide the interval from 0 to 1 into the following $n + 1$ parts $(0, t_1)$, (t_1, t_2), $\cdots$, $(t_{n-1}t_n)$, $(t_n, 1)$ which we shall call *blocks*. Since the joint distribution of the t_i is well known, that of the lengths of these $n - 1$ blocks is easily found. This distribution of lengths would be unimportant, if it were not at the same time the distribution of the fractions of the population covered by the blocks. As is shown later, this distribution of fractions covered, or, more simply, of *coverages*, has the following properties:

(i) the fractions covered add up to 1.

(ii) the distribution is completely symmetrical.

Property (ii) makes intuitive the result of Wilks [3, 1941] that the distributions of the coverage of regions obtained

(a) by removing the $k_1 + k_2$ left-most blocks,

(b) by removing the k_1 left-most and the k_2 right-most blocks

are identical. The specific distribution obtained satisfies

(iii) if the coverages are taken as barycentric coordinates on an n-simplex, the distribution over the simplex is uniform,

(iv) the sum of the coverages of any k *pre*selected blocks of the $n + 1$ has the well-known distribution

$$Pr\,\{\text{sum of } k \text{ coverages} < t\} = I_t\,(n - k + 1, k)$$

where $I_\beta\,(n, m)$ is the incomplete Beta function.

We shall call a set of blocks, derived from a sample, whose coverages behave in this general way a set of *statistically equivalent blocks*. Normally this will be abbreviated to se-blocks. (A precise definition is given in section 4.)

We shall concentrate much of our attention on all the blocks and their symmetrical character, rather than on the tolerance region formed by deleting k of them, since our results will then be applicable to many other problems.

Now we can generalize Wald's original procedure. Let $W_1, W_2, \cdots, W_n$ be a sample of n—we shall not need to consider its distribution—and let φ_1, $\varphi_2, \cdots, \varphi_n$ be n numerically valued functions of W, possibly alike, possibly distinct, such that $\varphi_1(W), \varphi_2(W), \cdots, \varphi_n(W)$ have a joint distribution. Proceed as follows:

Order the W_i according to the numbers $\varphi_1(W_i)$, select the W_i for which $\varphi_1(W_i)$ is largest and denote it by $W_{i(1)}$. The first block contains all W such that

$$\varphi_1(W) > \varphi_1(W_{i(1)}). \tag{2.1a}$$

Discarding $W_{i(1)}$, order the remaining W_i according to the values of $\varphi_2(W_i)$, and select as $W_{i(2)}$ the one giving the largest value. The second block contains all W such that

$$\begin{aligned} \varphi_1(W) &< \varphi_1(W_{i(1)}), \\ \varphi_2(W) &> \varphi_2(W_{i(2)}). \end{aligned} \tag{2.1b}$$

Continue this process. The mth block, for $m \leq n$ will be defined by

$$\begin{aligned} \varphi_j(W) &< \varphi_j(W_{i(j)}), \qquad j = 1, 2, \cdots, m-1, \\ \varphi_m(W) &> \varphi_m(W_{i(m)}), \end{aligned} \tag{2.1m}$$

and the $(n + 1)$st block by

$$\varphi_j(W) < \varphi_j(W_{i(j)}), \qquad j = 1, 2, \cdots, n. \tag{2.1n}$$

(A graphical example of this construction is given shortly.) This set of $n + 1$ blocks will be statistically equivalent whenever the cumulative distribution of each φ_i function is continuous.

To specialize this to the case described above, let W be a pair (x, y) of numbers and let

(i) the first k_1 φ's be the x-coordinate of W,
(ii) the next k_2 φ's be *minus* the x-coordinate of W,
(iii) the next k_3 φ's be the y-coordinate of W,
(iv) the next k_4 φ's be *minus* the y-coordinate of W,
(v) the remaining φ's be arbitrary.

