

NONPARAMETRIC TOLERANCE REGIONS

BY D. A. S. FRASER

University of Toronto

1. Summary. Nonparametric tolerance regions can be constructed from statistically equivalent blocks using published graphs by Murphy [8]. In this paper the procedure for obtaining the statistically equivalent blocks is generalized. The n 'cuts' used to form the $n + 1$ blocks need not cut off one block at a time, but at each stage may cut off a group of blocks, the group to be further divided at a later stage by a different type of cut in general. An example is given which indicates possible applications.

The results are also interpreted for discontinuous distributions by indicating the necessary modifications to the corresponding theorem in [7].

2. Introduction. The generality with which nonparametric tolerance regions can be formed has been successively treated by Wilks, Wald, Scheffé, Tukey, and others in a series of papers [1], [2], [3], [4], [5], [6], [7]. In each case the sample space for n observations from a continuous distribution is divided by these observations into $n + 1$ regions or blocks. Subject to mild restrictions on the procedure used to divide the sample space, the proportions of the population contained in these regions have an elementary distribution, a uniform distribution over a set prescribed by simple inequalities. Furthermore the marginal distribution of the proportion of the population which lies in a group of these regions has the Beta distribution. This enables the statistician to choose enough regions to make a probability statement such as the following: "In repeated sampling the probability is β that the region T contains at least α of the population." Graphs for obtaining the probabilities and the number of original regions to compound are given in a paper by Murphy [8].

In the previous papers the sample space was partitioned by forming a single block at a time; this restriction, however, is not necessary. The whole region corresponding to $n + 1$ blocks may be divided into r blocks and $n + 1 - r$ blocks; then each of these sets of blocks may be divided by a procedure depending on where and how the first division was made. The exact statement of the possible procedures is given by Theorem 6.1 in Section 6.

Advantages of this procedure are perhaps illustrated by the following example. A sample of 59 observations is made from a continuous bivariate distribution known to have two modes; a 50% tolerance region in two parts centering on the two modes is desired. From Murphy's graphs [8] a region formed from 36 blocks¹ is seen to have a 90% probability of containing at least 50%

Received 12/6/51.

¹ It is worth noticing that only 60% of the equivalent blocks yields 90% confidence in at least 50% of the population.

of the population, that is, 90% confidence that the region contains at least 50% of the population. The following procedure is proposed as a solution to obtaining the required region.

The 59 points are plotted in Fig. 1. The function y is used to remove two blocks by the cut c_2 ; two further blocks using the function $-y$ are removed by the cut c_4 . Similarly x and $-x$ are used to form cuts c_6 and c_8 . The rectangle so formed now corresponds to 52 blocks.

The rectangle is tentatively cut into eight sections formed by the two diagonals and the two lines through the center parallel to the x and y axis. For convenience, number these sections from one to eight clockwise starting at top center. In the first section cut off one block from the outside using a line making an angle of $-22\frac{1}{2}^\circ$ to the x axis, that is, we use the function $y + x \tan 22\frac{1}{2}^\circ$ to form the cut

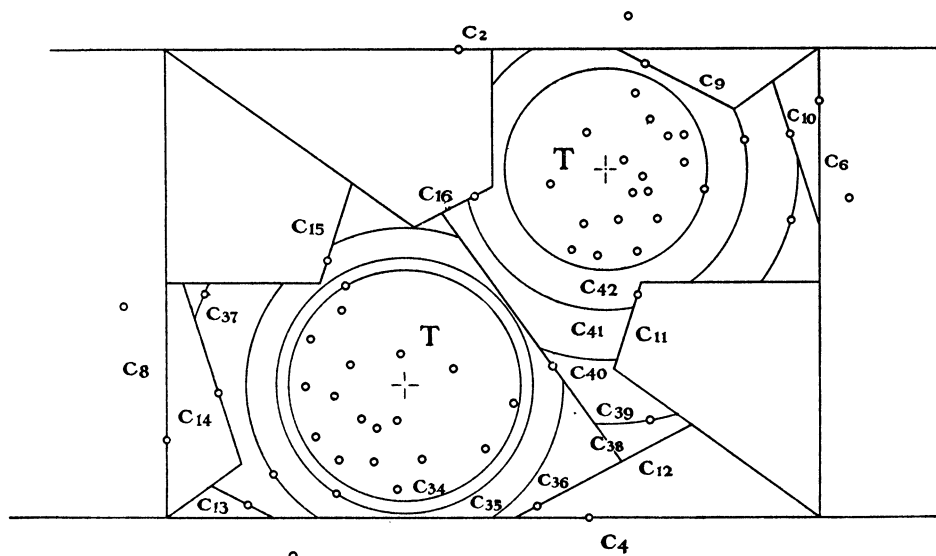


FIG. 1

c_9 . For the second section use the function $y + x \tan 67\frac{1}{2}^\circ$ to remove one block by the cut c_{10} . Apply a similar procedure to each of the 6 remaining sections, thus forming cuts c_{11} , c_{12} , c_{13} , c_{14} , c_{15} , and c_{16} . The region now remaining corresponds to 44 blocks.

