

ORDER STATISTICS

S. S. WILKS

CONTENTS

1. Introduction	7
2. Notation and preliminary definitions	10
3. Sampling distributions of coverages for the case of one dimension	13
4. Examples of direct applications of coverage distributions for the case of one dimension	14
(a) Confidence limits of medians, quartiles and other quantiles	14
(b) Population tolerance limits	15
5. Distribution of single order statistics	16
(a) Exact distribution of $X_{(r)}$	16
(b) Limiting distribution of largest (or smallest) value in a sample	17
(c) Limiting distribution of $X_{(r)}$	20
6. Joint distributions of several one-dimensional order statistics and applications	20
(a) Distribution of sample range	21
(b) Limiting distribution of two or more order statistics in large samples	22
(c) Estimation of population parameters by order statistics	22
(d) Application of order statistics to breaking strength of thread bundles	23
7. Confidence bands for the cdf $F(x)$	24
8. Sampling distributions of coverages for the case of two or more dimensions	25
(a) For regions determined by one-dimensional order statistics	26
(b) For rectangular regions in case of independent random variables	26
(c) Wald's results for rectangular regions in case the random variables are not independent	28
(d) Tukey's generalization of Wald's results	29
(e) Extensions of distribution theory of coverages and order statistics to the discrete case	32
9. Application of multi-dimensional coverages and order statistics to estimation problems	32
10. Order statistics in the testing of statistical hypotheses—the method of randomization	34
11. Examples of nonparametric statistical tests for one dimension based on the method of randomization	36
(a) Two-sample tests	36
(b) Tests of independence or "randomness" in ordered sequences; run tests	39
(c) One-dimensional parametric tests involving order statistics	42
(d) Analysis of variance tests by the method of randomization	42
12. Nonparametric tests for two or more dimensions based on the method of randomization	45
(a) Tests of independence based on correlation coefficients	45
(b) A "corner" test of association	46
References and literature	47

An address delivered before the Summer meeting of the Society in New Haven on September 4, 1947, by invitation of the Committee to Select Hour Speakers for Annual and Summer Meetings; received by the editors October 21, 1947.

1. Introduction. Within the past twenty-five years, a large body of statistical inference theory has been developed for samples from populations having normal, binomial, Poisson, multinomial and other specified forms of distribution functions depending on one or more unknown population parameters. These developments fall into two main categories: (i) statistical estimation, and (ii) the testing of statistical hypotheses. The theory of statistical estimation deals with the problem of estimating values of the unknown parameters of distribution functions of specified form from random samples assumed to have been drawn from such populations. The testing of statistical hypotheses deals with the problem of testing, on the basis of a random sample, whether a population parameter has a specified value, or more generally whether one or more specified functional relationships exist among two or more population parameters.

All of this theory has now been placed on a foundation of probability theory through the work of R. A. Fisher, J. Neyman, E. S. Pearson, A. Wald, and others. It has been applied to most of the common distribution functions which occur in statistical practice. Many statistical tables have been prepared to facilitate application of the theory.

There are many problems of statistical inference in which one is unable to assume the functional form of the population distribution. Many of these problems are such that the strongest assumption which can be reasonably made is continuity of the cumulative distribution function of the population. An increasing amount of attention is being devoted to statistical tests which hold for all populations having continuous cumulative distribution functions. Problems of this type in which the distribution function is arbitrary within a broad class are referred to as *non-parametric* problems of statistical inference. An excellent expository account of the theory of non-parametric statistical inference has been given by Scheffé [60].¹ In non-parametric problems it is being found that *order statistics*, that is, the ordered set of values in a random sample from least to greatest, are playing a fundamental rôle. It is to be particularly noted that the term *order* is not being used here in the sense of arrangement of sample values in a sequence as they are actually drawn.

There are both theoretical and practical reasons for this increased attention to nonparametric problems and order statistics. From a theoretical point of view it is obviously desirable to develop methods of statistical inference which are valid with respect to broad classes

¹ Numbers in brackets refer to the references cited at the end of the paper.

of population distribution functions. It will be seen that a considerable amount of new statistical inference theory can be established from order statistics assuming nothing stronger than continuity of the cumulative distribution function of the population. Further important large sample results can be obtained by assuming continuity of the derivative of the cumulative distribution function. From a practical point of view it is desirable to make the statistical procedures themselves as simple and as broadly applicable as possible. This is indeed the case with statistical inference theory based on order statistics. Order statistics also permit very simple "inefficient" solutions of some of the more important *parametric* problems of statistical estimation and testing of hypotheses. The solutions are inefficient in the sense that they do not utilize all of the information contained in the sample as it would be utilized by "best possible" (and computationally more complicated) methods. But this inefficiency can be offset in many practical situations where the size of the sample can be increased by a trivial amount of effort and expense.

It is the purpose of this paper to present some of the more important results in the sampling theory of order statistics and of functions of order statistics and their applications to statistical inference together with some reference to important unsolved problems at appropriate places in the paper. The results will be given without proofs, since these may be found in references cited throughout the paper. Before proceeding to the technical discussion it may be of interest to make a few historical remarks about order statistics.

One of the earliest problems on the sampling theory of order statistics was the *Galton difference problem* studied in 1902 by Karl Pearson [51]. The mathematical problem here, which was solved by Pearson, was to find the mean value of the difference between the r th and $(r+1)$ th order statistic in a sample of n values from a population having a continuous probability density function. In 1925 Tippett [72] extended the work of Pearson to find the mean value of the sample *range* (that is, the difference between least and greatest order statistics in a sample). In the same paper Tippett tabulated, for certain sample sizes ranging from 3 to 1000, the cumulative distribution function of the largest order statistic in a sample from a normal population having zero mean and unit variance. In 1928 Fisher and Tippett [14] determined by a method of functional equations and for certain restricted regularity conditions on the population distribution the limiting distribution of the greatest (and also least) value in a sample as the sample size increases indefinitely. R. von Mises [35] made a precise determination of these regularity conditions. Gumbel

has made further studies of these limiting distributions and has made various applications to such problems as flood flows [19] and maximum time intervals between successive emissions of gamma rays from a given source [18]. In 1932 A. T. Craig [4] gave general expressions for the exact distribution functions of the median, quartiles, and range of a sample size of n . Daniels [7] has recently made an interesting application of the sampling theory of order statistics to develop the probability theory of the breaking strength of bundles of threads.

One of the simplest and most important functions of order statistics is the sample cumulative distribution function $F_n(x)$, the fraction of values in a sample of n values not exceeding x . In 1933, Kolmogoroff [30] established a fundamental limit theorem in probability theory which enables one to set up from $F_n(x)$ a confidence band (for large n) for an unknown continuous cumulative distribution function $F(x)$ of the population from which the sample is assumed to have been drawn. Smirnov [66] extended Kolmogoroff's results to a treatment of the probability theory of the difference between two sample cumulative distribution functions $F_{n_1}(x)$ and $F_{n_2}(x)$ for large samples. Wald and Wolfowitz [76] have developed a method for determining exact confidence limits for $F(x)$ from a sample of n values.

In 1936 Thompson [70] showed how confidence limits for the median and other quantities of a population having a continuous cumulative distribution function could be established from order statistics in a sample from such a population. His result was discovered independently by Savur [59] in 1937. In 1941 the author [85] showed how the probability function of the portion of the distribution (continuous) of a population lying between two order statistics could be used to set up *tolerance limits* for the population from which the sample is assumed to have been drawn. These ideas were extended by Wald [75] to the determination of rectangular tolerance regions for populations having distribution functions of several variables. More recently Tukey [73] has extended Wald's ideas to the determination of more general tolerance regions. Tukey's extensions give promise of a variety of applications in statistical inference.

One of the most important properties of the probability distribution for a sample of n values from a population having a continuous cumulative distribution function is that the probabilities associated with the $n!$ different permutations of sample values are equal. Fisher [12, 13] initiated the idea of utilizing this property to develop the *randomization method* for constructing statistical tests and illustrated

his ideas in several examples. Friedman [16], Hotelling and Pabst [26], Pitman [55, 56, 57], Olmstead and Tukey [45] and Welch [83, 84] have used the randomization method and its extensions to several samples for developing various statistical tests valid for populations with continuous cumulative distribution functions. Wald and Wolfowitz [77] used the idea to develop a test of the statistical hypothesis that two samples have come from populations having identical continuous cumulative distribution functions. Wolfowitz [88] has proposed an extension of the Neyman-Pearson likelihood ratio method (a standard method for determining test criteria in parametric problems) for systematically determining test criteria in non-parametric problems, making use of the randomization principle.

2. Notation and preliminary definitions. Throughout this paper we shall consider only populations with continuous cumulative distribution functions. At certain points in the paper, which will be indicated, we shall consider cumulative distribution functions (cdf) with derivatives which are continuous except possibly for a set of points of measure zero; we shall refer to such derivatives as continuous with this understanding.

A one-dimensional continuous cdf $F(x)$ satisfies the following conditions:

- (a) $F(x)$ is continuous on the entire one-dimensional x space.
- (b) $F(+\infty) = 1$.
- (c) $F(-\infty) = 0$.
- (d) $F(x)$ is nondecreasing.

If X is such that the probability that $X \leq x$ is $F(x)$, or briefly if

$$\Pr(X \leq x) = F(x),$$

then we say that X is a *random variable* which has the cdf $F(x)$. If $F(x)$ has a continuous derivative $f(x)$, then $f(x)dx$ is called the *probability element* of X , and $f(x)$ the *probability density function* (pdf) of X .

Since we are considering continuous cdf's it should be noted that $\Pr(X \leq x) = \Pr(X < x) = F(x)$.

Similarly, a k -dimensional continuous cdf $F(x_1, x_2, \dots, x_k)$ satisfies the following conditions:

- (a) $F(x_1, x_2, \dots, x_k)$ is continuous over the entire k -dimensional (Euclidean) space R_k .
- (b) $F(+\infty, +\infty, \dots, +\infty) = 1$.
- (c) $F(-\infty, x_2, \dots, x_k) = F(x_1, -\infty, x_3, \dots, x_k) = \dots = F(x_1, x_2, \dots, x_{k-1}, -\infty) = 0$.

(d) $\Delta_{x_1} [\Delta_{x_2} \cdots [\Delta_{x_k} F(x_1, x_2, \cdots, x_k)] \cdots] \geq 0$ where $\Delta_{x_i} G(x_1, x_2, \cdots, x_k) = G(x_1, x_2, \cdots, x_{i-1}, x_i + h_i, x_{i+1}, \cdots, x_k) - G(x_1, x_2, \cdots, x_k)$, and $h_i > 0$. If X_1, X_2, \cdots, X_k are such that the probability that all the inequalities $X_1 \leq x_1, X_2 \leq x_2, \cdots, X_k \leq x_k$ hold is $F(x_1, x_2, \cdots, x_k)$, or more briefly if

$$\Pr (X_1 \leq x_1, X_2 \leq x_2, \cdots, X_k \leq x_k) = F(x_1, x_2, \cdots, x_k),$$

then we say that X_1, X_2, \cdots, X_k are *random variables* having cdf $F(x_1, x_2, \cdots, x_k)$. If $\partial^k F(x_1, x_2, \cdots, x_k) / \partial x_1 \partial x_2 \cdots \partial x_k = f(x_1, x_2, \cdots, x_k)$, say, is continuous, $f(x_1, x_2, \cdots, x_k) dx_1 dx_2 \cdots dx_k$ is called the *probability element* of the k random variables, and $f(x_1, x_2, \cdots, x_k)$ is the *probability density function* (pdf) of the random variables.

In general, we shall denote random variables by capital letters and the variables in the cdf and pdf by the corresponding lower case letters.

If X is a random variable having cdf $F(x)$, and if $G(x)$ is a Borel-measurable function, then $G(X)$ is a random variable having a cdf $H(y)$ defined by the Lebesgue-Stieltjes integral

$$H(y) = \Pr (G(x) \leq y) = \int_S dF(x)$$

where S is the set of points for which $G(x) \leq y$. The *mean value* of $G(X)$ is defined by

$$(1) \quad E(G(X)) = \int_{-\infty}^{\infty} G(x) dF(x).$$

Similarly, in the case of k variables having cdf $F(x_1, x_2, \cdots, x_k)$ the cdf $H(y_1, y_2, \cdots, y_r)$ of r functionally independent Borel-measurable functions of the X 's, say $\phi_i(X_1, X_2, \cdots, X_r)$, $i = 1, 2, \cdots, r$ ($r \leq k$), is defined by the Lebesgue-Stieltjes integral

$$(2) \quad H(y_1, y_2, \cdots, y_r) = \int_S dF(x_1, x_2, \cdots, x_k)$$

where S is the set of points in R_k for which

$$\phi_i(x_1, x_2, \cdots, x_k) \leq y_i \quad (i = 1, 2, \cdots, r).$$

The mean value of $\phi_i(X_1, X_2, \cdots, X_k)$ is defined by

$$(3) \quad E(\phi_i) = \int_{R_k} \phi_i(x_1, x_2, \cdots, x_k) dF.$$

If $F(x)$ is a cdf and if X_1, X_2, \dots, X_n are random variables having cdf

$$(4) \quad F(x_1) \cdot F(x_2) \cdot \dots \cdot F(x_n),$$

then X_1, X_2, \dots, X_n is said to be a *random sample* O_n of size n from a *population* having cdf $F(x)$. The n -dimensional Euclidean space R_n of the x 's in the case of random sampling is called *sample space*. All sets of points and all functions considered in this paper are assumed to be Borel-measurable.

It should be noted that R_n is constructed as a product space from n one-dimensional spaces and has a cdf which is the product of the corresponding n one-dimensional (and identical) cdf's.

Similarly, one defines a random sample of size n from a population having a k -dimensional cdf. In this case the sample space would be kn -dimensional.

Suppose X_1, X_2, \dots, X_n is a sample of size n from a population having continuous cdf $F(x)$. It can be shown that the probability of two or more of the X 's being equal is zero. Hence we shall always consider the X 's in a sample as all having different values. Let the X 's be arranged in increasing order of magnitude from least to greatest and denoted by $X_{(1)} < X_{(2)} < \dots < X_{(n)}$. These ordered values of the X 's are called *order statistics*, $X_{(r)}$ being the r th order statistic, $r = 1, 2, \dots, n$.