Then the first k_1 blocks will contain all W for which

$$x = \varphi_j(W) > \varphi_j(W_{i(j)}), \qquad j = 1, 2, \cdots, k_1$$

that is, for which

$$x > x_u = \varphi_{k_1}(W_{i(k_1)}).$$

Similarly, the next $k_2 + k_3 + k_4$ blocks will contain all W with

$$\begin{aligned} &x < x_l, \\ &y > y_u, \qquad x_l \leq x \leq x_u, \\ &y < y_l, \qquad x_l \leq x \leq x_u, \end{aligned}$$

respectively, and the removal of these $k_1 + k_2 + k_3 + k_4$ blocks leaves Wald's tolerance region (plus the boundaries where $x = x_u$, $x = x_l$, $y = y_u$, $y = y_l$).

There would be no point in this more general wording, if it did not include new cases of some interest. We give now, in graphic terms, an example of such a case.

We deal with a sample of n bivariate observations, which we think of as plotted on a *map* so that we can use geographical language. The number n is rather large, and we wish to construct a tolerance region by deleting 12 blocks. We proceed as follows:

Find the most northerly point, draw an East-West line through it, and shade the area North of the line. Find the most easterly point in the unshaded area, draw a North-South line through it, and shade the unshaded area East of the

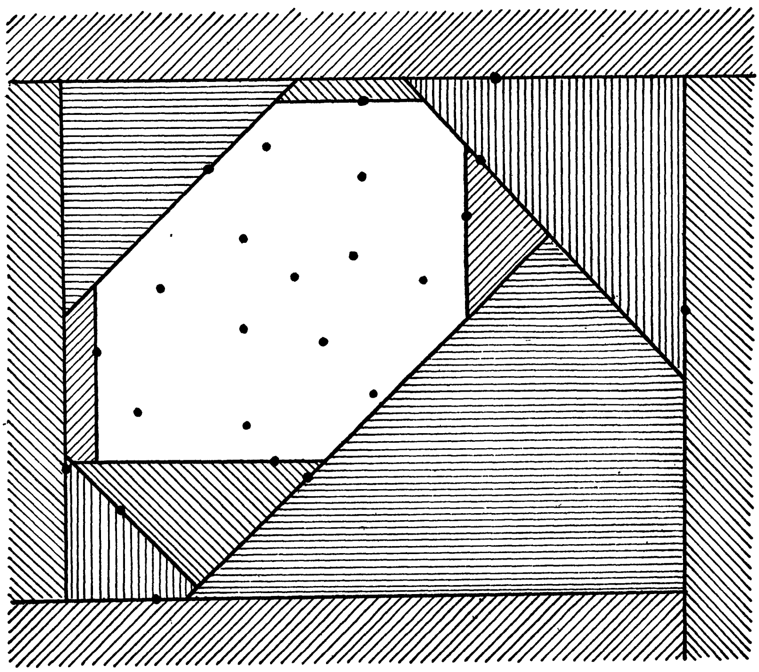

FIG. 1

line. Find the most southerly point, (always working in the unshaded area), draw an East-West line through it and shade the area South of the line. Find the most westerly point, draw a North-South line through it, and shade the area West of the line. Find the most northeasterly point, draw a NW-SE line through it and shade the area northeast of the line. Find the most southeasterly point, draw a NE-SW line through it, and shade the area southeast of the line. Repeat this 6 times more, choosing in succession the most southwesterly, northwesterly, northerly, easterly, southerly, and westerly points. The remaining points will now lie in an unshaded area surrounded by a polygon, which will have 8 (or perhaps fewer) sides. The inside of this polygon is the desired tolerance region.

Figure 1 shows the final result, starting from $n = 25$. The practicing statistician is invited to try an example of his own with n at least 100.

Other newly accessible cases can easily be invented by the reader, after he considers this example carefully.

The use of a single W and n functions φ_i has two virtues; it simplifies notation and frees the intuition, as compared with the use of n chance quantities $Z_i = \varphi_i(W)$.

If the bivariate situation above were regarded as a 12-variate situation, where the variates were, in order, $(y, x, -y, -x, x + y, x - y, -x - y, -x + y, y, x, -y, -x)$ then the original Wald procedure with $k_1 = k_3 = \cdots = k_{23} = 1$; $k_2 = k_4 = \cdots = k_{24} = 0$ would apply to construct the same region. Yet even if x and y had a bivariate normal distribution, Wald's proof would not apply without extension. For the 12-dimensional distribution is highly singular (it is concentrated on a 2-dimensional plane in 12-dimensional space) and there is no hope of a density function. An extension of Wald's result to the case where the 12-dimensional joint cumulative distribution function is continuous —as is the case in this example when x and y have a continuous joint cumulative —is clearly needed.