The eight sections originally were of equal area. Each section has had a block removed thus reducing the areas to the values a_1, \dots, a_8 , say. Further cutting will depend on these areas, they being an indication of the relative positions of the two modes. Consider the total area of an adjacent pair of reduced sections and of the opposite pair; for example, total area equals $a_1 + a_2 + a_5 + a_6$. Do this for each of the four possible selections. From the diagram it is easily seen that the group with minimum total area corresponds to Sections 3, 4, 7, 8. These are the sections which presumably tend to separate the two modes; hence we

divide the remaining region by a line with slope -1 . If the blocks had been 2, 3, 6, 7 we would have used a line with slope 0. The rough reasoning behind this cutthroat procedure should be apparent.

Using the function $y + x$, divide the 44 block region into parts corresponding to 22 blocks and 22 blocks, that is, we choose the point giving the 22nd largest value to the function $y + x$ and make the cut c_{38} . The two regions formed by this cut are further reduced with the objective being to form two circular regions each corresponding to 18 blocks.

Use the function $(y - \eta)^2 + (x - \xi)^2$ to remove four blocks from the right-hand region. As center of the circle (ξ, η) a reasonable choice would be the center, marked 'x', of the largest circle which can be inscribed in that region. Cuts c_{39} , c_{40} , c_{41} , and c_{42} are made by this function. We apply a similar procedure to the left-hand region.

The resultant two circles form a region T composed of 36 blocks and hence in repeated sampling have 90% confidence of being at least a 50% tolerance region. It should be noted that the two parts of T will not always be circular; they will be circular with perhaps indentations. (See, for example, cut c_{41} .)

3. Notation. Let W symbolize a probability distribution over a general space S and let w be an arbitrary point of this space. By the coverage of a set $S \subset S$ we mean the probability measure of S with respect to the distribution W , that is, $P_w(S) = P(S)$ is the coverage of the set S . If the set S is random, then the coverage will be random.

4. Conditional probabilities when a sample is ordered by a real function. For the proofs in Section 5 we shall need to know the form of certain conditional distributions. In [7] these conditional distributions were assumed without proof, here they are more complicated and a proof of their structure is given in this section.

Let $\psi(w)$ be a real-valued measurable function over the space S . For a sample of n elements from this distribution of W over S , we wish to determine the conditional distribution of the sample given that the j th largest value of $\psi(w)$ is equal to t . If several values of w_i have $\psi(w_i) = t$, the procedure in ordering these sample elements is to make each permutation equally likely and select a permutation at random. That the conditional distribution exists is easily seen from the fact that in a product space with a product measure the conditional distribution of $n - 1$ coordinates obtained by conditioning the remaining coordinate is just the marginal distribution of those $n - 1$ coordinates.

Consider (w_1, \dots, w_n) as a point in the product space $\prod_{i=1}^n S_i$, where each S_i is identical to S . The probability measure over this space will be the power measure of the given measure of W over S . We partition the space S into disjoint sets as follows:

$$S = \bar{S} \cup J \cup S,$$

where

$$\bar{\mathfrak{S}} = \{w \mid \psi(w) > t\},$$

$$\mathfrak{J} = \{w \mid \psi(w) = t\},$$

$$\underline{\mathfrak{S}} = \{w \mid \psi(w) < t\}.$$

The conditional distribution we wish to obtain will be a distribution over a subset of the following region in the product space:

$$X = \bigcup_{j=1}^n \left\{ \prod_{i=1}^{j-1} \mathfrak{S}_i \times \mathfrak{J}_j \times \prod_{i=j+1}^n \mathfrak{S}_i \right\}.$$

The components of this union are not disjoint. Consider the following decomposition into disjoint sets:

$$X = X_1 \cup X_2 \cup \cdots \cup X_n,$$

where

$$\begin{aligned} X_1 &= \bigcup_{j=1}^n X_{1j} = \bigcup_{j=1}^n \left\{ \prod_{i=1}^{j-1} (\mathfrak{S}_i - \mathfrak{J}_i) \times \mathfrak{J}_j \times \prod_{i=j+1}^n (\mathfrak{S}_i - \mathfrak{J}_i) \right\}, \\ X_2 &= \bigcup_{j=1}^n X_{2j} = \bigcup_{j=1}^n \left\{ \bigcup_{k \in I_j} [\mathfrak{J}_j \times \mathfrak{J}_k \times \prod_{i \in I_j - (k)} (\mathfrak{S}_i - \mathfrak{J}_i)] \right\}, \\ &\vdots \\ X_{r+1} &= \bigcup_{j=1}^n X_{r+1j} = \bigcup_{j=1}^n \left\{ \bigcup_{k_1 < \cdots < k_r \in I_j} \left[\mathfrak{J}_j \times \prod_{\alpha=1}^r \mathfrak{J}_{k_\alpha} \times \prod_{i \in I_j - (k_1, \dots, k_r)} (\mathfrak{S}_i - \mathfrak{J}_i) \right] \right\}, \\ &\vdots \\ X_n &= \bigcup_{j=1}^n X_{nj} = \bigcup_{j=1}^n \left\{ \mathfrak{J}_j \times \prod_{i \in I_j} \mathfrak{J}_i \right\} \\ &= \prod_{i=1}^n \mathfrak{J}_i, \end{aligned}$$

and $I_j = (1, \dots, j-1, j+1, \dots, n)$. The sets X_1, \dots, X_n are disjoint but the components of each X_i ($i = 1$ excepted) are not disjoint.

The advantage of this decomposition is in the symmetry possessed by the sets X_{ij} . X_{ij} is symmetric in the $n-1$ coordinates other than w_j ; also X_{ij} is identical to $X_{i'j}$ if we consider only the $n-1$ coordinates remaining after deleting respectively w_j or $w_{j'}$. Therefore consider the following decomposition:

$$X = X^1 \cup X^2 \cup \cdots \cup X^n,$$

where

$$X^j = \bigcup_{i=1}^n X_{ij}, \quad j = 1, \dots, n.$$

The sets X^1, X^2, \dots, X^n are symmetric in the $n-1$ coordinates obtained by omitting w_1, w_2, \dots, w_n , respectively; also as far as these $n-1$ coordinates are concerned the sets are identical. However, the sets are not disjoint, but

overlap in a simple manner indicated by the assignment of probability measure below.

Since X is composed of n identical components as far as the " $n - 1$ coordinates" are concerned, we can treat any one of them, say X^1 . Assign a probability measure to X^1 as follows: Let P_1^* be the measure such that the measure over X_{11} is the original power measure over S^n , the measure over X_{21} is the original power measure over S^n reduced by the factor $\frac{1}{2}$, the measure over X_{i1} is the original power measure over S^n reduced by the factor $1/i$, and the measure over X_{n1} is the original power measure over S^n reduced by the factor $1/n$. When the n identical cases X^1, \dots, X^n are compounded the original measure over X is reproduced; the sets overlap in such a manner that for points with measure reduced by the factor $1/i$, i sets overlap reproducing the original measure. If the probability measure of \mathfrak{J} is equal to zero, we need only consider the first set X_{11} . Its measure is zero, but if we consider the $n - 1$ coordinates, w deleted, the marginal measure gives us the conditional measure of those $n - 1$ elements of the sample. This particular case is covered again after the following simplification of the general case.

Since we are interested only in the $n - 1$ coordinates and since the distribution is homogeneous with respect to the n th coordinate, we now work in the product space of those $n - 1$ coordinates. The measures P_i^* will be altered by a factor if $P(\mathfrak{J}) \neq 0$; otherwise we have the particular case mentioned above and the conditional distribution is identical to the marginal. Therefore, letting $I = (2, \dots, n)$, we have

$$\begin{aligned}\bar{X}_1 &= \prod_{i \in I} (S_i - \mathfrak{J}_i), \\ \bar{X}_2 &= \bigcup_{k \in I} [\mathfrak{J}_k \times \prod_{i \in I - (k)} (S_i - \mathfrak{J}_i)], \\ &\vdots \\ \bar{X}_n &= \prod_{i \in I} \mathfrak{J}_i;\end{aligned}$$

and \bar{P}^* is the measure such that the measure over \bar{X}_1 is the original measure over S^{n-1} , the measure over \bar{X}_2 is the original measure over S^{n-1} reduced by the factor $\frac{1}{2}$, the measure over \bar{X}_n is the original measure over S^{n-1} reduced by the factor $1/n$.

The conditional distribution for which we are looking is obtained from the distribution over the space $Y \subset \bigcup_{j=1}^n \bar{X}_j$, where

$$Y = \bigcup_{\underline{P}} \left\{ \prod_{i=2}^j (\bar{S}_i \cup \mathfrak{J}_i) \times \prod_{i=j+1}^n (\bar{S}_i \cup \mathfrak{J}_i) \right\},$$

and the union is over all combinations \underline{P} of $j - 1$ integers chosen from the $n - 1$. These $j - 1$ integers index the coordinates for which the projection is $\bar{S} \cup \mathfrak{J}$.