If I is any interval or set determined from the X 's of a sample, the integral $C = \int_I dF(x)$ will be called the *population coverage* of I , or simply the *coverage* of I . For example, $F(X_{(j)}) - F(X_{(i)})$, for $i < j$, is the coverage of the interval $(X_{(i)}, X_{(j)})$. (If this paper were written completely abstractly, that is, devoid of statistical interpretation, a better term for C would be *measure of I*.) The coverage of I is the fraction of the population or amount of probability in the population distribution function contained in I . Since I is a function of the sample X 's, the coverage of I is a random variable. In the case of a population having a continuous k -dimensional cdf $F(X_1, X_2, \dots, X_k)$ a similar definition holds for a coverage C_k over a k -dimensional region or set S_k in the variate space determined by the sample.

We shall find it convenient to introduce the following definition at this stage:

DEFINITION. *Suppose a sample of size n is drawn from a k -variate population having a continuous cdf $F(x_1, x_2, \dots, x_k)$, and let S_k be a region (or set) in the x_1, x_2, \dots, x_k space having a boundary determined by the sample. If C_k , the coverage of S_k , has a distribution which does not depend on $F(x_1, x_2, \dots, x_k)$ we shall say that C_k is *distribution-free*.*

If the cdf has a specified functional form but depends on certain parameters $\theta_1, \theta_2, \dots, \theta_s$ and if S_k can be determined by the sample so that the coverage C_k has a distribution which does not depend on the θ 's, we shall say that C_k is *parameter-free*. Such coverages have been determined by the author [85], and by Paulson [46], for the case of populations having a normal distribution. However, we shall not be interested in parameter-free coverages in this paper since interesting parameter-free coverages do not, in general, depend on order statistics for their determination. It is obvious, of course, that if a coverage is distribution-free it is also parameter-free.

3. Sampling distributions of coverages for the case of one dimension. One of the most important, useful, and well known properties of a continuous cdf $F(x)$ may be stated as follows:

If X is a random variable having a continuous cdf $F(x)$ then $F(X)$ is a random variable such that

$$\Pr(F(X) \leq p) = p, \quad 0 \leq p \leq 1.$$

Thus, if X_1, X_2, \dots, X_n is a sample of size n from a population having a continuous cdf, the probability element of $P_i = F(X_i)$, $i = 1, 2, \dots, n$, is

$$(5) \quad dp_1 dp_2 \cdots dp_n,$$

the sample space of the P_i being the n -dimensional cube $0 \leq p_i \leq 1$ ($i = 1, 2, \dots, n$). Hence, it is seen that the random variables P_i are independently and uniformly distributed on the interval $(0, 1)$.

Now consider the order statistics $X_{(i)}$, $i = 1, 2, \dots, n$, in a sample of size n from a population having a continuous cdf $F(x)$. Represent these order statistics as points on the x -axis, where x is the variate in the cdf $F(x)$. These points cut the x -axis into $n+1$ intervals $I_i = [X_{(i)}, X_{(i-1)}]$, $i = 1, 2, \dots, n+1$, where $X_{(0)} = -\infty$, and $X_{(n+1)} = +\infty$. Let C_i be the coverage of $I_{(i)}$. These coverages are all random variables whose sum is unity. The probability element of any n of these coverages, for convenience the first n , is

$$(6) \quad n! dc_1 dc_2 \cdots dc_n,$$

where $c_i \geq 0$ and $\sum_{i=1}^n c_i \leq 1$. In other words, the pdf of the C_i is uniform over the n -dimensional simplex in the space of the c_i with corners at the $n+1$ points $(0, 0, \dots, 0)$, $(1, 0, 0, \dots, 0)$, \dots , $(0, 0, \dots, 0, 1)$. If a set of $n+1$ coverages C_1, C_2, \dots, C_{n+1} for which $\sum_{i=1}^{n+1} C_i = 1$ have probability element (6), we shall say that this is a set of *elementary coverages*. The most important property of this distribution as far as statistical inference is concerned is that it *does not depend on $F(x)$* , which is usually unknown in statistical infer-

ence problems. In other words, *the coverages for the intervals I_i are distribution-free.*

The probability element given by (6), although very simple, is basic in the sampling theory of coverages and order statistics.

If we let U be the coverage composed of the sum of any r ($r \leq n$) of the $n+1$ elementary coverages C_i , we can verify from (6) that U has the probability element

$$(7) \quad \frac{\Gamma(n+1)}{\Gamma(r)\Gamma(n-r+1)} u^{r-1}(1-u)^{n-r} du, \quad 0 \leq u \leq 1.$$

If U is the sum of r elementary coverages and V is the sum of s elementary coverages ($r+s \leq n$) so that U and V contain no elementary coverages in common, then the probability element of U and V is

$$(8) \quad \frac{\Gamma(n+1)}{\Gamma(r)\Gamma(s)\Gamma(n-r-s+1)} u^{r-1}v^{s-1}(1-u-v)^{n-r-s} du dv,$$

where $u \geq 0$, $v \geq 0$, and $u+v \leq 1$.

Similarly, one could set up the probability element of more than two disjoint sums of elementary coverages according to any partition scheme.

At this point it may be of interest to examine a few applications of the coverage distributions given by (7) and (8) and their immediate extensions.

4. Examples of direct applications of coverage distributions for the case of one dimension.

(a) *Confidence limits of medians, quartiles and other quantiles.* Suppose ξ_p is the p -quantile, that is, the value of x for which $F(\xi_p) = p$. If $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ are the order statistics in a sample of size n from a population with continuous cdf $F(x)$, we can construct confidence limits for ξ_p from any two order statistics, $X_{(i)}$ and $X_{(j)}$, $i < j$. For we have

$$(9) \quad \begin{aligned} \Pr(X_{(i)} < \xi_p < X_{(j)}) &= \Pr[F(X_{(i)}) < p < F(X_{(j)})] \\ &= \Pr\{F(X_{(i)}) < p < F(X_{(i)}) + [F(X_{(j)}) - F(X_{(i)})]\} \\ &= \Pr(U < p < U + V), \end{aligned}$$

where U and V are distributed according to (8) with $r=i$ and $s=j-i$. Evaluating this probability, one finds

$$(10) \quad \Pr(X_{(i)} < \xi_p < X_{(j)}) = I_p(i, n-i+1) - I_p(j, n-j+1)$$

where $I_x(p, q)$ is the incomplete beta function defined by

$$(11) \quad I_x(p, q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \int_0^x t^{p-1}(1-t)^{q-1} dt.$$

Thus, the probability can be calculated from the incomplete beta function tables [54] that the *confidence interval* $(X_{(i)}, X_{(j)})$ covers the unknown p -quantile ξ_p . In setting up confidence intervals for practical purposes one would choose a *confidence coefficient* α (say $\alpha = .95$) and then for each value of n select the pair of values of i and j closest together (or one of these pairs in accordance with practical considerations if there are more than one pair), such that the expression on the right of (10) is not less than α .

If $p = 0.5$, $\xi_{0.5}$ is the *median* of the cdf $F(x)$. In this case one would choose $i = k$ and $j = n - k + 1$, $k < (n+1)/2$ (that is, the k th largest and the k th smallest order statistics) and then select a value of k for each n in accordance with the method described above. Expression (10) becomes, in this case,

$$(12) \quad \Pr(X_{(k)} < \xi_{0.5} < X_{(n-k+1)}) = 1 - 2I_{0.5}(k, n - k + 1).$$

Similarly, one could deal with the *lower quartile* $\xi_{0.25}$ and *upper quartile* $\xi_{0.75}$ or any other quantile. Expression (12) was established in 1936 by Thompson [70], and discovered independently by Savur [59]. Nair [42] has tabulated the largest value of k for which $\Pr(X_{(k)} < \xi_{0.5} < X_{(n-k+1)}) \geq 0.95$ and also ≥ 0.99 for $n = 6, 7, 8, \dots, 81$.

(b) *Population tolerance limits*. The following problem, pointed out by Shewhart [63], arose in the statistical aspects of a mass production problem. Suppose a mass production operation is yielding a sequence of product pieces such that measurements on a given dimension are behaving so they can be idealized as values of a random variable from a population having some (unknown) continuous cdf $F(x)$. Can one find a sample size n and two *tolerance limits* $L_1(X_1, X_2, \dots, X_n)$, $L_2(X_1, X_2, \dots, X_n)$ from the sample values X_1, X_2, \dots, X_n so that no matter what continuous cdf $F(x)$ is, the probability is α that the fraction of the population covered by the interval (L_1, L_2) is at least β , that is,

$$(13) \quad \Pr[(F(L_2) - F(L_1)) > \beta] = \alpha?$$

That is, what functions L_1 and L_2 will produce an interval (L_1, L_2) such that the population coverage for (L_1, L_2) is *distribution-free*? Values of α and β of practical interest are high (that is, 0.90 to 0.99). It was shown by the author [86] that this problem could be solved by choosing L_1 and L_2 as order statistics. For example, if $L_1 = X_{(k)}$ and $L_2 = X_{(n-k+1)}$ (in practice k would be taken as 1, 2, 3 or some small

value compared with n) then it follows immediately from (7) that $F(X_{(n-k+1)}) - F(X_{(k)}) = U$ has the probability element

$$(14) \quad \frac{\Gamma(n+1)}{\Gamma(2k)\Gamma(n-2k+1)} u^{n-2k}(1-u)^{2k-1} du.$$

In this case the problem reduces to the determination of the smallest value of n for which

$$(15) \quad 1 - I_{\beta}(n-2k+1, 2k) \geq \alpha.$$

Note that this solution is approximate. In general there exists no value of n for which the probability in (13) is exactly equal to α for fixed α and β and specified order statistics, but there is always a smallest n for which (15) holds.

A question which naturally arises here is whether there are functions L_1 and L_2 of the sample X_1, X_2, \dots, X_n other than order statistics for which the problem of tolerance limits can be solved if nothing is assumed about the population cdf $F(x)$ beyond continuity. This is still an open question. However, if $F(x)$ has a derivative, Robbins [58] has proved that the only (distribution-free) tolerance limits which are symmetric in the sample values X_1, X_2, \dots, X_n are order statistics.

5. Distribution of single order statistics. The problem of determining the sampling theory of an order statistic consists of applying an appropriate transformation to the coverage distribution (7). In this section we shall consider some special cases of interest.

(a) *Exact distribution of $X_{(r)}$.* Suppose X_1, X_2, \dots, X_n is a sample from a population having a continuous cdf $F(x)$ which has a derivative (pdf) $f(x)$. The probability element $\phi(x_{(r)})dx_{(r)}$ for the r th order statistic $X_{(r)}$ is given by applying the transformation $u = F(x_{(r)})$ to (7), thus yielding

$$(16) \quad \begin{aligned} & \phi(x_{(r)})dx_{(r)} \\ &= \frac{\Gamma(n+1)}{\Gamma(r)\Gamma(n-r+1)} [F(x_{(r)})]^{r-1} [1-F(x_{(r)})]^{n-r} f(x_{(r)}) dx_{(r)}. \end{aligned}$$

It should be noted that as soon as we go from (7) to (16) via the transformation $u = F(x_{(r)})$ we pass from consideration of nonparametric problems to parametric ones. That is, we cannot calculate probabilities of $X_{(r)}$ falling in specified intervals from (16) without knowing $F(x)$, whereas we can calculate similar probabilities pertaining to $F(X_{(r)})$ from (7).

A. T. Craig [4] gave special cases of (16) in 1932, namely, for the

median and the two quartiles of the sample. For samples of size $2m+1$, $X_{(m+1)}$ is the *sample median*, and for samples of size $4k+3$, $X_{(k+1)}$ and $X_{(3k+3)}$ are the *lower* and *upper quartiles of the sample*.

(b) *Limiting distribution of largest (or smallest) value in a sample.* Limiting forms of the distribution (16) for $r=n$ (or 1) as $n \rightarrow \infty$ and properties of these limiting forms for various classes of cdf's $F(x)$ have been considered by a number of authors, particularly Dodd [10], Smirnoff [65], Fisher and Tippett [14], Fréchet [15], Gumbel [17], and von Mises [35]. One of the earliest studies of these extreme values was made in 1923 by Dodd [10], who investigated stochastic limits of greatest values in samples of size n as $n \rightarrow \infty$. Dodd's results, although of interest in the theory of probability, are not particularly useful in statistical inference, and therefore no attempt will be made to summarize them in detail here. It will, perhaps, be sufficient to state a typical one of Dodd's theorems:

If $f(x)$ is the pdf of X , such that $0 < k_1 < x^{1+\alpha}f(x) < k_2 < \infty$ for $x > 0$, and $f(x) = 0$ for $x < 0$, where $\alpha > 0$, then as the sample size $n \rightarrow \infty$,

$$(17) \quad \lim_{n \rightarrow \infty} \Pr (n^{1/\alpha-\epsilon} < X_{(n)} < n^{1/\alpha+\epsilon}) = 1$$

for any $\epsilon > 0$.

Dodd established five other theorems similar to this one, depending on various assumptions regarding the behavior of $f(x)$, particularly for large values of x .

One of the most important items in Dodd's paper is his remark to the effect that if $\xi_{0.5}^{(n)}$ is the median of the pdf $\phi(x_{(n)})$ of the largest sample value $X_{(n)}$ in a sample from a population having cdf $F(x)$ and pdf $f(x)$, then $\xi_{0.5}^{(n)}$ may be determined by solving the equation $[F(\xi_{0.5}^{(n)})]^n = 0.5$. The immediate extension of this is that if $\xi_p^{(n)}$ is the p -quantile of the cdf of $X_{(n)}$ then the probability of all values in the sample being less than $\xi_p^{(n)}$ is equal to p , that is,

$$(18) \quad [F(\xi_p^{(n)})]^n = p,$$

from which $\xi_p^{(n)}$ is given by solving the equation

$$(19) \quad F(\xi_p^{(n)}) = p^{1/n}.$$

In 1925 Tippett [72], using the immediate extension (18) of Dodd's results, tabulated the cdf $\int_{-\infty}^x \phi(x_{(n)}) dx_{(n)}$ of the largest value in samples of size n from a normal population for which

$$(20) \quad F(x) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^x e^{-t^2/2} dt.$$

Thus, the basic formula used by Tippett in his tabulation was

$$(21) \quad \int_{-\infty}^x \phi(x_{(n)}) dx_{(n)} = \left[\frac{1}{(2\pi)^{1/2}} \int_{-\infty}^x e^{-t^2/2} dt \right]^n.$$

Tippett's tabulation was made for $n=3, 5, 10, 20, 50, 100, 200, 300, \dots, 1000$ and for x ranging from -2.5 to 6.5 by intervals of 0.1 .