When we come to deal with the case of where the cumulative needs not be continuous we shall meet a further difficulty, namely "ties". But if, as in the present case, the cumulative is continuous, it is easy to see that the probability that $\varphi_i(W_j) = \varphi_i(W_k)$ for any i, j, k is zero.

3. Terminology and notation. A quantity which has a probability distribution we call a *chance quantity* (it has frequently been called a *random variable*). The term chance quantity does not imply that its values are single real numbers, they may be single real numbers (when we also speak of a real chance quantity), sets of n real numbers, or more general objects. The cumulative distribution function, or *cumulative*, of a single real chance quantity, X, is defined by

$$F(t) = Pr\{X < t\},$$

except perhaps at the discontinuities of F. We have used here the notation $Pr\{k(X)\}$ to indicate the probability that $k(X)$ holds, and we have followed our policy of using capital letters for chance quantities and the corresponding lower case letters for their values.

The set of values of W, or, as we shall say, the W-set, for which, for example $\varphi(W) \leq 3$, will be denoted by

$$\{W \mid \varphi(W) \leq 3\}.$$

We shall wish to compute probabilities associated with one or more functions of a chance quantity; usually we will emphasize that these functions shall be measurable with respect to the probability measure underlying the distribution of W by asserting that they have a joint cumulative, which is defined by

$$F(t_1, t_2, \cdots, t_k) = Pr\{\varphi_k(W) < t_k\},$$

(except possibly at discontinuities of F) and which does not exist unless the φ_i are measurable with respect to the unknown underlying distribution of W. In cases where we neglect to remind the reader, it is still assumed that the functions are measurable.

The coverage of a W-set, which may itself be a chance quantity, is defined by

$$\text{Coverage of } S = Pr\{W \epsilon S\}.$$

When S is a chance quantity, its coverage is also a chance quantity. The barycentric simplex (of dimension n) is the set of points in $n + 1$-dimensional Euclidean space $(t_1, t_2, \cdots, t_{n+1})$ with $t_1 + t_2 + \cdots + t_{n+1} = 1$ and $0 \leq t_i \leq 1$. The name comes from the representation of the point $(t_1, t_2, \cdots, t_{n+1})$ as the center of gravity (in mechanical terms) or mean (in statistical terms) of the distribution where a fraction t_i is concentrated at the ith vertex. (In order, the vertices are $(1, 0, 0, \cdots, 0)$, $(0, 1, 0, \cdots, 0)$, etc.) The uniform distribution on this simplex has an (n-dimensional) density

$$n!dt_1dt_2 \cdots dt_n, \qquad (0 \leq t_1, t_2, \cdots, t_n, 1 - t_1 - t_2 \cdots - t_n \leq 1),$$

and the cumulative

$$T(x_1, x_2, \cdots, x_{n+1}) = n! \int\int \cdots \int dt_1 dt_2 \cdots dt_n$$

where the integration is over the range where $0 \leq t_i \leq x_i$ *and at the same time* $t_1 + t_2 + \cdots + t_{n-1} \leq 1$.

4. The blocks determined by n values of W. We deal now with a population of W's (a probability measure μ on the space $T \equiv \{w\}$), a family of functions $\varphi_1, \varphi_2, \cdots, \varphi_m$ of W with a joint cumulative (measurable with respect to μ) and a set of values $w_1, w_2, \cdots, w_n$, $(w_i \epsilon T)$.

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

$$S_1 = \{w \mid \varphi_1(w) > a_1\} \tag{4.2}$$

where $a_1 = \max_i \varphi_1(w_i) = \varphi_1(w_{i(1)})$, *which defines* $i(1)$.

$$S_2 = \{w \mid \varphi_1(w) < a_1, \varphi_2(w) > a_2\}, \tag{4.3}$$

where $a_2 = \max_{i \neq i(1)} \varphi_2(w_i) = \varphi_2(w_{i(2)})$, $i(2) \neq i(1)$, *which defines* $i(2)$. *And in general, for* $1 < k \leq \min(m, n)$,

$$S_k = \{w \mid \varphi_1(w) < a_1, \cdots, \varphi_{k-1}(w) < a_{k-1}, \varphi_k(w) > a_k\}, \tag{4.4}$$

where $a_k = \max_i' \varphi_k(w_i) = \varphi_k(w_{i(k)})$, *the maximum being taken over all* i except $i(1), i(2), \cdots, i(k-1)$; and $i(k)$ *being chosen distinct from all* $i(j), j < k$.