But since we are interested only in the distribution of the $j - 1$ coordinates

“above” and $n - j$ coordinates “below” with respect to the ranking introduced by $\psi(w)$, we introduce the following decomposition of Y :

$$Y = \bigcup_{r=0}^{j-1} \bigcup_{s=0}^{n-j} Y_{rs},$$

where

$$\begin{aligned} Y_{00} &= \bigcup_P \left\{ \prod_{i=2}^j \bar{s}_i \times \prod_{i=j+1}^n \underline{s}_i \right\} = \bigcup_P Y_{00}^1, \\ Y_{rs} &= \bigcup_P \left\{ \left[\bigcup_{P_r^{j-1}} \prod_{i=2}^{r+1} \mathfrak{J}_i \times \prod_{i=r+2}^j \bar{s}_i \right] \times \left[\bigcup_{P_s^{n-j}} \prod_{i=j+1}^{j+s} \mathfrak{J}_i \times \prod_{i=j+s+1}^n \underline{s}_i \right] \right\} \\ &= \bigcup_P Y'_{rs}, \\ Y_{j-1, n-j} &= \bigcup_P \left\{ \prod_{i=2}^n \mathfrak{J}_i \right\} = \bigcup_P Y_{j-1, n-j}^1, \end{aligned}$$

and P_r^{j-1} indicates that the union is taken over all combinations of r integers chosen from $j - 1$ integers (the set corresponding to a particular combination is placed after \bigcup).

Using this decomposition, the distribution \bar{P}^* can be broken into $\binom{n-1}{j-1}$ identical cases as far as the $j - 1$ coordinates “above” and $n - j$ coordinates “below” are concerned. A typical one of these cases is:

$$Y^1 = \bigcup_{r=0}^{j-1} \bigcup_{s=0}^{n-j} Y_{rs}^1,$$

with measure P^{**} defined so that the measure over Y_{00}^1 is the original measure over S^{n-1} (the \bar{P}^* measure), the measure over Y_{rs}^1 is the original measure over S^{n-1} reduced by the factor $1 / (r + s + 1) \binom{r+s}{r}$ (the \bar{P}^* measure reduced by the factor $1 / \binom{r+s}{r}$), the measure over $Y_{j-1, n-j}^1$ is the original measure over S^{n-1} reduced by the factor $1 / n \binom{n-1}{j-1}$ (the \bar{P}^* measure reduced by the factor $1 / \binom{n-1}{j-1}$). This measure when compounded with the other similar permutations is easily seen to reproduce the measure of \bar{P}^* .

Thus the conditional distribution of the sample given that the j th largest value of $\psi(w) = t$ (ties broken by equally likely random choice) is that of $j - 1$ coordinates $\bar{w}_2, \bar{w}_3, \dots, \bar{w}_j$, $n - j$ coordinates $\underline{w}_{j+1}, \dots, \underline{w}_n$, and one coordinate w^* with distribution as follows: $(w^*, \bar{w}_2, \dots, \bar{w}_j, \underline{w}_{j+1}, \dots, \underline{w}_n)$ takes its values in $\mathfrak{J} \times Y^1$ with probability measure obtained by normalizing the product measure of the original W measure truncated to \mathfrak{J} and the measure

P^{**} over Y^1 . The distribution is symmetric with respect to $\bar{w}_2, \dots, \bar{w}_j$ and with respect to $\underline{w}_{j+1}, \dots, \underline{w}_n$.

This conditional distribution is not that of samples of $j - 1$ and $n - j$ from truncated portions of the original distribution except in the particular case in which \mathfrak{J} has measure 0. However, we can express it in terms of samples of $j - 1$ and $n - j$ (with the accompanying simplicity of structure) by the procedure described below.

We replace the original distribution W by the combination (W, U) where U is a uniformly distributed random variable $[0, 1]$ and is independent of W . For any value of t for which $P_w(\psi(W) = t) \neq 0$, record along with the function $\psi(w)$ the observed value of U . For a sample of n from the distribution (W, U) , we order the elements according to $\psi(w)$ and if there are ties we order them according to the corresponding values of u .

We now find the conditional distribution of the sample given that the j th largest value of $(\psi(W), U)$ is (t, u) . The theory at the beginning of this section immediately extends itself even though the combination $(\psi(w), u)$ is not real-valued. Since $P(\psi(W) = t, U = u) = 0$, the distribution of the $n - 1$ elements is that of (1) a sample of $j - 1$ from the distribution having relative measure P_1^c which over $\bar{\mathfrak{s}}$ is the measure of W , and over \mathfrak{J} is the measure of W reduced by the factor $(1 - u)$, and (2) a sample of $n - j$ from the distribution having relative measure P_2^c which over $\underline{\mathfrak{s}}$ is the measure of W , and over \mathfrak{J} is the measure of W reduced by the factor u .

We now show that the previous conditional distribution is obtained from this distribution of samples of $j - 1$ and $n - j$. The conditional distribution given that the j th value has $\psi(w) = t$ is obtained by taking the marginal distribution over U ; it is easily seen to be a distribution over the space Y^1 defined above. We now evaluate the measure for this distribution.