In 1928 Fisher and Tippett [14] obtained limiting distributions, as $n \rightarrow \infty$, of the greatest (and least) values in a sample from a population having a cdf $F(x)$ and satisfying certain regularity conditions for large values of x . They used an ingenious method of functional equations, based on the fact that the largest value in a sample of size mn has the same distribution as the largest value in a sample of size n from a population of largest values of samples of size m . They arrived at three types of limiting distributions. Almost simultaneously, Fréchet [15] arrived at one of the types. R. von Mises [35] has set forth more explicitly the regularity conditions on which the Fisher-Tippett results depend. Fisher and Tippett obtained three important classes of limiting distributions of $X_{(n)}$, the largest value in a sample of size n from a population having cdf $F(x)$, which may be stated as follows:

CASE I. *Suppose $F(x)$ is a cdf which is less than 1 for every finite value of x , and which is continuous and has two derivatives for all x greater than some finite value x_0 , and that*

$$\lim_{x \rightarrow \infty} \frac{d}{dx} \left(\frac{1 - F(x)}{F'(x)} \right) = 0.$$

Then

$$(22) \quad \lim_{n \rightarrow \infty} \Pr [(X_{(n)} - l_n)nF'(l_n) \leq u] = e^{-e^{-u}}$$

for any u on the interval $(-\infty, +\infty)$, where l_n satisfies the condition

$$F(l_n) = (n-1)/n.$$

CASE II. *Suppose $F(x)$ is a cdf which is less than 1 for every finite value of x , and has a derivative $F'(x)$ for x greater than some finite value x_0 such that*

$$\lim_{x \rightarrow \infty} \frac{x F'(x)}{1 - F(x)} = k$$

where $k > 0$. Then

$$(23) \quad \lim_{n \rightarrow \infty} \Pr \left(\frac{X_{(n)}}{l_n} \leq u \right) = e^{-u^{-k}}$$

for any u on the interval $(0, \infty)$, where l_n is defined in Case I.

CASE III. Suppose $F(x)$ is a cdf such that $F(a) = 1$, and that the first $k-1$ derivatives of $F(x)$ exist and are 0 at $x=a$. Suppose $F^{(k)}(a) = (-1)^{k+1}c$, where $c > 0$, and that $F^{(k+1)}(x)$ is bounded on the interval $(a-\epsilon, a)$ for some $\epsilon > 0$. Then

$$(24) \quad \lim_{n \rightarrow \infty} \Pr \left((X_{(n)} - a) \left(\frac{nc}{k!} \right)^{1/k} \leq u \right) = e^{-(-u)^k}$$

for any u on the interval $(-\infty, 0)$.

In 1927 Fréchet [15] obtained the limiting distribution in Case II in a slightly different form. The issue of the journal in which his result was published was not printed until 1928, thus indicating that Fréchet's results and those by Fisher and Tippett actually appeared almost simultaneously.

To take an important example of Case I, suppose $F(x)$ is the normal cdf given by (20). After some calculation, neglecting terms of order l_n^{-2} and $(\log n)^{-1}$, the expression $(X_{(n)} - l_n)nF'(l_n)$ reduces to

$$(25) \quad \left[X_{(n)}(2 \log n)^{1/2} - \log \left(\frac{n^2}{(4\pi)^{1/2} \log n} \right) \right].$$

Thus, for large n , the probability element of this expression is approximately

$$(26) \quad e^{-u - e^{-u}} du.$$

If one were interested in the largest value $X_{(n)}$ in a sample from a normal population with mean m and variance σ^2 , he would merely replace $X_{(n)}$ in (25) by $(X_{(n)} - m)/\sigma$.

A rather direct approach to the problem of obtaining the limiting distribution of the largest or smallest values in a sample can be made from (16). Consider the case of the largest value. We have $r = n$ in (16). The probability element of Y , where $Y = n[1 - F(X_{(n)})]$ and where $X_{(n)}$ has probability element (16) is

$$(27) \quad (1 - y/n)^{n-1} dy \quad (0 \leq y \leq n)$$

and hence the limiting distribution as Y as $n \rightarrow \infty$ has probability element

$$(28) \quad e^{-y} dy.$$

The problem of finding the approximate distribution of $X_{(n)}$ for large

n is a matter of making the transformation $Y = n[1 - F(X_{(n)})]$ on (28).

(c) *Limiting distribution of $X_{(r)}$.* Results similar to these given in the preceding paragraphs can be written down for the limiting distribution for the r th largest (or r th smallest) order statistic for fixed r . For example, in the case of the r th largest statistic $X_{(n-r+1)}$, the limiting distribution of $Y = n[1 - F(X_{(n-r+1)})]$ has probability element

$$(29) \quad \frac{y^{r-1}e^{-y}}{\Gamma(r)} dy$$

which plays the fundamental role in deriving the limiting distribution of $X_{(n-r+1)}$ for any particular cdf $F(x)$ and pdf $f(x)$ satisfying suitable regularity conditions.

If, however, one allows r to increase with n so that $r/n = p$, and if $f(x)$ is continuous and not equal to 0 at the p -quantile ξ_p , then the order statistic $X_{(r)}$ is asymptotically normally distributed with mean ξ_p and variance $p(1-p)/n[f(\xi_p)]^2$, that is,

$$(30) \quad \lim_{n \rightarrow \infty} \Pr \left[\frac{(X_{(r)} - \xi_p)f(\xi_p)n^{1/2}}{(p(1-p))^{1/2}} \leq u \right] = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^u e^{-t^2/2} dt.$$

Of the authors who have considered the limiting distributions of order statistics, Smirnov [65] made a rather systematic and comprehensive study of them in 1935, although there is no evidence that he was aware of the results on largest values which had been reached earlier by Fisher and Tippett [14] and Fréchet [15].

6. Joint distributions of several one-dimensional order statistics and applications. There are problems in parametric statistical inference which involve the sampling theory of two or more order statistics. Determination of the sampling distributions of the sample range ($X_{(n)} - X_{(1)}$) and midrange $(X_{(1)} + X_{(n)})/2$ are examples of such problems. The exact sampling theory of two or more order statistics in samples from a population having a cdf $F(x)$ with a continuous derivative $f(x)$ is straightforward and presents no particular difficulties. (See [87].) For example, consider two order statistics $X_{(r)}$ and $X_{(r')}$ ($r < r'$). The probability element $\phi(x_{(r)}, x_{(r')})dx_{(r)}dx_{(r')}$ is obtained by applying the transformation

$$(31) \quad u = F(x_{(r)}), \quad v = F(x_{(r')}) - F(x_{(r)})$$

to (8) and setting $s = r' - r$. Setting $r' = r + 1$, we have the joint distribution of $X_{(r+1)}$ and $X_{(r)}$, from which Karl Pearson [51] obtained the solution to the Galton difference problem, namely

$$(32) \quad E(X_{(r+1)} - X_{(r)}) = \frac{\Gamma(n+1)}{\Gamma(n-r+1)\Gamma(r+1)} \int_{-\infty}^{\infty} F(x)^{n-r} [1-F(x)]^r dx.$$

(a) *Distribution of sample range.* For $r=1$, and $r'=n$, we have the probability element for the two extremes in a sample from which it follows that the probability element $g(r)dr$ of the exact distribution of the *sample range* $R = X_{(n)} - X_{(1)}$ is

$$(33) \quad g(r)dr = n(n-1) \left\{ \int_{-\infty}^{\infty} [F(r+x_{(1)}) - F(x_{(1)})]^{n-2} f(r+x_{(1)}) f(x_{(1)}) dx_{(1)} \right\} dr.$$

A similar expression holds for the probability element of the mid-range $M = (X_{(n)} + X_{(1)})/2$.

The problem of the distribution theory of the sample range was originally discussed by von Bortkiewicz [1] in 1921. In 1925 Tippett [72] showed that the mean value of R is given by

$$(34) \quad E(R) = \int_{-\infty}^{\infty} \{1 - [F(x)]^n - [1 - F(x)]^n\} dx.$$

He tabulated $E(R)$ and also the variance of R for samples from size 2 to 1,000 from a normal population in which $F(x)$ is given by (20). In 1933 McKay and E. S. Pearson [33] determined explicit expressions for the distributions of ranges in samples of size 2 and 3 from a normal population. More recently, Hartley [23] has tabulated the cdf of the range R for samples of size 2, 3, 4, \dots , 20 from a normal population with zero mean and unit variance.

Setting up new random variables Y and Z defined by

$$(35) \quad Y = nF(X_{(1)}), \quad Z = n[1 - F(X_{(n)})]$$

one finds Y and Z independently distributed in the limit as $n \rightarrow \infty$, which means that $X_{(1)}$ and $X_{(n)}$ are *asymptotically independently distributed in large samples*, and hence the problem of determining the limiting distribution of the range R as $n \rightarrow \infty$ is one of dealing with the difference between two asymptotically independently distributed random variables. This limiting distribution has been discussed by Gumbel [20], who has also dealt with what he calls *mth ranges* $(X_{(m)} - X_{(n-m+1)})$ and *mth midranges* $(X_{(m)} + X_{(n-m+1)})/2$. In particular, he shows that for fixed m , the order statistics $X_{(m)}$ and $X_{(n-m+1)}$ are independently distributed in the limit, as $n \rightarrow \infty$. More recently Gumbel [22] has studied the *reduced range* $R^* = g_n[R - 2l_n]$ where l_n is given by $F(l_n) = (n-1)/n$ and g_n is the ratio

$F'(l_n)/[1 - F(l_n)]$. He has made some tabulations for the case of samples of size n from a normal population. Carlton [3] has given the asymptotic distribution of the range of a sample from a population having probability element dx/θ on the interval $(0, \theta)$.

(b) *Limiting distribution of two or more order statistics in large samples.* Smirnoff [65] has shown that if n , n_1 , and n_2 increase indefinitely so that $n_1/n = p_1$, and $n_2/n = p_2$ ($0 < p_1, p_2 < 1$), then the order statistics $X_{(n_1)}$ and $X_{(n_2)}$ in a sample of size n from a population having a pdf $f(x)$ (which is greater than 0 and continuous at the quantiles ξ_{p_1} and ξ_{p_2}) are asymptotically jointly normally distributed with means ξ_{p_i} ($i=1, 2$), variances $p_i(1-p_i)/n[f(\xi_{p_i})]^2$ ($i=1, 2$), and covariance $p_1(1-p_2)/nf(\xi_{p_1})f(\xi_{p_2})$. It should be remarked that these variances and covariance were derived as early as 1920 by Karl Pearson [53]. However, he did not establish the fact that the limiting distribution of $X_{(n_1)}$ and $X_{(n_2)}$, as $n \rightarrow \infty$, is normal. Mosteller [39] has extended Smirnoff's result to a set of k order statistics, obtaining, of course, a normal k -variate limiting distribution.

(c) *Estimation of population parameters by order statistics.* Mosteller [39] has investigated the efficiency of various linear combinations of several order statistics in large samples for estimating the mean and variance of a normal distribution function. To pick one simple case, he shows that by choosing $p_1 = .2702$ and $p_2 = .7298$, the average of the two order statistics $x_{(np_1)}$ and $x_{(np_2)}$ have an efficiency 0.81 in estimating the population mean for large n . We are using efficiency here in the sense of R. A. Fisher, briefly and roughly stated as follows: if the "most efficient" method of estimating the population mean had been used (the arithmetic mean of all sample values in this case) a sample of size $0.81n$ would produce an estimate with the same variance, to terms of order $1/n$, as that for the estimate based on the two given order statistics. To state the matter another way, this means that the use of these two order statistics for estimating the population mean "utilized about 81% of the information contained in the sample." Among other results, Mosteller has shown that by using the average of 10 properly spaced order statistics one can obtain an estimate of the mean with an efficiency of more than .97. He considered the cases of 2, 3, 4, \dots , 10 symmetrically spaced order statistics in the estimation of the mean of a normal population. He also considered order statistics in the estimation of the variance σ^2 of a normal population. The use of the range for this purpose in small samples was investigated by Davies and Pearson [8]. They found that for samples of size 10 or less, the range (multiplied by a suitable factor) is almost as accurate for estimating σ as the cus-

tomary method based on the sum of squares of deviations of sample values from their mean.

The use of order statistics promises to furnish a simple and effective practical method for estimating parameters of the normal and other populations having continuous pdf's for both large and small samples—especially in situations where large samples are easily available. The development of this field will depend on a vast amount of computation, but the reward will be extremely simple statistical methods useful for both large and small samples under certain practical conditions.

(d) *Application of order statistics to breaking strength of thread bundles.* An interesting application of the joint distribution theory of a set of one-dimensional order statistics has been made by Daniels [7] to the statistical theory of breaking strength of bundles of threads, which illustrates a type of application of order statistics which may be important in the critical study of other physical phenomena of a similar nature.

Suppose the breaking strength of a population of thread lengths has a continuous cdf $F(x)$. Let the ordered breaking strengths (from least to greatest) of a sample be $X_{(1)}, X_{(2)}, \dots, X_{(n)}$. Now suppose these threads are made into a bundle of n parallel threads and let a load S be applied and consider the probability that S will break the bundle. It is evident that the bundle will break if the following conditions are satisfied:

$$(36) \quad \begin{aligned} 0 < X_{(1)} < S/n, \\ X_{(i-1)} < X_{(i)} < S/(n+1-i) \quad (i = 2, 3, \dots, n). \end{aligned}$$

If we denote by B the region in sample space for which these inequalities hold then the probability that S will break the bundle is

$$(37) \quad p_n(S) = n! \int_B dF(x_{(1)}) dF(x_{(2)}) \cdots dF(x_{(n)}).$$

But, setting $p_i = F(x_{(i)})$ ($i = 1, 2, \dots, n$), this is equivalent to the following integration

$$(38) \quad p_n(S) = n! \int_0^{b_n} \int_{p_1}^{b_{n-1}} \int_{p_2}^{b_{n-2}} \cdots \int_{p_{n-1}}^{b_1} dp_n dp_{n-1} \cdots dp_1$$

where $b_i = F(S/i)$, $i = 1, 2, \dots, n$, which yields

$$(39) \quad p_n(S) = \sum_r \frac{n!(b_1 - b_2)^{r_1} (b_2 - b_3)^{r_2} \cdots (b_{n-1} - b_n)^{r_{n-1}} b_n^{r_n}}{r_1! r_2! \cdots r_n!}$$

where \sum_r denotes summation over all values of r_1, r_2, \dots, r_n (positive integers or zero) for which

$$\sum_{j=1}^i r_j \leq i \quad (i = 1, 2, \dots, n - 1)$$

and $\sum_1^n r_i = n$.