If $m \geq n$, *then*

$$S_{n+1} = \{w \mid \varphi_1(w) < a_1, \cdots, \varphi_n(w) < a_n\}. \tag{4.5}$$

If $m \leq n$, *then*

$$S_{m|n+1} = \{w \mid \varphi_1(w) < a_1, \cdots, \varphi_m(w) < a_m\}. \tag{4.6}$$

The result of this definition is to use $w_1, \cdots, w_n$ and $\varphi_1, \cdots, \varphi_m$ to define $n + 1$ blocks (one more than there are w's) in case there are enough functions, and, in case there are not enough functions, to define one small block, S_i, for each function plus one large remainder $S_{m|n+1}$. We notice

(4.2) REMARK. *The blocks of* (4.1) *are well defined unless* $\varphi_1(w_j) = \varphi_i(w_k)$ *for some* i, j, k.

5. Statement of results for the statistician. The central results can be stated as follows:

(5.1) THEOREM $A_{m|n+1}$. *If* $W_1, W_2, \cdots, W_n$ *are a sample of* n *from a distribution, if* $\varphi_1, \varphi_2, \cdots, \varphi_m$, $(m \leq n)$, *are* m *functions such that*

$$\varphi_1(W), \varphi_2(W), \cdots, \varphi_m(W)$$

have a joint distribution which has a continuous cumulative, and if the blocks $S_1, S_2, \cdots, S_m$ *and* $S_{m|n+1}$ *are defined as in* (4.1), *then*

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

(ii) *the distribution of the coverages*

$$c_i = Pr\{w \text{ in } S_i\}, \qquad i = 1, 2, \cdots, m$$

and

$$c_{m|n+1} = Pr\ \{w \text{ in } S_{m|n+1}\}$$

is the same as that of $t_1, t_2, \cdots, t_m$ *and* $t_{m+1} + t_{m+2} + \cdots + t_{m+1}$ *where* t_i *are uniformly distributed on the barycentric simplex with* $n + 1$ *vertices.*

Conditions (5.1i) and (5.1ii) are the precise definition of *a partial family of statistically equivalent blocks of type* $n + 1$ *and an associated* $(m \mid n + 1)$ *tolerance region.*

(5.2) THEOREM B_{n+1}. *If* $W_1, W_2, \cdots, W_n$ *are a sample of* n *from a distribution, and if* $\varphi_1, \varphi_2, \cdots, \varphi_m$, $(m \geq n)$, *are* m *functions such that*

$$\varphi_1(W), \varphi_2(W), \cdots, \varphi_m(W)$$

have a joint distribution which has a continuous cumulative, and if the blocks $S_1, S_2, \cdots, S_{n+1}$ *are defined as in* (4.1), *then*

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

(ii) *the distribution of the coverages*

$$c_i = Pr\,\{w \text{ in } S_i\}, \qquad i = 1, 2, \cdots, n + 1$$

is the same as that of $t_1, t_2, \cdots, t_{n+1}$, *where the* t_i *are uniformly distributed on the barycentric simplex with* $n + 1$ *vertices.*

Conditions (5.2i) and (5.2ii) are the precise definition of *a complete family of statistically equivalent blocks*. In Paper III we shall have to widen these notions a little, and this form will then be qualified by the phrase "in the narrow sense".