The marginal distribution of the u coordinate for the j th largest value is needed. For continuous distributions the j th largest value has a mapping of the Beta distribution with parameters $n - j + 1$ and j . However, the mapping is linear for the portion in which we are interested and therefore has as density function

$$K \cdot [P(\bar{\mathfrak{s}}) + (1 - u)P(\mathfrak{J})]^{j-1} [P(\underline{\mathfrak{s}}) + uP(\mathfrak{J})]^{n-j}.$$

The relative probability measure over Y^1 is obtained by considering the different subsets Y_{rs}^1 . The measure over Y_{rs}^1 given the u value for the j th largest element is the original product measure over \mathfrak{s}^{n-1} reduced by the factor

$$\frac{(1 - u)^{r-1} u^s}{[P(\bar{\mathfrak{s}}) + (1 - u)P(\mathfrak{J})]^{j-1} [P(\underline{\mathfrak{s}}) + uP(\mathfrak{J})]^{n-j}}.$$

The factor in the denominator obtains from the normalization of the distributions of samples of $j - 1$ and $n - j$ from P_1^c and P_2^c . Taking the marginal distribution with respect to u , the factor becomes

$$\int_0^1 \frac{(1-u)^r u^s K [P(\bar{s}) + (1-u)P(\mathfrak{z})]^{j-1} [P(\underline{s}) + uP(\mathfrak{z})]^{n-j}}{[P(\bar{s}) + (1-u)P(\mathfrak{z})]^{j-1} [P(\underline{s}) + uP(\mathfrak{z})]^{n-j}} du$$

$$= K \frac{\Gamma(r+1)\Gamma(s+1)}{\Gamma(r+s+2)} = \frac{K}{(r+s+1) \binom{r+s}{r}}$$

The constant K is present to normalize the distribution over Y^1 . Thus the model in terms of samples of $j - 1$ and $n - j$ reproduces the conditional distribution, since the marginal distribution over U of the samples of $j - 1$ and $n - j$ reproduces the distribution P^{**} .

5. Definition of the Blocks for the continuous case. In order to form blocks with the generality described in the introduction, we not only need a sequence of functions which describe how the cuts are made but also a sequence of integers to indicate through which point each cut is to be made. The procedure no longer produces $n + 1$ blocks one at a time, but rather n divisions or cuts are made which successively divide the space into sets, each set being equivalent to the union of a number of blocks.

Consider n points w_1, w_2, \dots, w_n in the space \mathfrak{S} . To define the $n + 1$ blocks the following two sequences are needed; the first,

$$\varphi_1(w), \varphi_2(w | \varphi_1), \dots, \varphi_n(w | \varphi_1, \dots, \varphi_{n-1}),$$

is a sequence of real-valued functions over \mathfrak{S} ; the second,

$$p_1, p_2(\varphi_1), p_3(\varphi_1, \varphi_2), \dots, p_n(\varphi_1, \dots, \varphi_{n-1}),$$

is a permutation of the integers $(1, 2, \dots, n)$. In each case the elements of the sequences depend as indicated on values of previous elements in the first sequence. The value of φ to be inserted in the functions at any stage is described in the definition 5.1 below.

For a set of real numbers (x_1, \dots, x_i) we define $\max_j^r x_j$ to be the r th largest value in the set, and $i(\max_j^r x_j)$ to be the integer i for which $x_i = \max_j^r x_j$. If $i(\max_j^r x_j)$ is not uniquely determined, the context in which the symbol is used will indicate which of the available integers is to be chosen.

DEFINITION 5.1. *The set (w_1, w_2, \dots, w_n) of points in \mathfrak{S} and the functions $\{\varphi_i(w | \varphi_1, \dots, \varphi_{i-1})\}$ and $\{p_i(\varphi_1, \dots, \varphi_{i-1})\}$ define a partition of \mathfrak{S} into disjoint blocks S_1, \dots, S_{n+1} and cuts T_1, \dots, T_n as follows:*

(i) *For the first stage,*

$$T_{p_1} = \{w | \varphi_1(w) = \varphi_1^{p_1}\},$$

$$\bigcup_{p \leq p_1} S_p \cup \bigcup_{p < p_1} T_p = \{w | \varphi_1(w) > \varphi_1^{p_1}\},$$

$$\bigcup_{p > p_1} S_p \cup \bigcup_{p > p_1} T_p = \{w | \varphi_1(w) < \varphi_1^{p_1}\},$$

where $\varphi_1^{p_1} = \max_{j=1, \dots, n}^{p_1} \varphi_1(w_j)$, and $i(\max^{p_1} \varphi_1(w_j)) = i(\varphi_1^{p_1})$ is any one of the integers available. $\beta_1(S; p)$, $\beta_1(T; p)$ are used to designate for the S 's, T 's respectively the index sets which contain p and over which the set union operator is applied. Corresponding to each $\beta_1(T; p)$ there is a set of integers $\gamma_1(p)$ which index the points w_j to be used for the cuts indicated by $\beta_1(T; p)$; if $p > p_1$, $< p_1$ then $\gamma_1(p)$ consists of $n - p_1$, $p_1 - 1$ integers j for which $\varphi_1(w_j) \leq \varphi_1^{p_1}$, $\geq \varphi_1^{p_1}$.