For a specified sample of size n , let S^* be the largest value of S for which the inequalities (36) are not all true. Then S^* is a random variable representing the *breaking load* of the bundle of n threads for this sample. Daniels has shown that if, in addition to the continuity of $F(x)$, we assume that $\lim_{x \rightarrow \infty} x[1 - F(x)] = 0$ and $x[1 - F(x)]$ has a maximum for $x = x^*$, then

$$(40) \quad \lim_{n \rightarrow \infty} \Pr \left\{ \frac{S^* - nx^*[1 - F(x^*)]}{x^*(nF(x^*)[1 - F(x^*)])^{1/2}} \leq u \right\} = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^u e^{-t^2/2} dt.$$

In other words, S^* is approximately normally distributed in large samples with mean $nx^*[1 - F(x^*)]$ and variance $nx^{*2}F(x^*)[1 - F(x^*)]$.

7. Confidence bands for the cdf $F(x)$. A fundamental problem in nonparametric statistical inference is that of finding confidence bands for a continuous cdf $F(x)$ from a sample of size n . The natural approach to this problem is to make use of the sample cdf $F_n(x)$ in determining these confidence bands. $F_n(x)$ is a function of order statistics defined as follows:

$$(41) \quad \begin{aligned} F_n(x) &= 0, & x < X_{(1)}, \\ F_n(x) &= i/n, & X_{(i)} \leq x < X_{(i+1)}, \\ F_n(x) &= 1, & x \geq x_{(n)}. \end{aligned}$$

The sample cdf $F_n(x)$ has been referred to by Cramér [5] as the *statistical image* of the population cdf $F(x)$.

Consider the transformation $T_{(1)} = F(X_{(1)})$, $T_{(2)} = F(X_{(2)})$, \dots , $T_{(n)} = F(X_{(n)})$. It is clear that $T_{(1)}, \dots, T_{(n)}$ are order statistics in a sample of size n from a population having probability element dt on the interval $(0, 1)$. The cdf of T is t . Denoting the sample cdf by $G_n(t)$, it is seen that $\text{Sup}_{(t)} [t - G_n(t)] = \text{Sup}_{(x)} [F(x) - F_n(x)]$.

By working with the simplified random variable T and using the fact just stated, Kolmogoroff [30] showed in 1933 that

$$(42) \quad \lim_{n \rightarrow \infty} \Pr \left\{ \sup_{(x)} [F(x) - F_n(x)] n^{1/2} \leq \lambda \right\} = \Phi(\lambda)$$

where

$$(43) \quad \Phi(\lambda) = \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2\lambda^2}.$$

Thus, for a given confidence coefficient α , approximate confidence bands for $F(x)$ are furnished by the two functions $F_n(x) \pm \lambda_\alpha/n^{1/2}$, where λ_α is determined so that $\Phi(\lambda) = \alpha$. Thus, we can say that for large n , the probability is approximately α that the curve $y = F(x)$ will be contained in the band bounded by the four graphs: $y = F_n(x) \pm \lambda_\alpha/n^{1/2}$, $y = 0$ and $y = 1$. Kolmogoroff tabulated values of λ_α for $\alpha = .95, .98, .99, .995, \text{ and } .999$.

A result closely related to Kolmogoroff's and obtained by Smirnov [66] relates to the difference between two sample cdf's. Suppose $F_m(x)$ and $F_n(x)$ are sample cdf's for samples of sizes m and n respectively from a population having a continuous cdf $F(x)$, and the ratio m/n remains between two fixed positive numbers a_1 and a_2 . Then Smirnov showed that

$$(44) \quad \lim_{n \rightarrow \infty} \Pr \left\{ \sup_{(x)} \frac{[F_m(x) - F_n(x)]}{(1/m + 1/n)^{1/2}} \leq \lambda \right\} = \Phi(\lambda).$$

Wald and Wolfowitz [76] have considered the problem of setting up exact confidence bands for $F(x)$. They have shown, for confidence coefficient α and constant d_α , how to construct confidence bands for $F(x)$ of the form $F_n(x) \pm d_\alpha$. Thus, the probability is α that the region or band bounded by the graphs of $y = F_n(x) \pm d_\alpha$, $y = 0$, and $y = 1$ contains the curve $y = F(x)$. There are many ways of determining the d_α , but their evaluation is rather laborious. It would be very desirable to have an investigation of sets of values of the d_α which would produce "best" confidence bands, perhaps of smallest average vertical width weighted in some sense, and to have suitable tabulations to facilitate their application.

8. Sampling distributions of coverages for the case of two or more dimensions. In §3 we discussed the sampling distribution of the set of elementary coverages produced by the order statistics in a sample of size n from a population having a one-dimensional continuous cdf $F(x)$. We recall that these elementary coverages have the extremely simple probability element (6), which does not depend on the population cdf $F(x)$, from which one can readily determine the distribution of two or more sums of elementary coverages. We shall now deal with the problem of coverages for the case of two or more dimensions.

For the sake of simplicity, we shall deal with the case of two

dimensions. Extension of the results to more than two dimensions is straightforward. Thus, suppose we have a sample O_n of size n , say $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$, from a population having a continuous cdf $F(x, y)$. This sample can be represented as n points in the xy -plane. The basic question is this: *What kinds of two-dimensional intervals or regions S determined by the sample O_n are there, if any, which have distribution-free coverages?*

We shall discuss several methods of constructing such regions.

(a) *For regions determined by one-dimensional order statistics.* It is obvious from the one-dimensional case that the *slices* produced by the n lines parallel to the y axis passing through the n sample points are regions having distribution-free coverages. In fact, the coverages (two-dimensional) for these regions have the probability element given by (6). This is evident from the fact that this particular slicing operation is equivalent to considering only the X 's of the sample in precisely the same manner as we did in arriving at (6). Entirely similar results hold, of course, if the slicing is carried out with the n lines through the sample points parallel to the x -axis. In this case we would be considering only the Y 's of the sample.

The preceding considerations suggest a more general method of slicing the xy -plane into $n+1$ regions or sets. Suppose $h(x, y)$ is a function such that if X and Y are random variables with continuous cdf $F(x, y)$, the cdf of $h(X, Y)$ ($=Z$, say) is continuous. For the n sample points there will be n values of Z , namely $Z_i = h(X_i, Y_i)$. Suppose $Z_{(i)}$ ($i=1, 2, \dots, n$) are the ordered values of the Z_i . Now there will be $n+1$ regions or sets of points in the xy -plane, S_1, S_2, \dots, S_{n+1} , where S_1 is the set of points for which $h(x, y) < Z_{(1)}$, S_2 the set for which $Z_{(1)} < h(x, y) < Z_{(2)}$ and so on. Let C_1, C_2, \dots, C_{n+1} be the coverages associated with S_1, S_2, \dots, S_{n+1} , respectively. Then these coverages are distribution-free and any n of them have probability element (6). This follows at once from the fact that under the conditions stated, $Z = h(X, Y)$ is a random variable having a continuous cdf and that we are simply considering the elementary coverages determined by the order statistics $Z_{(1)}, Z_{(2)}, \dots, Z_{(n)}$.

But there are more general regions than these which are determined by the sample and for which the coverages are distribution-free as we shall see in the following paragraphs.

(b) *For rectangular regions in case of independent random variables.* In case X and Y are independent in the probability sense, that is, when $F(x, y) = F_1(x) \cdot F_2(y)$ ($F_1(x)$ and $F_2(y)$ being continuous cdf's of X and Y respectively), one can construct rectangular regions having distribution-free coverages by setting up a product distribution in the

product space of X and Y as follows: Suppose the sample O_n is ordered with respect to the X 's. The order statistics $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ determine a set of one-dimensional elementary coverages having probability element (6). The sum U of any r ($r \leq n$) of these elementary coverages has probability element (7). Similarly, if the sample is ordered with respect to the Y 's, a set of one-dimensional elementary coverages is determined such that the sum V of any s ($s \leq n$) of them has probability element (7) with u replaced by v and r by s . Since U and V are functions of independently distributed random variables, they are themselves independently distributed. If I_x is the set of x intervals belonging to U and I_y is the set of y intervals belonging to V , then $I_x \cdot I_y$ is a two-dimensional product set of rectangles, say S . The sum of the two-dimensional coverages over S is simply the product UV . Hence the probability element of the two-dimensional coverage $UV = W$, say, can be calculated from the joint distribution of U and V ,

$$(45) \quad \frac{\Gamma^2(n+1)}{\Gamma(r)\Gamma(s)\Gamma(n-r+1)\Gamma(n-s+1)} u^{r-1}v^{s-1}(1-u)^{n-r}(1-v)^{n-s} du dv$$

by integrating u and v over the region for which $w < uv < w + dw$ and neglecting terms of order $(dw)^2$ and higher.

To take a case in which the two-dimensional region S is very simple, suppose U is taken as the sum of the elementary coverages for the intervals $(0, X_{(1)}), (X_{(1)}, X_{(2)}), \dots, (X_{(r-1)}, X_{(r)})$, then $U = F_1(X_{(r)})$. Similarly, let $V = F_2(Y_{(s)})$. Then the two-dimensional set S is the "corner" for which $x < X_{(r)}, y < Y_{(s)}$, and the two-dimensional coverage over S is $F_1(X_{(r)}) \cdot F_2(Y_{(s)})$. This two-dimensional coverage is the product of two independently distributed random variables U and V which have distributions independent of $F_1(x)$ and $F_2(y)$. Consequently, the distribution of $F_1(X_{(r)}) \cdot F_2(Y_{(s)})$ is independent of $F_1(x)F_2(y)$.

This product method of constructing two-dimensional regions having distribution-free coverages can readily be extended to any number of dimensions. It was devised by the author [86] in connection with the problem of tolerance limits for two independent variables. To take a simple case, the rectangular region S for which $X_{(1)} < x < X_{(n)}$ and $Y_{(1)} < y < Y_{(n)}$ is such that coverage W over it has probability element

$$(46) \quad n^2(n-1)^2 w^{n-2} [2(w-1) - (w+1) \log w] dw.$$

By integrating this expression from β to 1 and setting the result equal to α one has an equation involving the three variables α, β

and n . By fixing two of them one may solve for the third. For instance if $\alpha=0.95$ and $n=500$ one finds $\beta=0.991$. Thus, we may say that the probability is 0.95 that, in a sample of size 500, at least 99.1% of the population will be included in the rectangle determined by the range of the X 's and the range of the Y 's in the sample.

By using the product method one can find the joint distribution of one or more X order statistics and one or more Y order statistics in a sample from a two-dimensional population in which the two random variables are independent, and similarly for higher dimensions. But in the case of dependence this problem is extremely complicated. Mood [37] has given some results on the joint distribution of the median of X 's and the median of Y 's in a sample from a population in which x and y are not independent. He showed that the asymptotic joint distribution of the medians for the multivariate case is normal and determined the matrix of the variances and covariances for the two and three dimensional cases.

(c) *Wald's results for rectangular regions in case the random variables are not independent.* The product method of constructing regions described above is severely restricted by the requirement of independence, that is, that $F(x, y) = F_1(x) \cdot F_2(y)$. The problem of constructing rectangular regions having distribution-free coverages for the case of samples from populations having any continuous cdf in two or more dimensions was solved by Wald [75] in his study of the problem of tolerance limits, when the random variables for the population are not assumed to be independent.

Let us consider the two-dimensional case. We have n sample pairs represented as n points in the xy -plane. Consider the X order statistics $X_{(1)}, X_{(2)}, \dots, X_{(n)}$. The lines $x = X_{(r_1)}$ and $x = X_{(s_1)}$ ($r_1 < s_1$) cut the xy -plane into three slices. There are $s_1 - r_1 - 1$ sample points lying within the middle slice. Let the Y order statistics for the $s_1 - r_1 - 1$ points of the middle slice be denoted by $Y'_{(r_1+1)}, Y'_{(r_1+2)}, \dots, Y'_{(s_1-1)}$. Consider the region S consisting of the portion of the middle slice which lies between the lines $y = Y'_{(r_2)}$ and $y = Y'_{(s_2)}$ ($r_2 < s_2$). S is therefore a rectangle defined by $X_{(r_1)} < x < X_{(s_1)}$, and $Y'_{(r_2)} < y < Y'_{(s_2)}$. Within S there are $s_2 - r_2 - 1$ sample points. Wald's basic result is as follows:

The region S has a distribution-free coverage W which has probability element

$$(47) \quad \frac{\Gamma(n+1)}{\Gamma(s_2 - r_2)\Gamma(n - s_2 + r_2 + 1)} w^{s_2 - r_2 - 1} (1 - w)^{n - s_2 + r_2} dw.$$

The proof of this statement depends on an ingenious application of conditional probability theory.

Wald extended this result so that the region S could consist of several rectangles rather than a single one. More specifically, suppose k ($< n$) vertical lines $x = X_{(m_i)}$ ($i = 1, 2, \dots, k$) are drawn. The xy -plane will be cut into $k + 1$ slices. Within the left-most slice there will be $m_1 - 1$ sample points. Within the next slice there will be $m_2 - m_1 - 1$ sample points, and so on. Now suppose we draw a horizontal line through each of the $m_1 - 1$ sample points within the left-most slice, each horizontal line being drawn completely across this slice but not extending into the other slices. We thus cut the first slice into m_1 rectangular blocks. (Note that two of the blocks in this slice have two infinite boundaries each, while $m_1 - 2$ of the blocks have one infinite boundary each.) We do a similar operation on each of the remaining slices. The total number of rectangular blocks into which the xy -plane has been cut is $n + 1$. Let R_{ij} be the j th block in the i th slice, and let S^* be a set consisting of s of these rectangular blocks. Let W^* be the sum of the coverages for these s blocks, that is, $W^* = \int_{S^*} dF(x, y)$.