6. Statement of results for the measure theorist. The construction of (4.1) maps the product $T^n \times U^n$ into E^{n+1} where T is the set of w's (and hence T^n is the set of ordered n-tuples of w's), U is the space of all real-valued functions defined over T, measurable with respect to a fixed probability measure μ, and possessing a continuous cumulative, (i.e. $\mu(\{w \mid \varphi(w) = c\}) = 0$ for all real c), and hence U^n is the space of ordered n-tuples of such functions, and E^{n+1} is Euclidean n-dimensional space. More precisely, the mapping is into the barycentric simplex with $n + 1$ vertices, a subset of E^{n+1}, and is well defined except for a set in T^n of measure zero with respect to μ^n, the power measure of μ. In these terms, we may restate theorem B as follows:

(6.1) THEOREM B_{n+1}. *Hold the n functions* $\varphi_1, \varphi_2, \cdots, \varphi_n$ *and the probability measure fixed, then* T^n *is mapped into* B_n *and the power measure* μ^n *is carried by that mapping into a measure on* B_n. *This measure is always* $n!$ *times Lebesgue measure.*

7. Wald's principle. The essential principle behind Wald's process of discarding observations is sufficiently fundamental to warrant a name of its own. It can be stated, quite generally, in the two following forms:

(7.1) WALD'S PRINCIPLE. (discrete form.) *Let* W *be a chance quantity, and consider samples of n. Fix disjoint w-sets* $A_1, A_2, \cdots, A_m, B$. *Consider those samples of n for which exactly one value falls in each* A_i *and the remaining n-m fall in B. The distribution of the n-m falling in B is that of a random sample of n-m from the distribution of W restricted to B.* (i.e. $\mu_B(X) = [\mu(B)]^{-1}\mu(BX)$.)

(7.2) WALD'S PRINCIPLE. (conditional form.) *Let* W *be a chance quantity, and* φ *a function such that each value of* $\varphi(W)$ *has probability zero. Consider samples of n. Then the conditional distribution of the* w_i, *given that*

$$\max_i \varphi(w_i) = a,$$

is that of one w_{i_0} *with* $\varphi(w_{i_0}) = a$ *and a sample of* $n-1$ *other* w_i *from the distribution of W restricted to* $B = \{w \mid \varphi(w) < a\}$.

(7.3) CENTRAL LEMMA. *Let* W *be a chance quantity and let* $\varphi_1, \cdots, \varphi_n$ *be functions with a joint cumulative such that* $\varphi_i(w) = a$ *has probability zero for each i and a (i.e. the joint cumulative is continuous). Then the conditional distribu-*

tion of the remaining $n-k$ w's, after k blocks have been chosen according to (4.1) *is that of a sample from the distribution of W restricted to*

$$B = \{w \mid \varphi_1(w) < a_1, \cdots, \varphi_k(w) < a_k\},$$

where $k = 1, 2, \cdots, n$.

The proofs of these statements are elementary and direct. To establish (7.1) we have only to show that given two sets in B^{n-k}, their probabilities on the assumption that one w_i is in each A_i are in the ratio of their probabilities for an unrestricted sample of $n-k$. But the probability of finding the $n-k$ w_i in a set R, contained in B^{n-k}, and one w_i in each A_i, is exactly

$$\frac{n!}{(n-k)!}\mu(A_1)\mu(A_2)\cdots\mu(A_k)$$

times the probability that $n-k$ w_i, known to be in B^{n-k}, will fall in R. This establishes (7.1).

In order to prove (7.2) we must show that the probability of a set R of n-tuples $w_1, w_2, \cdots, w_n$ is the same whether calculated directly or calculated by the proposed conditional distribution. To this end, it is natural to decompose R as follows:

$$R = R(1) + R(2) + \cdots + R(n) + Z,$$

where $R(i)$ contains those $(w_1, \cdots, w_n)$ in R for which $\varphi(w_i) > \varphi(w_j)$ for all $j \neq i$, and Z contains the remaining $(w_1, \cdots, w_n)$; which must involve at least one tie $\varphi(w_j) = \varphi(w_k)$, $j \neq k$. Since Z has probability zero, it will suffice to establish the equality of the two calculations for sets of the form $R(i)$, and because of symmetry we may restrict ourselves to sets of the form $R(1)$.