(ii) For the second stage,

$$T_{p_2(\varphi_1^{p_1})} = \{w \mid \varphi_2(w \mid \varphi_1^{p_1}) = \varphi_2^{p_2^*}; \Xi(w)\}$$

$$\bigcup_{\substack{p \leq p_2(\varphi_1^{p_1}) \\ p \in \beta_1(S; p_2(\varphi_1^{p_1}))}} S_p \cup \bigcup_{\substack{p < p_2(\varphi_1^{p_1}) \\ p \in \beta_1(T; p_2(\varphi_1^{p_1}))}} T_p = \{w \mid \varphi_2(w \mid \varphi_1^{p_1}) > \varphi_2^{p_2^*}; \Xi(w)\}$$

$$\bigcup_{\substack{p > p_2(\varphi_1^{p_1}) \\ p \in \beta_1(S; p_2(\varphi_1^{p_1}))}} S_p \cup \bigcup_{\substack{p > p_2(\varphi_1^{p_1}) \\ p \in \beta_1(T; p_2(\varphi_1^{p_1}))}} T_p = \{w \mid \varphi_2(w \mid \varphi_1^{p_1}) < \varphi_2^{p_2^*}; \Xi(w)\}$$

where $\varphi_2^{p_2^*} = \max_{j \in \gamma_1(p_2)}^{p_2^*} \varphi_2(w_j \mid \varphi_1^{p_1})$, $p_2^* = p_2(\varphi_1^{p_1}) - \min \beta_1(T; p_2(\varphi_1^{p_1})) + 1$, and $i(\varphi_2^{p_2^*})$ is any one of the integers available. $\beta_2(S; p)$, $\beta_2(T; p)$ are used to designate for the S 's, T 's respectively the smallest index sets, containing p , over which the set union operator is applied. $\Xi(w)$ in each case stands for the condition that the points of the sets being defined at that stage conform to restrictions imposed on the points of those same sets at the earlier stages. Thus in effect $\Xi(w)$ stands for all the inequalities at the previous stages which apply to points of the sets under consideration.

(iii) For the r th stage,

$$T_{p_r(\varphi_1^{p_1}, \dots, \varphi_{r-1}^{p_{r-1}})} = \{w \mid \varphi_r(w \mid \varphi_1^{p_1}, \dots, \varphi_{r-1}^{p_{r-1}}) = \varphi_r^{p_r^*}; \Xi(w)\}$$

$$\bigcup_{\substack{p \leq p_r(\varphi_1^{p_1}, \dots, \varphi_{r-1}^{p_{r-1}}) \\ p \in \beta_{r-1}(S; p_r(\varphi_1^{p_1}, \dots, \varphi_{r-1}^{p_{r-1}}))}} S_p \cup \bigcup_{\substack{p < p_r(\varphi_1^{p_1}, \dots, \varphi_{r-1}^{p_{r-1}}) \\ p \in \beta_{r-1}(T; p_r(\varphi_1^{p_1}, \dots, \varphi_{r-1}^{p_{r-1}}))}} T_p$$

$$= \{w \mid \varphi_r(w \mid \varphi_1^{p_1}, \dots, \varphi_{r-1}^{p_{r-1}}) > \varphi_r^{p_r^*}; \Xi(w)\},$$

$$\bigcup_{\substack{p > p_r(\varphi_1^{p_1}, \dots, \varphi_{r-1}^{p_{r-1}}) \\ p \in \beta_{r-1}(S; p_r(\varphi_1^{p_1}, \dots, \varphi_{r-1}^{p_{r-1}}))}} S_p \cup \bigcup_{\substack{p > p_r(\varphi_1^{p_1}, \dots, \varphi_{r-1}^{p_{r-1}}) \\ p \in \beta_{r-1}(T; p_r(\varphi_1^{p_1}, \dots, \varphi_{r-1}^{p_{r-1}}))}} T_p$$

$$= \{w \mid \varphi_r(w \mid \varphi_1^{p_1}, \dots, \varphi_{r-1}^{p_{r-1}}) < \varphi_r^{p_r^*}; \Xi(w)\},$$