Wald's extended result is this:

W^* is a distribution-free coverage with probability element

$$(48) \quad \frac{\Gamma(n + 1)}{\Gamma(s)\Gamma(n - s + 1)} (w^*)^{s-1} (1 - w^*)^{n-s} dw^*.$$

Actually, it can be shown that any n of the coverages for the $n + 1$ blocks R_{ij} have probability element (7). The extension of Wald's two-dimensional results to the k -dimensional case is straightforward, and, in fact, has been carried out by Wald himself.

(d) *Tukey's generalization of Wald's results.* Tukey [73] has generalized Wald's ideas to a stage which shows the full possibilities of the use of order statistics in cutting up the xy -plane into regions having distribution-free coverages. Although Tukey has generalized Wald's results for any finite number of dimensions we shall again, for the sake of simplicity, confine ourselves to the two-dimensional case.

Let $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ be a sample of size n from a population having cdf $F(x, y)$, and consider the n points in the xy -plane which represent the sample. Suppose, for each value of i ($i = 1, 2, \dots, n$), that $h_i(x, y)$ is a function such that $h_i(X, Y)$ is a random variable having a continuous cdf. Two or more of these functions may be identical. Thus, $h_i(x, y)$ may be used to order the sample points so that the values $h_i(X_j, Y_j) = Z_{ij}$ ($j = 1, 2, \dots, n$) will be represented as n points on some interval (a_i, b_i) . The Z_i order statistics will be $Z_{i(1)}, Z_{i(2)}, \dots, Z_{i(n)}$. The "curve" $h_1(x, y) = Z_{1(n)}$ cuts the xy -plane into two regions or sets (we exclude boundaries;

the probability contained in them is zero since the cdf's of the $h_i(X, Y)$ are continuous):

$$S_1: \text{ for which } h_1(x, y) > Z_{1(n)},$$

and

$$S_1': \text{ for which } h_1(x, y) < Z_{1(n)}.$$

The "curve" $h_2(x, y) = Z_{2(n-1)}$ cuts S_1' into two sets

$$S_2: \text{ for which } h_2(x, y) > Z_{2(n-1)},$$

and

$$S_2': \text{ for which } h_2(x, y) < Z_{2(n-1)}.$$

Similarly, we cut S_2' into two sets, S_3 and S_3' , by the "curve" $h_3(x, y) = Z_{3(n-2)}$, and so on. The xy -plane therefore will be cut up into $n+1$ disjoint sets, S_1, S_2, \dots, S_{n+1} which are completely defined as follows:

$$(49) \quad \begin{aligned} S_1: & \{ h_1(x, y) > Z_{1(n)} \} \\ S_2: & \{ h_1(x, y) < Z_{1(n)}, h_2(x, y) > Z_{2(n-1)} \} \\ S_3: & \{ h_1(x, y) < Z_{1(n)}, h_2(x, y) < Z_{2(n-1)}, h_3(x, y) > Z_{3(n-2)} \} \\ & \vdots \\ S_n: & \{ h_1(x, y) < Z_{1(n)}, \dots, h_{n-1}(x, y) < Z_{n-1(2)}, h_n(x, y) > Z_{n(1)} \} \\ S_{n+1}: & \{ h_i(x, y) < Z_{i(n+1-i)}, i=1, 2, \dots, n \}. \end{aligned}$$

If C_1, C_2, \dots, C_{n+1} are the coverages (two-dimensional) for S_1, S_2, \dots, S_{n+1} , respectively, we have $\sum_{i=1}^{n+1} C_i = 1$. It is to be noted that there are $n!$ ways in which the xy -plane can be sliced up into $n+1$ sets of type S_1, S_2, \dots, S_{n+1} , depending on the order in which the functions $h_i(x, y)$ are selected. Tukey's basic *slicing principle* is as follows:

For any given order in which the sets S_1, S_2, \dots, S_{n+1} are determined, the coverages C_1, C_2, \dots, C_{n+1} associated with these sets are distribution-free and have probability element (6).

This principle is more general than it may first appear, for we could have cut the xy -plane into two sets by using a "cutting curve" determined by a general order statistic $Z_{1(t_1)}$, that is, $h_1(x, y) = Z_{1(t_1)}$, at the first step. The xy -plane is thus cut into two sets containing $t_1 - 1$ and $n - t_1$ sample points respectively. Similarly, each of these sets can be cut into two sets by using a "cutting curve" determined

by a general order statistic $Z_{2(t_2)}$, that is, $h_2(x, y) = Z_{1(t_1)}$, and so on, until the xy -plane has been cut into $n+1$ disjoint sets.

The possibilities of applying Tukey's slicing principle to statistical inference problems look promising but are still to be explored. One of the most immediate applications is that the method provides a more flexible "trimming" procedure for constructing tolerance regions than one can construct by Wald's method of combining rectangles discussed earlier. As an example, consider n sample points represented in the xy -plane as shown in Fig. 1.

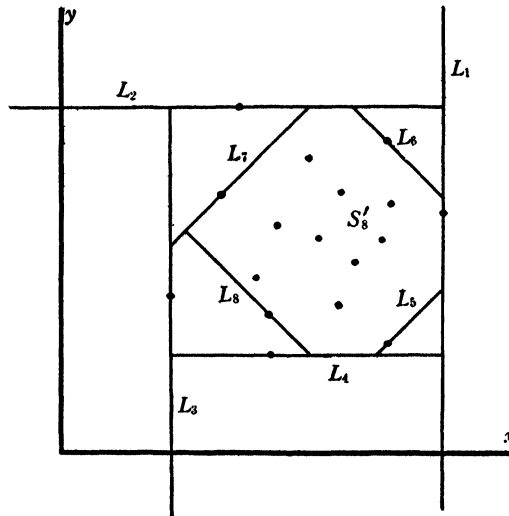


FIG. 1

Now suppose the lines (or line segments) L_1, L_2, \dots, L_8 , as shown in the figure, are drawn in that order. Lines L_1 and L_3 are vertical, L_2 and L_4 horizontal, L_5 and L_7 are 45° with respect to the x axis and L_6 and L_8 are 135° . Note that eight sample points have been used in constructing these eight lines, and that $n-8$ points remain in the interior "residual" region S'_8 . It follows from Tukey's results that the coverage W for S'_8 is distribution-free and has probability element

$$(50) \quad \frac{\Gamma(n+1)}{\Gamma(n-7)\Gamma(8)} w^{n-8}(1-w)^7 dw.$$

Thus, the probability that S'_8 , when used as a tolerance region, contains at least $100\beta\%$ of the population is obtained by integrating (50) from β to 1. Note that by continuing the slicing of S'_8 in any manner in accordance with Tukey's principle one would finally

cut S'_i into $n-7$ regions, thus making a total $(n-7)+8$ or $n+1$ regions having distribution-free coverages with probability element (6).

(e) *Extensions of distribution theory of coverages and order statistics to the discrete case.* It should be noted that all of the results concerning coverages for any number of dimensions which have been discussed heretofore depend on the continuity of the population cdf. It is important to have analogous results for the case in which the cdf is not continuous. This importance stems from the fact that, because of limitations of making measurements, the observational data, more strictly idealized, must be considered as a sample from a *discrete* population, that is, one having a step-function cdf rather than a continuous one. In actual applications ties *do* occur among the values of a sample.

Tukey, in some work not yet published, has shown how the probability theory of coverages and order statistics can be extended to the discrete case in any number of dimensions by substituting inequalities for equalities in the probability statements for the continuous case. Scheffé and Tukey [62] discussed the problem for the one-dimensional case in 1945.

9. Application of multi-dimensional coverages and order statistics to estimation problems. In §6c it was pointed out that some studies had been made of the problem of making "inefficient" estimates of the mean and variance of a normal population by means of order statistics. Analogous problems exist for the case of normal distributions of two or more variables. A few results have been obtained on these problems but much remains to be done.

Let us consider the case of two dimensions. Suppose $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ is a sample from a normal bivariate distribution having means m_x and m_y , variances σ_x^2, σ_y^2 and correlation coefficient ρ . Estimates of the mean and variance of x (or y) may be made from the X (or Y) order statistics along lines described in §6c. But the problem of estimating ρ from order statistics is one which must be considered in terms of using both X and Y order statistics. Mosteller [39] and Hotelling and Pabst [26] have made some contributions to this problem. Mosteller considered the numbers of sample points falling into four regions $S_{11}, S_{12}, S_{21},$ and S_{22} of the xy -plane defined as follows. Draw the vertical lines $x = X_{(m+1)}$ and $x = X_{(n-m)}$ ($2m < n$). There will be m sample points to the left of the first line and m to the right of the second. Draw any horizontal line $y = y'$ such that m of these $2m$ points lie above this line and m below. Consider the four "corner" regions:

$$\begin{aligned}
 (51) \quad S_1: & \quad x > X_{(n-m)}, & y > y', \\
 S_2: & \quad x > X_{(n-m)}, & y < y', \\
 S_3: & \quad x < X_{(m+1)}, & y > y', \\
 S_4: & \quad x < X_{(m+1)}, & y < y'.
 \end{aligned}$$

Let m^* be the number of sample points falling in S_1 . Then considering only the numbers of sample points falling into these four regions, the maximum likelihood estimate of ρ , say R^* , is given by solving the equation,

$$(52) \quad \frac{m^*}{m - m^*} = \frac{np^*}{m - np^*}$$

where

$$(53) \quad p^* = \frac{1}{2\pi(1-\rho^2)^{1/2}} \int_0^\infty \int_k^\infty \exp \left\{ -\frac{1}{2(1-\rho^2)} [x^2 + y^2 - 2\rho xy] \right\} dx dy$$

and k is given by

$$(54) \quad \frac{m}{n} = \frac{1}{(2\pi)^{1/2}} \int_k^\infty e^{-t^2/2} dt.$$

The variance of R^* for large samples is

$$(55) \quad \sigma_{R^*}^2 = \frac{p^*(m - np^*)}{2nm p^*}$$

and the efficiency of this method of estimating ρ is about 51.5% (the Pearson product-moment correlation being 100% efficient), in the case of samples from a normal population in which the true correlation ρ between X and Y is zero. Mosteller has given a graph to facilitate the solution of equation (52) for ρ in practical applications.

Suppose the sample pairs are ordered with respect to the X 's, that is, $(X_{(1)}, Y_{(\alpha_1)}), (X_{(2)}, Y_{(\alpha_2)}), \dots, (X_{(n)}, Y_{(\alpha_n)})$. Then the α 's will be some permutation of the integers $1, 2, \dots, n$. Let R' be the *rank correlation* coefficient in the sample between X 's and Y 's, that is, defined as the ordinary Pearsonian correlation coefficient between the *ranks* $1, 2, \dots, n$ and $\alpha_1, \alpha_2, \dots, \alpha_n$ (see (75)). Pursuing a lead suggested by the work of Karl Pearson [52] on rank correlation, Hotelling and Pabst [26] have proposed an estimate R' of ρ defined by

$$(56) \quad R' = 2 \sin \pi r' / 6 \cdot$$

The variance of R' for the case of $\rho = 0$ is $\pi^2/9n$, and the efficiency of ρ' is 91.2%.

It would be desirable to have further studies of the problem of estimating ρ by other schemes depending on order statistics and simple functions of order statistics. There are probably other very simple "inefficient" methods of estimating ρ which would be useful in situations where large samples can be obtained with very little effort. This is true in such fields as psychological testing where large amounts of data are available on punch cards which have already been used for operational purposes.

In normal multivariate statistical theory it would be desirable to have similar studies made of the problem of estimating correlation coefficients from samples in which only ranks are given for sample values of some of the variables.

10. Order statistics in the testing of statistical hypotheses—the method of randomization. The use of order statistics in the construction of tests for nonparametric statistical hypotheses (valid for populations having any continuous cdf) have been receiving an increasing amount of attention during recent years. A considerable number of results have already been obtained and many more can be expected within the next few years. The main value of nonparametric tests of statistical hypotheses lies in the fact that they are free from the frequently uncomfortable assumption that a population distribution is normal or has some other specific functional form.

One of the simplest and most elegant results in nonparametric statistical inference is the confidence interval theory of the median and other quantiles of a continuous cdf based on order statistics, as discussed in §4(a). Confidence band theory for a continuous cdf $F(x)$ provided by Kolmogoroff's theorem (for large n) and by the Wald-Wolfowitz method (for small n), discussed in §7, furnishes a satisfactory solution to the problem of testing the hypothesis that a sample is from a population with a specified continuous cdf $F_0(x)$.

Beyond these results, there is yet to be developed a satisfactory general theory of nonparametric statistical inference, particularly test theory. The only nonparametric tests which have thus far been developed are rather special cases in which intuition and experience with the well-established parametric test theory have served as guides. What is needed is a general theory of constructing nonparametric tests and a satisfactory method of characterizing the power of these tests. Wolfowitz [88] has made a step in this direction by proposing an extension of the likelihood ratio method, used in deriving parametric statistical tests. He has illustrated this method

with several examples, but more remains to be done before the method can be considered generally useful.

If there is a single property of order statistics which has played the greatest role in the development of nonparametric tests thus far it may be stated as follows:

Suppose X_1, X_2, \dots, X_n is a sample from a population having a continuous cdf $F(x)$. Consider the region $S_{\{\alpha_i\}}$ in the sample space R_n for which the inequality

$$(57) \quad X_{\alpha_1} < X_{\alpha_2} < \dots < X_{\alpha_n}$$

holds, where $\alpha_1, \alpha_2, \dots, \alpha_n$ is any one of the $n!$ permutations of the integers $1, 2, 3, \dots, n$. Then the amount of probability contained in $S_{\{\alpha_i\}}$ is $1/n!$.