Given an integer N, we decompose the range of $\varphi(w)$ into Nn segments of equal probability, which we may do because the cumulative of φ is continuous. There are then Nn values b_k, $(b_o = -\infty, b_{Nn} = +\infty)$ such that

$$Pr\{b_{k-1} < \varphi(w) < b_k\} = 1/Nn.$$

We now decompose our set R (which is of the form $R(1)$ as follows:

$$R = R_2 + \cdots + R_{Nn} + Y,$$

where R_k contains those n-tuples

$$(w_1, \cdots, w_n) \text{ for which } b_{k-1} < \varphi(w_1) < b_k$$

and $\varphi(w_i) < b_{k-1}$ for all $i > 1$. The remaining set Y contains n-tuples where the two largest $\varphi(w_i)$, $(i = 1$ and $i = i_o)$, belong to the same interval. The probability of this is less than

$$\frac{n(n-1)}{2}\left(\frac{1}{nN}\right)^2 \leq \frac{1}{2N^2}$$

as calculated from the known distribution. Calculating from the conditional distribution, we find immediately a bound of

$$\sum_k \left\{ \left(\frac{k}{Nn}\right)^n - \left(\frac{k-1}{Nn}\right)^n \right\} \frac{n-1}{k} = \frac{n-1}{(Nn)^n} \sum \frac{k^n - (k-1)^n}{k}$$

$$= \frac{n-1}{(Nn)^n} \left((nN)^{n-1} + \sum_{k<nN} \frac{k^{n-1}}{k-1} \right)$$

$$\leq \frac{n-1}{(Nn)^n} (A_n (nN)^{n-1})$$

$$= \frac{A_n}{N},$$

where A_n is a constant depending only on n. Thus, as N increases, the probability of the successive sets Y tend to zero—calculated either way. To show the equivalence of the two calculations it is now sufficient to show that they agree for the sets R_k. But this is a case of (7.1) and the lemma is proved.

Now (7.3) follows by induction, applying (7.2) at each step.

8. Proof of theorems. We notice that Theorem B_n is equivalent to Theorem $A_{m|n+1}$, since, according to (4.1) $S_{n|n+1} = S_{n+1}$.

We have only to prove theorem $A_{m|n+1}$, which we do by induction on m. For $m = 1$, it is exactly Wilks' [3, 1941] original one-dimensional theorem, and is known. Let us assume it for $m = k$ and demonstrate it for $m = k + 1$, for by induction this will complete the proof.

We must deal with the blocks $S_1, S_2, \cdots, S_k, S_{k+1}$ and $S_{k+1|n+1}$, (notation as in (4.1) and (5.1)). We need the obvious

(8.1) Lemma. *Since the cumulative of* φ_{k+1} *is continuous, the union of* S_{k+1} *and* $S_{k+1|m+1}$ differs from $S_{k|m+1}$ *by a set of zero probability.*

Hence

$$c_{k|n+1} \equiv c_{k+1} + c_{k+1|n+1}.$$

Since we know from the induction hypothesis that $c_1, c_2, \cdots, c_k$ and $c_{k|n+1}$ have the correct joint distribution, we have only to show that c_{k+1} and $c_1, c_2, \cdots, c_k$ have the correct joint distribution. Fix $c_1, c_2, \cdots, c_k$. Then $a_1, a_2, \cdots, a_k$ must be fixed, and so (7.3) applies to the $n-k$ w_i's not discarded after $a_1, a_2, \cdots, a_k$ have been fixed. The conditional distribution of c_{k+1} must be that of a fixed number $(1 - c_1 - c_2 - \cdots - c_k)$, which is the probability attached to $S_{k|n+1}$, times the coverage of one block based on a sample of $n-k$, since the remaining $n-k$ w's behave like a sample.

Consider the very particular case where w is uniformly distributed between zero and one and $\varphi_i(w) \equiv w$, all that we have said in the last paragraph applies —the conditional distribution of c_{k+1} given $c_1, c_2, \cdots, c_k$ is the same in the two cases—hence the joint distribution of $c_1, c_2, \cdots, c_k, c_{k+1}$ is the same in both cases—but in this very particular case the joint distribution is known to be that required by theorem $A_{k+1|n+1}$.

REFERENCES

[1] H. Scheffé and J. W. Tukey, "Non-parametric estimation. I. Validation of order statistics", *Annals of Math. Stat.*, Vol. 16 (1945), pp. 187–192. (Also cited as Paper I.)

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

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