where $\varphi_r^{p_r^*} = \max_{j \in \gamma_{r-1}(p_r)}^{p_r^*} \varphi_r(w_j \mid \varphi_1^{p_1}, \dots, \varphi_{r-1}^{p_{r-1}})$ and $p_r^* = p_r(\varphi_1^{p_1}, \dots, \varphi_{r-1}^{p_{r-1}}) - \min \beta_{r-1}(T; p_r(\varphi_1^{p_1}, \dots, \varphi_{r-1}^{p_{r-1}})) + 1$. $\beta_r(S; p)$, $\beta_r(T; p)$ are the smallest index sets for the S 's, T 's respectively which contain the integer p ; the index sets considered are those over which the indices of the unions in stages 1 to r range. Corresponding to each $\beta_r(T; p)$ is a set of integers $\gamma_r(p)$ which indexes the points w_j to be used for the cuts indicated by $\beta_r(T; p)$; $\gamma_r(p)$ is identical to $\gamma_{r-1}(p)$ unless $\beta_{r-1}(T; p)$ contains $p_r(\varphi_1^{p_1}, \dots, \varphi_{r-1}^{p_{r-1}})$; in which case $\beta_{r-1}(T; p)$ is partitioned into $(p_r(\varphi_1^{p_1}, \dots, \varphi_{r-1}^{p_{r-1}}))$ and two $\beta_r(T; p)$'s, and correspondingly $\gamma_{r-1}(p)$ is partitioned into $i(\varphi_r^{p_r^*})$ and two $\gamma_r(p)$'s according as $\varphi_r(w_j) = \varphi_r^{p_r^*}$, $\leq \varphi_r^{p_r^*}$, $\geq \varphi_r^{p_r^*}$. $\Xi(w)$ is as defined in stage (ii).

The procedure for the r th stage is applied for $r = 2, \dots, n$.

6. General results for the continuous case.

THEOREM 6.1. *If $\varphi_1(w | \varphi_1^{p_1}, \dots, \varphi_{i-1}^{p_{i-1}})$ has a continuous conditional distribution for all values of $\varphi_1^{p_1}, \dots, \varphi_{i-1}^{p_{i-1}}$ (except for a set of probability measure (W) zero and for all i , and if for a sample of n (w_1, \dots, w_n) from the distribution W , blocks S_1, \dots, S_{n+1} are defined according to Definition 5.1, then*

- (i) *The blocks are random,*
 (ii) *The coverage (c_1, \dots, c_{n+1}), where $c_i = P_W(S_i)$, are random and have a uniform distribution over the region in R^{n+1} defined by the linear conditions*

$$\sum_{i=1}^{n+1} c_i = 1, \quad c_i \geq 0 \quad (i = 1, \dots, n+1).$$

An example of the uniform distribution referred to in the theorem is the distribution of $U_1, U_2 - U_1, \dots, U_n - U_{n-1}, 1 - U_n$, where U_1, \dots, U_n are the order statistics of a sample of n from the uniform distribution $[0, 1]$.

The proof outlined below does not presuppose the results obtained in the earlier papers [1] to [7] in the References, but rather it follows directly from the results of Section 5 and a simple probability mapping.

PROOF. For a sample of n from the uniform distribution $[0, 1]$ we define coverages c'_1, \dots, c'_{n+1} as follows.

$$\begin{aligned} c'_1 &= U_1, \\ c'_2 &= U_2 - U_1, \\ &\vdots \\ c'_{n+1} - 1 &= U_n, \end{aligned}$$

where U_1, \dots, U_n are the order statistics of the sample. It is well known that these coverages have the distribution described in the theorem.

We consider the density of this distribution as a product of conditional and marginal densities. It can be written as the marginal distribution of $\sum_{i=1}^{p_1} c'_i = C_{p_1}$ which is a Beta distribution, multiplied by the conditional distribution of c'_1, \dots, c'_p , and the conditional distribution of $c'_{p_1+1}, \dots, c'_{n+1}$. These two conditional distributions are independent. The first is the distribution of coverages (each reduced by the factor C_{p_1}) obtained from a sample of $p_1 - 1$ from the uniform distribution $[0, 1]$. The second is that of coverages (reduced by $1 - C_{p_1}$) obtained from a sample of $n - p_1$ from the uniform distribution $[0, 1]$. Similarly each of these conditional coverage distributions can be written as a product of a marginal and two conditionals, and so on. These results are immediate consequences of the geometry of the region in R^{n+1} .

Consider now a sample of n from the uniform distribution over $[0, 1]^n$, the unit cube in R^n . Order the sample with respect to the first coordinate u_1 from the smallest to largest and pick the p_1 th point. The marginal distribution of the first coordinate of this point is that of C_{p_1} in the previous situation. It is easily seen that the conditional distribution of the remainder of the sample is that of a sample of $p_1 - 1$ from the uniform distribution over $[0, 1]^n$, the first coordinate

(and any derived coverages) being reduced by factor C_{p_1} , and of a sample of $n - p_1$ from the uniform distribution over $[0, 1]^n$ the first coordinate being reduced by the factor $1 - C_{p_1}$. Similarly each of these conditional distributions can be broken into a marginal and two conditionals and so on. The important thing is that using the conditional approach the coverage distributions are easily seen to be the same in the two somewhat different situations.