The truth of this statement is clear from the fact that

$$(58) \quad \begin{aligned} & \Pr (X_{\alpha_1} < X_{\alpha_2} < \dots < X_{\alpha_n}) \\ &= \Pr \left(\int_{-\infty}^{X_{\alpha_1}} dF(x) < \int_{-\infty}^{X_{\alpha_2}} dF(x) < \dots < \int_{-\infty}^{X_{\alpha_n}} dF(x) \right) \\ &= \Pr (P_{\alpha_1} < P_{\alpha_2} < \dots < P_{\alpha_n}) \end{aligned}$$

where $P_{\alpha_1}, P_{\alpha_2}, \dots, P_{\alpha_n}$ are defined in §3, and have probability element (5). Integrating (5) over the region for which $0 < p_{\alpha_1} < p_{\alpha_2} < \dots < p_{\alpha_n} < 1$ gives the desired result $1/n!$.

A method of devising nonparametric tests of statistical hypotheses on the basis of order relations among the values of a sample known as the *method of randomization* has been suggested by Fisher [12, 13]. The general idea here is to construct a *rejection region* W in sample space R_n by taking enough regions in the set $\{S_{\{\alpha_i\}}\}$ (each of which has probability $1/n!$) so that the probability contained in W will be a specified amount ϵ . This method has been used in a number of special problems, the method of selecting regions from $\{S_{\{\alpha_i\}}\}$ being more or less intuitive in each case. We should note, however, that the randomization method cannot be used to test the hypothesis that $F(x)$ has a specified form $F_0(x)$, against the class of alternatives consisting of all continuous cdf's different from $F_0(x)$. For, suppose we construct a rejection region W for H_0 in sample space R_n by taking the sum of enough of the regions in the set $\{S_{\{\alpha_i\}}\}$ to make the probability contained in W equal to ϵ (or approximately so if ϵ is not a multiple of $1/n!$) when $F(x) \equiv F_0(x)$. Such a region will contain exactly the same probability ϵ for any other continuous cdf $F(x)$, and the test would have no *power* for testing the hypothesis that $F(x) \equiv F_0(x)$ over any of the alternatives $F(x) \neq F_0(x)$. In other words, the probability of the rejection of an alternative cdf will always be ϵ no matter which

alternative it is. The method of randomization for constructing statistical tests therefore must be applied to situations in which it is desired to test whether certain subsets of values in a sample come from populations having continuous cdf's differing in some respect.

The construction of rejection regions for nonparametric tests by the method of randomization has not yet been placed on a satisfactory general foundation as has been done in the case of parametric tests. The best way to discuss the method is perhaps to consider some examples. These examples will show how intuition plays an important role in each case and will emphasize the need of a general theory.

Before considering applications of the method of randomization to the construction of nonparametric statistical tests, it should be emphasized that the method of randomization has other interesting applications. It furnishes a way of dealing with any problem of determining probabilities of specified order relations among values in two or more samples which have been drawn from the same population having a continuous cdf.

For example we may ask what is the probability that a pair of samples O_m (first sample) and O_n (second sample) from a population with continuous cdf $F(x)$ will be such that all the values of O_n will fall within the range established by O_m ? What is the probability that the pair of samples will be such that the largest value in O_n will exceed the largest value in O_m ? Suppose the values in O_n are drawn one by one until a value exceeds the largest value in O_m . For pairs of samples drawn in this way what is the probability function of n , the size of the second sample which must be drawn to satisfy the conditions of the problem? There are many problems of this type which can be dealt with by the randomization method. They all reduce to combinatorial analysis of the set of equally probable permutations of the $m+n$ values of the two samples, enumerating the subset which satisfies the conditions of the problem. Similarly, one could deal with order relations problems for three or more samples. Also, one can raise analogous questions for the case of two or more samples from populations having continuous cdf's of two or more dimensions.

11. Examples of nonparametric statistical tests for one dimension based on the method of randomization.

(a) *Two-sample tests.* First, let us consider the two-sample problem treated originally in a very specific example by Fisher [13] and more generally treated by Pitman [55]. Let X_1, X_2, \dots, X_m and $X_{m+1}, X_{m+2}, \dots, X_{m+n}$ be two independent samples O_m and O_n of sizes m and n from populations having continuous cdf's $F(x)$ and $G(x)$ respectively. Let the values in O_m and O_n be pooled together into a

single sample Z_1, Z_2, \dots, Z_{m+n} of size $m+n$ and let the order statistics in the pooled sample be denoted by $Z_{(1)}, Z_{(2)}, \dots, Z_{(m+n)}$. If $F(x) \equiv G(x)$ then the probabilities associated with all permutations of the sample values are all equal to $1/(m+n)!$. Let a *separation* be a division of the pooled sample into two sets, consisting of m values and n values respectively. There are $C_{m+n, n}$ possible separations. For any particular separation let Z_1, Z_2, \dots, Z_m and Z'_1, Z'_2, \dots, Z'_n be the two sets of values. Let \bar{Z} and \bar{Z}' be the means of the sets. Let $S = \sum_{i=1}^m (Z_i - \bar{Z})^2$ and $S' = \sum_{j=1}^n (Z'_j - \bar{Z}')^2$. Pitman proposed

$$(59) \quad V = \frac{mn(\bar{Z} - \bar{Z}')^2}{(m+n)(S + S') + mn(\bar{Z} - \bar{Z}')^2}$$

as a criterion for testing the difference between the two means \bar{Z} and \bar{Z}' , large values of V being *significant*. He determined the first three moments of the distribution of V over the set of all separations (equally weighted) and for large m and n he showed the probability element of the cdf of V to be approximately

$$(60) \quad \frac{\Gamma((m+n+1)/2)}{\Gamma(1/2)\Gamma((m+n)/2)} v^{-1/2}(1-v)^{(m+n)/2-1} dv.$$

It is to be noted that the equal weighting of the separations comes from the fact that when $F(x) \equiv G(x)$ (that is, when the *null hypothesis* is true) the $(m+n)!$ regions $\{S_{\{\alpha_i\}}\}$ corresponding to the $(m+n)!$ permutations of the type $Z_{\alpha_1} < Z_{\alpha_2} < \dots < Z_{\alpha_{m+n}}$ all have probabilities equal to $1/(m+n)!$.

In this example it is to be noted that a rejection region W in the sample space R_n is constructed by taking those regions from the set $\{S_{\{\alpha_i\}}\}$ for which V has its largest values, and enough of them to produce a total probability as close as possible to ϵ , the significance level. More precisely if v_ϵ is the *critical* value of V at significance level ϵ , then W is the region in sample space constructed of sets from $\{S_{\{\alpha_i\}}\}$ for which $V \geq v_\epsilon$, where v_ϵ is the smallest number for which $\Pr(V \geq v_\epsilon) \leq \epsilon$, when $F(x) \equiv G(x)$.

If we are to subject the test criterion V and the rejection region W to a more penetrating analysis we must ask for what class Ω of pairs of continuous cdf's $\{F(x), G(x)\}$, including the null hypothesis $F(x) \equiv G(x)$, is the test (consisting of V and the rejection region W) "satisfactory"? This question has not been considered. Some might consider as a minimum requirement for the test to be satisfactory that it be *unbiased*, that is, that the probability (approximately equal to ϵ) associated with the rejection region W be smaller when $F(x) \equiv G(x)$ than for any other pair of cdf's in Ω . Actually, the question of

whether a test is satisfactory or not has to be determined from a consideration of the whole *power function* of the test, that is, the probability contained in W expressed as a function of the pair $\{F(x), G(x)\}$. If W and W' are two regions such that the power functions are equal when $F(x) \equiv G(x)$ and the power function of W' is greater than that for W for all pairs $F(x), G(x)$ which are different, then the test based on W' is "better" than that based on W .

The Pitman test is unbiased if $G(x) = F(x+m)$, that is, if Ω consists of all pairs of continuous *cdf's* of the form $\{F(x), F(x+m)\}$. In other words, the test would be satisfactory as a *slippage* or *location* test.

A second rather simple and elegant test for the problem of two samples has been proposed by Wald and Wolfowitz [77]. To describe their results we again consider two samples, O_m and O_n of m and n values from populations having continuous *cdf's* $F(x)$ and $G(x)$ respectively. There is no loss of generality in assuming $m \leq n$. As before, let the pooled sample be Z_1, Z_2, \dots, Z_{m+n} and the ordered values of the Z 's be $Z_{(1)}, Z_{(2)}, \dots, Z_{(m+n)}$. This sequence of order statistics will be a mixture of X 's from the two samples. We shall define a *run* of length r from O_m in this sequence as an uninterrupted subsequence of X 's from O_m , with a similar definition of a run of length s from O_n . The sequence of order statistics will then consist of a certain number of runs of X 's from O_m and a certain number from O_n . Let U be the *total* number of runs. Wald and Wolfowitz proposed U as the criterion for testing the hypothesis $F(x) \equiv G(x)$, where small values of U are significant. This is an intuitively reasonable test since small values of U correspond to bunching or poor mixing of values from the two samples, and hence furnish an observational basis for doubt of the hypothesis that $F(x) \equiv G(x)$. More precisely, the rejection region W in the sample space R_n consists of all regions from $\{S_{\{\alpha_i\}}\}$ for which $U \leq u_\epsilon$, where u_ϵ is the smallest integer for which $\Pr(U \leq u_\epsilon) \geq \epsilon$ when $F(x) \equiv G(x)$. The problem of determining u_ϵ is solved, of course, if the exact distribution of U is found when $F(x) \equiv G(x)$. This reduces to a completely combinatorial problem. The result obtained by Wald and Wolfowitz is

$$\Pr(U = 2k) = \frac{2C_{m-1, k-1}C_{n-1, k-1}}{C_{m+n, m}} \quad (k = 1, 2, \dots, m),$$

$$\Pr(U = 2k - 1) = \frac{C_{m-1, k-1}C_{n-1, k-2} + C_{m-1, k-2}C_{n-1, k-1}}{C_{m+n, m}} \quad (k = 2, 3, \dots, m + 1).$$

Swed and Eisenhart [69] have tabulated $\Pr(U \leq u')$ for $m \leq n \leq 20$, and for all values of u' in each case. They have also tabulated the smallest integer u_ϵ for which $\Pr(U \leq u_\epsilon) \geq \epsilon$ for $\epsilon = .005, .01, .025, .05$, and the largest integer u_ϵ for which $\Pr(U \leq u_\epsilon) \leq \epsilon$ for $\epsilon = .95, .975, .99$ and $.995$.

An important property of the U test established by Wald and Wolfowitz for certain regularity conditions on $F(x)$ and $G(x)$ is that the test is *consistent* for all pairs of cdf's $\{F(x), G(x)\}$ in the limit as m and $n \rightarrow \infty$ so that $m/n = r \neq 0$. This means that the probability of a sample point falling in the rejection region W in sample space when $F(x) \neq G(x)$ approaches 1 as m and $n \rightarrow \infty$ in a constant ratio $m/n = r \neq 0$.

The results of Wald and Wolfowitz can be readily extended to the case of three or more samples. Such an extension would reduce to the theory of runs for three or more kinds of elements as developed by Mood [36].

Other simple tests of the two-sample problem using order statistics have been devised by Thompson [71], Dixon [9], and Mathisen [32]. The tests proposed by Thompson and Mathisen are not consistent for hypotheses involving all pairs of continuous cdf's $\{F(x), G(x)\}$, although it can be conjectured that they are consistent for the important case of a slippage test of unimodal cdf's, that is, for all pairs $\{F(x), F(x+m)\}$, when $F'(x)$ exists and has only one maximum.

(b) *Tests of independence or "randomness" in ordered sequences; run tests.* It was pointed out earlier that no satisfactory statistical test based on order relations among the values in a sample could be devised for testing the hypothesis that $F(x)$ has a specified form $F_0(x)$. This does not mean that there are not important hypotheses which can be tested by using order relations among the values in a single sample.

One of the most important problems in this category is the question of whether $F(x)$ is changing from drawing to drawing, or whether the sample values, X_1, X_2, \dots, X_n , in the ordered sequence as drawn are "random." In this case we would consider the class Ω of all n -dimensional continuous cdf's $F(x_1, x_2, \dots, x_n)$, or some subset of Ω , and the null hypothesis would state that

$$(62) \quad F(x_1, x_2, \dots, x_n) \equiv F(x_1) \cdot F(x_2) \cdot \dots \cdot F(x_n).$$

This hypothesis of *independence* or "randomness" is basic to the whole theory of random sampling. The practical importance of this problem as one to be investigated before applying random sampling theory has been strongly emphasized by Shewhart [63] on the basis of his

experience with sampling in industry. In dealing with this problem various tests of independence based on order statistics have been proposed. Several of them, now referred to as *run tests* and based on intuition have been studied. We shall briefly describe two of these tests: (i) *runs above and below the median*, and (ii) *runs up and down*.

First let us examine (i). Consider a sample of values $X_1, X_2, \dots, X_{2n+1}$ drawn in the order indicated. Note that these are *not* order statistics. In this sequence let each X less than the median, $X_{(n+1)}$, be denoted by a , and let each X greater than $X_{(n+1)}$ be denoted by b . Then deleting the median from $X_1, X_2, \dots, X_{2n+1}$ we have $2n$ X 's left which are now replaced by some arrangement of n a 's and n b 's. There will be r_{1i} runs of a 's of length i , and r_{2i} runs of b 's of length i , $i=1, 2, \dots, n$. Under the hypothesis of independence, the probability theory of r_{1i} and r_{2i} reduces to a consideration of the set of $(2n)!$ permutations of the n a 's and n b 's, all of which have equal probability $1/(2n)!$. The median is ignored as far as determining runs is concerned. In case of a sample of $2n$ values one can select any number between $X_{(n)}$ and $X_{(n+1)}$ as the value to separate X 's into a 's and b 's, and the run theory is the same as that for the case just considered. Thus, the problem of finding the distribution function of the r_{1i} and r_{2i} is completely combinatorial, and has been solved by Mood [36]. He has also solved the analogous distribution problem for more than two kinds of elements, such as would arise if one used several arbitrary order statistics to cut up the set of sample values into more than two sets of elements.

Mosteller [38], using Mood's basic distribution theory of runs, has considered the length L of the longest run of a 's as a criterion for testing randomness, large values of L being significant. The critical value of L for probability level ϵ is the smallest integer l_ϵ such that $\Pr(L \geq l_\epsilon) \leq \epsilon$. A similar criterion exists for the b 's. He also considered a criterion L' defined as the length of the longest run of a 's or b 's. He tabulated critical values of l_ϵ and l'_ϵ for $2n=10, 20, 30, 40, 50$ and for $\epsilon=.01$ and $.05$.

It should be noted that the Wald-Wolfowitz U test may also be considered as a test of type (i) for testing the hypothesis of independence. In this case U would be the total number of runs of a 's and b 's.

Now let us consider a run test of type (ii) treated by Wolfowitz and Levene [89] and by Olmstead [44]. In this case let the sample values be X_1, X_2, \dots, X_n in the order drawn. We now set up a sequence of $n-1$ $+$'s and $-$'s defined as follows: If $X_i < X_{i+1}$ we write down a $+$, if $X_i > X_{i+1}$ we write down a $-$, $i=1, 2, \dots, n-1$. We

now consider runs of +’s and -’s. The test criterion L'' used in this test is the length of the longest run of +’s or -’s, large values of L'' being significant. The probability theory of L'' is much more complicated than that of L or L' , and was worked out by Levene and Wolfowitz. Olmstead [44], making use of a recursion relation, has been able to tabulate the exact probability function of L'' for sample sizes ranging from 2 to 14 and has been able to find the approximate distribution for larger values of n . A criterion for testing independence in an ordered sequence proposed by Young [90] and based on the method of randomization is given by

$$(63) \quad C = 1 - \frac{\sum_{i=1}^{n-1} (X_i - X_{i+1})^2}{2 \sum_{i=1}^n (X_i - \bar{X})^2}.$$

Young found the first four moments of the distribution of C assuming independence as expressed by (62).

Wald and Wolfowitz [79] have considered a criterion similar to that studied by Young. Their criterion is defined as

$$(64) \quad D = \sum_{i=1}^{n-1} X_i X_{i+1} + X_1 X_n$$

and was proposed as a test for randomness, the null hypothesis being given by (62). The method of randomization was used in dealing with the distribution of D . The mean and variance of D was found when the null hypothesis stated by (62) is true and it was shown that D is asymptotically normally distributed for large n in this case.

It should be noted that these run tests have been devised from intuitive considerations for the purposes of detecting the presence of “causes” of slowly changing “secular changes” or slippage in the population cdf during the course of drawing a sample. The probability theory of all of these run tests is based on the assumption that the null hypothesis of independence expressed by (62) holds. It is likely that these run tests are satisfactory for alternatives to the hypothesis of random sampling in which the population cdf changes by slippage during the course of sampling, but this is yet to be explored.

The problem of testing for independence or for “randomness” in successive drawings from a population, which is fundamental in the theory of random sampling, deserves more theoretical attention than it has received. Possibly a consideration of the problem from the point of view of Wald’s theory of sequential analysis would be profitable.

(c) *One-dimensional parametric tests involving order statistics.* A number of authors have investigated parametric statistical tests based on order statistics for samples from a normal population. E. S. Pearson and Hartley [50] have studied the "Studentized" range

$$(65) \quad q = \frac{(m-1)^{1/2}(X_{(n)} - X_{(1)})}{\left[\sum_{i=1}^m (X'_i - \bar{X}')^2 \right]^{1/2}}$$

determined by two samples X_1, X_2, \dots, X_n and X'_1, X'_2, \dots, X'_m from a normal population having mean m and variance σ^2 , where \bar{X}' is the mean of the second sample. They have tabulated the 1% and 5% points of the distribution of q for all values of n from 3 to 20, and for $m = 11, 12, \dots, 20, 24, 30, 40, 60,$ and 120. The $100\alpha\%$ point in this case is defined as the value of q_α for which $\Pr|q| > q_\alpha = \alpha$ when both samples are from the same normal population.

Daly [6] has considered the quantity

$$(66) \quad G = \frac{\bar{X} - m}{X_{(n)} - X_{(1)}}$$

as an alternative to the Student t -test for testing the hypothesis that the population mean m in a normal population has a specified value m_0 . He has found that for sample sizes less than 10, a test based on G is almost as powerful as the Student t -test. Walsh [82] has considered an even simpler alternative to the t -test for small samples from a normal population, namely,

$$(67) \quad H = \frac{(X_{(n)} + X_{(1)})/2 - m}{X_{(n)} - X_{(1)}}.$$

This test is almost as powerful as the Student t -test for values of n less than 10.

Walsh, in a paper not yet published, has devised very simple tests based on order statistics for testing the hypothesis that the median of a continuous cdf has a specified value. In the case of a normal population he has found the *power efficiency* of his tests to be at least 90%, when compared with the Student t -test, for samples of less than 15. The power efficiency of $E\%$ as used by Walsh means that the power function of the test for a sample of size n is approximately the same as that of a Student t -test for a sample of size $En/100$.

(d) *Analysis of variance tests by the method of randomization.* Among other one-dimensional nonparametric tests should be mentioned

analysis of variance tests. In 1937 Friedman [16] devised an analysis of variance test based on ranks. In his test rs "observations" X_{ij} ($i=1, 2, \dots, r; j=1, 2, \dots, s$) are arranged in r rows and s columns. The sample values within each row are ordered and only the *ranks* (1, 2, \dots , s) of the ordered values in each row are used in the test. The hypothesis tested (null hypothesis) is that there are no column "effects," that is, that the sample values within each row are from a single population with some continuous cdf. This means that all possible permutations of ranks within a row are equally probable. If α_{ij} is the rank associated with the i th row and j th column, where $\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{is}$ ($i=1, 2, \dots, r$) is some permutation of the integers 1, 2, \dots , s , the test criterion proposed by Friedman is

$$(68) \quad \chi_r^2 = \frac{12}{rs(s+1)} \sum_{j=1}^s \left(\sum_{i=1}^r \alpha_{ij} \right)^2 - 3r(s+1).$$

If there are no column "effects," the quantity χ_r^2 is, for large r , approximately distributed according to the chi-square law with $s-1$ degrees of freedom, that is, with probability element

$$(69) \quad \frac{(\chi_r^2/2)^{(s-3)/2}}{2\Gamma((s-1)/2)} e^{-\chi_r^2/2} d(\chi_r^2).$$

The test is a randomization test in which the null hypothesis is that the s values X_{ij} ($j=1, 2, \dots, s$) in the i th row ($i=1, 2, \dots, r$) is a sample from a population having some continuous cdf. We therefore have a sample of size s from each of r populations. Kendall and Smith [29] and also Wallis [80] have studied the same problem but used a criterion η_r^2 related to χ_r^2 by the expression $\chi_r^2 = r(s-1)\eta_r^2$.

In 1937 Pitman [57] investigated the standard analysis of variance test for testing column "effects" in a layout of r rows and s columns under the assumption that the observations in each row constitute a sample of size s from a population having a continuous cdf, the cdf's for the rows being identically the same except for slippage. Let X_{ij} be the observation for the i th row and j th column. Consider the ratio

$$(70) \quad W = \frac{\sum_{i,j} (\bar{X}_{.j} - \bar{X})^2}{\sum_{i,j} (X_{ij} - \bar{X}_{.j} - \bar{X}_{i.} + \bar{X})^2 + \sum_{i,j} (\bar{X}_{.j} - \bar{X})^2}$$

where $\sum_{i,j}$ denotes summation from $i=1$ to r and $j=1$ to s , $\bar{X}_{i.}$, $\bar{X}_{.j}$ are the averages of the X 's in the i th row and j th columns respec-

tively, and \bar{X} is the average of all r s X 's. Pitman, using the method of randomization, determined the first four moments of the distribution of W under the set of all permutations of the X 's within rows, and he presented some evidence to the effect that the distribution of W could be reasonably well fitted by the beta distribution

$$(71) \quad \frac{\Gamma(r(s-1)/2)}{\Gamma((s-1)/2)\Gamma((r-1)(s-1)/2)} w^{(s-3)/2}(1-w)^{(r-1)(s-1)/2-1} dw,$$

which is the same distribution as that obtained under the usual assumption of normality of the analysis of variance test. In 1938 Welch [84] made a study of the standard correlation ratio test for homogeneity from the point of view of the method of randomization. This problem is very similar to the analysis of variance problem treated by Pitman. Welch [83] also carried out an investigation of the analysis of variance test for randomized blocks and Latin squares from the point of view of the method of randomization.

All of these analysis of variance tests are essentially extensions of the usual analysis of variance test to the case of more general populations. One type of nonparametric test which needs consideration is a k -sample slippage or location test for testing the hypothesis that a sample which has the largest "observed slippage" to the right, let us say, as measured in some sense by order statistics, does, in fact, come from a population centered farther to the right than the populations from which the other samples come. Here it would be assumed that the cdf's of the k populations are of the form $F(x+m_i)$ ($i=1, 2, \dots, k$), and the null-hypothesis would be that the m_i are all equal. Mosteller [40] has proposed one simple test for this situation in the case of k samples, each of size n . To define his test we select from among the k samples that sample having the largest X . Let r be the number of X 's in this sample which exceed the largest X in any of the remaining samples; r is the test criterion. The problem of determining the probability function of r under the null hypothesis (that is, when all m_i are equal) is an application of the method of randomization—it is completely combinatorial, and the probability that $r=r'$ under these conditions is

$$(72) \quad \Pr(r=r') = \frac{k(n!)(kn-r')!}{(kn)!(n-r')!},$$

which is approximately $k^{1-r'}$ for large n . The important feature of this test is that one allows himself to select the sample with the largest "observed slippage" and then to ask whether there are enough values in it exceeding the largest value in the remaining samples to

be "significant." A full analysis of this test—like other randomization tests—requires a study of its power function for all sets of cdf's of the form $\{F(x+m_1), F(x+m_2), \dots, F(x+m_k)\}$. Mosteller has obtained a few inequalities on the power of the test.

12. Nonparametric tests for two or more dimensions based on the method of randomization. The principle of randomization can be extended to populations involving two or more variables. We shall not attempt to discuss the general problem here. Several examples of tests for the two-dimensional case will indicate the nature of such tests.

(a) *Tests of independence based on correlation coefficients.* Pitman [56], by using the principle of randomization, for two dimensions, developed a test for the independence of the two variables in a bivariate population having a continuous cdf $F(x, y)$. In this case the sample O_n is $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$. The hypothesis of independence is that $F(x, y) = F_1(x) \cdot F_2(y)$. If the hypothesis is true, one essentially has two samples: namely, X_1, X_2, \dots, X_n and Y_1, Y_2, \dots, Y_n , in which the pairing of values has occurred "at random." In the X sample space there are $n!$ regions of the type $X_{\alpha_1} < X_{\alpha_2} < \dots < X_{\alpha_n}$ where $\alpha_1, \alpha_2, \dots, \alpha_n$ is a permutation of the integers $1, 2, \dots, n$, and similarly for the Y sample space. Therefore in the product space of the two samples there are $(n!)^2$ product regions in each of which an X inequality $X_{\alpha_1} < X_{\alpha_2} < \dots < X_{\alpha_n}$ and a Y inequality $Y_{\beta_1} < Y_{\beta_2} < \dots < Y_{\beta_n}$ hold simultaneously. The probability contained in each of these product regions is $(1/n!)^2$ when the hypothesis of independence is true. This means that the probability associated with any particular matching $(X_{\alpha_1}, Y_{\beta_1}), (X_{\alpha_2}, Y_{\beta_2}), \dots, (X_{\alpha_n}, Y_{\beta_n})$ is $1/(n!)^2$. Pitman has considered the distribution of the ordinary correlation coefficient

$$(73) \quad R = \frac{\sum_{i=1}^n (X_{\alpha_i} - \bar{X})(Y_{\beta_i} - \bar{Y})}{\left(\sum_{i=1}^n (X_{\alpha_i} - \bar{X})^2 \cdot \sum_{i=1}^n (Y_{\beta_i} - \bar{Y})^2 \right)^{1/2}}$$

over the $(n!)^2$ matchings of X 's and Y 's, where \bar{X} and \bar{Y} are the means of the X 's and Y 's respectively. He obtained the first four moments of R and for large n fitted a beta distribution to that of R , obtaining as the pdf of the fitted distribution

$$(74) \quad \frac{\Gamma(n/2 - 1/2)}{\Gamma(1/2)\Gamma(n/2 - 1)} (1 - t^2)^{n/2-2} dt,$$

for the interval $(-1, +1)$.

In 1936, Hotelling and Pabst [26], using the method of randomization, investigated the rank correlation coefficient as a criterion for testing independence of X and Y , defined by

$$(75) \quad R' = \frac{\sum_{i=1}^n (\alpha_i - (n+1)/2)(\beta_i - (n+1)/2)}{\left(\sum_{i=1}^n (\alpha_i - (n+1)/2)^2 \cdot \sum_{i=1}^n (\beta_i - (n+1)/2)^2 \right)^{1/2}}$$

which has the same approximate distribution as R' , namely (74). Each of the quantities $n^{1/2}R$ and $n^{1/2}R'$ is asymptotically normally distributed with zero mean and unit variance for large n . Olds [43] has tabulated the distribution of a quantity S related to R' by the relation $R' = 1 - 6S/(n^3 - n)$ for values of n from 2 to 10.

(b) A "corner" test of association. Olmstead and Tukey [45] have recently devised a simple test for the independence of two random variables on the basis of a sample of size $2n+1$. Their test is as follows: consider the sample $(X_1, Y_1), (X_2, Y_2), \dots, (X_{2n+1}, Y_{2n+1})$ as represented by $2n+1$ points in the xy -plane. Then draw the horizontal line $y = Y_{(n+1)}$ and the vertical line $x = X_{(n+1)}$. Designate the quadrants as $+, -, +, -$, beginning in the upper right-hand quadrant and moving counterclockwise. Now begin with the sample point farthest to the right and move toward the left as long as sample points are on the same side of the horizontal median line as the original one encountered. Let this number of points be P_1 , attaching a $+$ or $-$ sign depending on whether the points are in a $+$ or $-$ quadrant. Do a similar counting operation moving from the top sample point downward, a similar one moving from the left-most to the right and a similar one moving from the bottom sample point upward. Let the numbers of points be P_2, P_3 and P_4 , respectively, with attached $+$ or $-$ signs depending on whether the points in each case fall in a $+$ or $-$ quadrant. Let S be the algebraic sum of the four P 's with their signs attached. Then if X and Y are independent, S is a function of only the order relations among the X 's and Y 's and the problem of determining the distribution function of S is therefore combinatorial. Olmstead and Tukey have shown that, for large n ,

$$(76) \quad \lim_{n \rightarrow \infty} \Pr(|S| \geq k) = \frac{9k^3 + 9k^2 + 168k + 208}{216(2)^k}.$$

They have also tabulated exact values of $\Pr(|S| \geq k)$ for $2n = 2, 4, 6, 8, 10, 14$, and for $k = 1, 2, 3, \dots, 30$. An important feature of this

test is that $\Pr(|S| \geq k)$ approaches its limiting form very rapidly.