The coverages referred to in the theorem can be obtained by a mapping from the coverages in the latter case above, and since those coverages have been shown to have the distribution described in the theorem, the proof follows. The mapping referred to is the following.

$\varphi_1(W)$ has a continuous distribution. Consider a monotone nonincreasing function $g_1(u_1)$ such that $g_1(U)$ has the same distribution as $\varphi_1(W)$ when U has the uniform distribution $[0, 1]$. Apply this mapping to the second situation above; the region having $u_1 < C_{p_1}$ maps into the set of points of $\varphi_1(w)$ which corresponds to $U_1^{p_1}S_i$ (the T_i have probability measure zero). Thus the marginal distribution of C_{p_1} is identical to that of $\sum_1^{p_1}c_i$. The conditional distribution of the remainder of the sample given that $\sum_1^{p_1}C_i$ is fixed ($\max_1^{p_1}\varphi_1(w_i)$ fixed) is by Section 4 that of samples of $p_1 - 1, n - p_1$ from the original distribution restricted to points for which $\varphi_1(w) > \max_1^{p_1}\varphi_1(w_i), \varphi_1(w) < \max_1^{p_1}\varphi_1(w_i)$. Since we are left with the probability distribution of samples of $p_1 - 1, n - p_1$ for which any derived coverages are reduced by the factor $C_{p_1}, 1 - C_{p_1}$, we can split up these conditional distributions just as we did the original distribution.

At each stage in the formation of the blocks S_i , we define a mapping from a coordinate (taken in the order u_1, \dots, u_n) of $[0, 1]^n$ to the range of the φ being considered. The mapping is chosen to reproduce the required conditional distribution of that φ . Thus the splitting up by successive cuts of the space S corresponds to splitting by cuts in $[0, 1]^n$. The successive cuts are made parallel to the coordinate planes $u_1 = 0, u_2 = 0, \dots, u_n = 0$. Thus the distribution of the c 's is reproduced in the c 's by using conditional distributions successively, corresponding to the steps in which the cuts were made. This completes the proof.

7. Main results for the discontinuous case. The results for the discontinuous case correspond very closely to those given in [7]. However, for the proof of the main theorem the results of Section 4 in the present paper are needed and can no longer be assumed to be sufficiently obvious. The mappings from the uniform (continuous) distribution to the discontinuous distributions supplies the necessary randomization at cuts having finite probability. This permits the conditional distributions to be used in the form of samples from a truncated distribution. The proof as a whole follows the pattern set in Section 8 of [7] with the sort of modification indicated by the continuous case proof in Section 6 of this paper.

The definition of the m -system of functions carries over. The dependence of any Φ on other Φ 's with smaller subscripts is through the values of those Φ 's at their respective cuts.

The definition of the blocks, cuts and coverages is as given in Section 5 of this paper with the following modifications.

- (i) The φ 's are replaced by Φ 's.
- (ii) When more than one point falls on a cut the point should be chosen at random, perhaps most conveniently in such a manner that each point has the same probability of being selected.
- (iii) The closed blocks $\bar{S}_1, \dots, \bar{S}_{n+1}$ are defined by the expressions for S_1, \dots, S_{n+1} where throughout $<$ is replaced by \leq and $>$ by \geq .
- (iv) Definitions (7.4) and (7.5) in [7] are used to define block groups and coverages.

The main theorem for the discontinuous case reads as Theorem 8.1 in [7] with the obvious modifications as indicated by the theorem for the continuous case in the present paper.

REFERENCES

- [1] S. S. WILKS, "Determination of sample sizes for setting tolerance limits," *Ann. Math. Stat.*, Vol. 12 (1941), pp. 91-96.
- [2] A. WALD, "An extension of Wilks' method for setting tolerance limits," *Ann. Math. Stat.*, Vol. 14 (1943), pp. 45-55.
- [3] H. SCHEFFÉ AND J. W. TUKEY, "Non-parametric estimation. I. Validation of order statistics," *Ann. Math. Stat.*, Vol. 16 (1945), pp. 187-192.
- [4] J. W. TUKEY, "Nonparametric estimation, II. Statistically equivalent blocks and tolerance regions—the continuous case," *Ann. 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," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 30-39.
- [6] D. A. S. FRASER AND R. WORMLEIGHTON, "Nonparametric estimation, IV," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 294-298.
- [7] D. A. S. FRASER, "Sequentially determined statistically equivalent blocks," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 372-381.
- [8] R. B. MURPHY, "Nonparametric tolerance limits," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 581-589.