This test has also been extended to the case of more than two dimensions.

REFERENCES AND LITERATURE

1. L. von Bortkiewicz, *Variationsbreite und mittlerer Fehler*, Berlin Math. Ges. Sitzungsber. vol. 21 (1921) pp. 3–11.
2. ———, *Die Variationsbreite beim Gauss'schen Fehlergesetz*, Nordisk Statistisk Tidsskrift vol. 1 (1922) pp. 11–38, 193–220.
3. A. G. Carlton, *Estimating the parameters of a rectangular distribution*, Ann. Math. Stat. vol. 17 (1946) pp. 355–358.
4. Allen T. Craig, *On the distributions of certain statistics*, Amer. J. Math. vol. 54 (1932) pp. 353–366.
5. Harald Cramér, *Mathematical methods of statistics*, Princeton University Press, 1946.
6. J. F. Daly, *On the use of the sample range in an analogue of Student's t-test*, Ann. Math. Stat., vol. 17 (1946) pp. 71–74.
7. H. E. Daniels, *The statistical theory of the strength of bundles of threads*, Proc. Roy. Soc. London Ser. A. vol. 183 (1945) pp. 405–435.
8. O. L. Davies and E. S. Pearson, *Methods of estimating from samples the population standard deviations*, Journal of the Royal Statistical Society Supplement vol. 1 (1934) pp. 76–93.
9. W. J. Dixon, *A criterion for testing the hypothesis that two samples are from the same population*, Ann. Math. Stat. vol. 11 (1940) pp. 199–204.
10. Edward L. Dodd, *The greatest and the least variate under general laws of error*, Trans. Amer. Math. Soc. vol. 25 (1923) p. 525–539.
11. G. Elfving, *The asymptotical distribution of range in samples from a normal population*, Biometrika vol. 34 (1947) pp. 111–119.
12. R. A. Fisher, *On the random sequence*, Quarterly Journal of the Royal Meteorological Society vol. 52 (1926) p. 250.
13. ———, *Coefficient of racial likeness and the future of craniometry*, Journal of the Royal Anthropological Institute vol. 66 (1936) pp. 57–63.
14. R. A. Fisher and L. H. C. Tippett, *Limiting forms of the frequency distribution of the largest or smallest member of a sample*, Proc. Cambridge Philos. Soc. vol. 24 (1298) pp. 180–190.
15. Maurice Fréchet, *Sur la loi de probabilité de l'écart maximum*, Annales de la Société Polonaise de Mathématique vol. 6 (1927) (printed in 1928) pp. 92–116.
16. M. Friedman, *The use of ranks to avoid the assumption of normality*, Journal of the American Statistical Association vol. 32 (1937) pp. 675–701.
17. E. J. Gumbel, *Les valeurs extrêmes des distributions statistiques*, Annales de l'Institut Henri Poincaré vol. 4 (1935) pp. 115–158.
18. ———, *Les intervalles extrêmes entre les émissions radioactives*, Journal de Physique (7) vol. 8 (1937) pp. 321–329, 446–452.
19. ———, *The return period of flood flows*, Ann. Math. Stat., vol. 12 (1941) pp. 163–190.
20. ———, *Ranges and midranges*, Ann. Math. Stat. vol. 15 (1944) pp. 414–422.
21. ———, *On the independence of the extremes in a sample*, Ann. Math. Stat. vol. 17 (1946) pp. 78–81.
22. ———, *The distribution of the range*, Ann. Math. Stat. vol. 18 (1947) pp. 384–412.

23. H. O. Hartley, *The probability integral of the range in samples of n observations from a normal population*, *Biometrika* vol. 32 (1942) pp. 301–308.
24. ———, *The range in random samples*, *Biometrika* vol. 32 (1942) pp. 334–348.
25. Cecil Hastings, Jr., Frederick Mosteller, John W. Tukey, and Charles P. Winsor, *Low moments for small samples: A comparative study of order statistics*, *Ann. Math. Stat.* vol. 18 (1947) pp. 413–426.
26. H. Hotelling and M. R. Pabst, *Rank correlation and tests of significance involving no assumptions of normality*, *Ann. Math. Stat.* vol. 7 (1936) pp. 29–43.
27. J. O. Irwin, *The further theory of Francis Galton's individual difference problem*, *Biometrika* vol. 17 (1925) pp. 100–128.
28. M. G. Kendall, *A new measure of rank correlation*, *Biometrika* vol. 30 (1938) pp. 81–93.
29. M. G. Kendall and B. B. Smith, *The problem of m rankings*, *Ann. Math. Stat.* vol. 10 (1939) pp. 275–287.
30. A. Kolmogoroff, *Sulla determinazione empirica di una legge di distribuzione*, *Giornale dell' Istituto Italiano degli Attuari* vol. 4 (1933) pp. 83–91.
31. ———, *Confidence limits for an unknown distribution function*, *Ann. Math. Stat.* vol. 12 (1941) pp. 461–463.
32. Harold C. Mathisen, *A method of testing the hypothesis that two samples are from the same population*, *Ann. Math. Stat.* vol. 14 (1943) pp. 188–194.
33. A. T. McKay and E. S. Pearson, *A note on the distribution of range in sample sizes of n* , *Biometrika* vol. 25 (1933) pp. 415–420.
34. A. T. McKay, *Distribution of the difference between the extreme observations and the sample mean in samples of n from a normal universe*, *Biometrika* vol. 27 (1935) pp. 466–471.
35. R. von Mises, *La distribution de la plus grande de n valeurs*, *Revue de l'Union Interbalkanique* vol. 1 (1936) pp. 1–20.
36. A. M. Mood, *The distribution theory of runs*, *Ann. Math. Stat.* vol. 11 (1940) pp. 367–392.
37. ———, *On the joint distribution of the medians in samples from a multivariate population*, *Ann. Math. Stat.* vol. 12 (1941) pp. 268–278.
38. Frederick Mosteller, *Note on application of runs to quality control charts*, *Ann. Math. Stat.* vol. 12 (1941) pp. 228–232.
39. ———, *On some useful "inefficient" statistics*, *Ann. Math. Stat.* vol. 17 (1946) pp. 377–408.
40. ———, *A k -sample slippage test for an extreme population*, to appear in *Ann. Math. Stat.* vol. 19 (1948).
41. K. R. Nair, *The median in tests by randomization*, *Sankhyā* vol. 4 (1940) pp. 543–550.
42. ———, *Table of confidence intervals for the median in samples from any continuous population*, *Sankhyā* vol. 4 (1940) pp. 551–558.
43. E. G. Olds, *Distribution of sums of squares of rank differences for small numbers of individuals*, *Ann. Math. Stat.* vol. 9 (1938) pp. 133–148.
44. P. S. Olmstead, *Distribution of sample arrangements for runs up and down*, *Ann. Math. Stat.* vol. 17 (1946) pp. 24–33.
45. P. S. Olmstead and John W. Tukey, *A corner test for association*, *Ann. Math. Stat.* vol. 18 (1947) pp. 495–513.
46. Edward Paulson, *A note on tolerance limits*, *Ann. Math. Stat.* vol. 14 (1943) pp. 90–93.
47. E. S. Pearson, *Some aspects of the problem of randomization*, *Biometrika* vol. 29 (1937) pp. 53–64; vol. 30 (1938) pp. 159–179.

48. ———, *The probability integral of the range in samples of n observations from a normal population*, *Biometrika* vol. 32 (1942) pp. 301–308.
49. E. S. Pearson and H. O. Hartley, *The probability integral of the range in samples of n observations from a normal population*, *Biometrika* vol. 32 (1942) pp. 301–310.
50. ———, *Tables of the probability integral of the "Studentized" range*, *Biometrika* vol. 33 (1943) pp. 89–99.
51. Karl Pearson, *Note on Francis Galton's problem*, *Biometrika* vol. 1 (1902) pp. 390–399.
52. ———, *On further methods of determining correlation*, *Drapers' Company Research Memoirs*, *Biometric Series IV*, Galton Laboratory, London, 1907.
53. ———, *On the probable errors of frequency constants*, *Biometrika* vol. 13 (1920) pp. 113–132.
54. ———, *Tables of the incomplete beta function*, Cambridge University Press, 1932.
55. E. J. G. Pitman, *Significance tests which may be applied to samples from any populations*, *Journal of the Royal Statistical Society Supplement* vol. 4 (1937) pp. 119–130.
56. ———, *Significance tests which may be applied to samples from any populations*, II. *The correlation coefficient test*, *Journal of the Royal Statistical Society Supplement* vol. 4 (1937) pp. 225–232.
57. ———, *Significance tests which may be applied to samples from any populations*, III. *The analysis of variance test*, *Biometrika* vol. 29 (1938) pp. 322–335.
58. Herbert Robbins, *On distribution-free tolerance limits in random sampling*, *Ann. Math. Stat.* vol. 15 (1944) pp. 214–216.
59. S. R. Savur, *The use of the median in tests of significance*, *Proceedings of the Indian Academy of Sciences (Section A)* vol. 5 (1937) pp. 564–576.
60. Henry Scheffé, *Statistical inference in the non-parametric case*, *Ann. Math. Stat.* vol. 14 (1943) pp. 305–332.
61. Henry Scheffé and John W. Tukey, *A formula for sample sizes for population tolerance limits*, *Ann. Math. Stat.* vol. 15 (1944) pp. 217.
62. ———, *Non-parametric estimation*, I. *Validation of order statistics*, *Ann. Math. Stat.* vol. 16 (1945) pp. 187–192.
63. W. A. Shewhart, *Statistical method from the viewpoint of quality control*, U. S. Department of Agriculture, Washington, 1939.
64. N. Smirnov, *Sur la dépendance des membres d'un série de variations*, *Bull. Math. Univ. Moscow* vol. 1 (1937–1938) pp. 1–12.
65. ———, *Über die verteilung des allgemeinen Gliedes in der variationsreche*, *Metron* vol. 12 (1935) pp. 59–81.
66. ———, *On the estimation of the discrepancy between empirical curves of distribution for two independent samples*, *Bull. Math. Univ. Moscow* vol. 2 (1939).
67. ———, *Sur les écarts de la courbe de distribution empirique*, *Rec. Math. (Mat. Sbornik)* N.S. vol. 6 (1939) pp. 3–26.
68. W. L. Stevens, *Distribution of groups in a sequence of alternatives*, *Annals of Eugenics* vol. 9 (1939) pp. 10–17.
69. Frieda S. Swed and Churchill Eisenhart, *Tables for testing randomness of grouping in a sequence of alternatives*, *Ann. Math. Stat.* vol. 14 (1943) pp. 66–87.
70. William R. Thompson, *On confidence ranges for the median and other expectation distributions for populations of unknown distribution form*, *Ann. Math. Stat.* vol. 7 (1936) pp. 122–128.
71. ———, *Biological applications of normal and range associated significance*

- tests in ignorance of original distribution forms, *Ann. Math. Stat.* vol. 9 (1938) pp. 281–287.
72. L. H. C. Tippett, *On the extreme individuals and the range of samples taken from a normal population*, *Biometrika* vol. 17 (1925) pp. 264–387.
73. John W. Tukey, *Non-parametric estimation*, II. *Statistically equivalent blocks and tolerance regions—the continuous case*, *Ann. Math. Stat.* vol. 18 (1947) pp. 529–539.
74. A. Wald, *Setting of tolerance limits when the sample is large*, *Ann. Math. Stat.* vol. 13 (1942) pp. 389–399.
75. ———, *An extension of Wilks' method for setting tolerance limits*, *Ann. Math. Stat.* vol. 14 (1943) pp. 45–55.
76. A. Wald and J. Wolfowitz, *Confidence limits for continuous distribution functions*, *Ann. Math. Stat.* vol. 10 (1939) pp. 105–118.
77. ———, *On a test whether two samples are from the same population*, *Ann. Math. Stat.* vol. 11 (1940) pp. 147–162.
78. ———, *Note on confidence limits for continuous distribution functions*, *Ann. Math. Stat.* vol. 12 (1941) pp. 118–119.
79. ———, *An exact test for randomness in the non-parametric case based on serial correlation*, *Ann. Math. Stat.* vol. 14 (1943) pp. 378–388.
80. W. A. Wallis, *The correlation ratio for ranked data*, *Journal of the American Statistical Association* vol. 34 (1939) pp. 533–538.
81. John E. Walsh, *Some significance tests based on order statistics*, *Ann. Math. Stat.* vol. 17 (1946) pp. 44–52.
82. ———, *Some significance tests for the mean using the sample range and mid-range*, to appear in *Ann. Math. Stat.* vol. 19 (1948).
83. B. L. Welch, *On the z-test in randomized blocks and Latin squares*, *Biometrika* vol. 29 (1937) pp. 21–52.
84. ———, *On tests of homogeneity*, *Biometrika* vol. 30 (1938) pp. 149–158.
85. S. S. Wilks, *On the determination of sample sizes for setting tolerance limits*, *Ann. Math. Stat.* vol. 12 (1941) pp. 91–96.
86. ———, *Statistical prediction with special reference to the problem of tolerance limits*, *Ann. Math. Stat.* vol. 13 (1942) pp. 400–409.
87. ———, *Mathematical statistics*, Princeton University Press, 1943.
88. J. Wolfowitz, *Additive partition functions and a class of statistical hypotheses*, *Ann. Math. Stat.* vol. 13 (1942) pp. 247–279.
89. J. Wolfowitz and H. Levene, *The covariance matrix of runs up and down*, *Ann. Math. Stat.* vol. 15 (1944) pp. 58–69.
90. L. C. Young, *On randomness in ordered sequences*, *Ann. Math. Stat.* vol. 12 (1941) pp. 293–300.