ADDITIVE PARTITION FUNCTIONS AND A CLASS OF STATISTICAL HYPOTHESES

By J. Wolfowitz

New York City

1. Introduction. The purpose of the first part of this paper is to prove several theorems about a class of functions of partitions which are additive in structure and subject to mild restrictions. These theorems may be regarded as contributions to the theory of numbers, but if one makes certain assignments of probabilities to the partitions the theorems may be expressed as statements about asymptotic distributions. It is in this latter, probabilistic language, that we shall carry out the proofs, for the following reasons. The discussion will be more concise and certain circumlocutions will be avoided. The theorems have statistical application and a number of theorems discussed recently in statistical literature are corollaries of one of our theorems.

In the second part of this paper the theory of testing statistical hypotheses where the form of the distribution functions is totally unknown and only continuity is assumed, will be discussed. The exact extension of the likelihood ratio criterion to this case will be given. Approximations to the application of this criterion in two problems will be proposed, one of which applies the results mentioned above. Lastly, in connection with the second problem, a combinatorial problem will be solved which is new and has interest per se.

2. Partitions of a single integer. Let n be a positive integer and $A = (a_1, a_2, \cdots, a_s)$ be any sequence of positive integers a_i $(i = 1, 2, \cdots, s)$, where $\sum_{i=1}^{s} a_i = n$, and s may be any integer from 1 to n. Two sequences A which have different elements or the same elements arranged in different order are to be considered distinct, so it is easy to see that there are 2^{n-1} sequences A. We shall consider the sequence A as a stochastic variable and assign to all sequences A the same probability, which is therefore 2^{-n+1}. Let r_j be the number of elements a in A which equal j $(j = 1, 2, \cdots, n)$, so that r_j is a stochastic variable. Let k be an integer $\leq n$. Then the joint distribution of the stochastic variables $r_1, r_2, \cdots, r_k$ is given as follows: The probability that $r_i = b_i$ $(i = 1, 2, \cdots, k)$ is

$$2^{-n+1}\left(\sum_{r=1}^{n} \sum \frac{r!}{(b_1)!(b_2)! \cdots (b_k)!(r_{(k+1)})! \cdots (r_n)!}\right), \tag{2.1}$$

where the inner summation is carried out over all sets of non-negative integers $r_{(k+1)}, \cdots, r_n$ such that

$$b_1 + b_2 + \cdots + b_k + r_{(k+1)} + \cdots + r_n = r, \tag{2.2}$$

$$b_1 + 2b_2 + \cdots + kb_k + (k+1)r_{(k+1)} + \cdots + nr_n = n. \tag{2.3}$$

(The b_i, of course, are non-negative integers.)

Let $r = \sum_{i=1}^{n} r_i$, and

$$r'_{(k+1)} = \sum_{i=k+1}^{n} r_i, \qquad (k < n),$$

so that r and r'_{k+1} are both stochastic variables. The probability that at the same time

$$r_i = b_i, \qquad (i = 1, \cdots, k), \tag{2.4}$$

and

$$r'_{(k+1)} = b'_{(k+1)}, \tag{2.5}$$

is given by (2.1) with the restriction

$$r_{(k+1)} + \cdots + r_n = b'_{(k+1)}, \tag{2.6}$$

added to the restrictions (2.2) and (2.3). With this added restriction the summation in (2.1) may be performed as follows: Let $t = \sum_{i=1}^{k} ib_i$. It is easy to see that the number of sequences A where every $a_i > k$, $r = r'_{(k+1)} = b'_{(k+1)}$, and $\Sigma a_i = n - t$, is given by the coefficient of x^{n-t} in the purely formal expansion in x of

$$(x^{k+1} + x^{k+2} + x^{k+3} + \cdots)^{b'_{(k+1)}} = x^{(k+1)b'_{(k+1)}}\left(\frac{1}{1-x}\right)^{b'_{k+1)}},$$

and is

$$\binom{n - t - kb'_{(k+1)} - 1}{b'_{(k+1)} - 1}.$$

Hence $P\{(2.4) \text{ and } (2.5)\}$, where this symbol will always denote the probability of the relation in braces, is seen to be

$$2^{-n+1} \frac{\left(\sum_{i=1}^{k} b_i + b'_{(k+1)}\right)!}{(b'_{(k+1)})! \prod_{i=1}^{k} (b_i)!} \binom{n - t - kb'_{(k+1)} - 1}{b'_{(k+1)} - 1}. \tag{2.7}$$

If X is a stochastic variable, let $E(X)$ and $\sigma^2(X)$ denote, respectively, the mean and variance of X (if they exist), and if Y is another stochastic variable, let $\sigma(XY)$ be the covariance between X and Y. Also let $\bar{X} = \dfrac{X - E(X)}{\sigma(X)}$. By the distribution of X we shall mean a function $\varphi(x)$ such that $P\{X < x\} \equiv \varphi(x)$. These conventions being established, we seek first to evaluate $E(r_i)$. This may be done by differentiating with respect to $\bar{y}$ the coefficient of x^n in the purely formal expansion in x of $2^{-n+1}(x + x^2 + \cdots + x^{i-1} + yx^i + x^{i+1} + \cdots)^r$, setting $y = 1$ and summing over all values of r. We have therefore to evaluate

$$2^{-n+1} \cdot \sum_{r=2}^{n} r \binom{n - i - 1}{r - 2},$$

which is easily seen to give us the result

$$E(r_i) = (n - i + 3)2^{-i-1}, \qquad (i < n), \tag{2.8}$$

while it is obvious that

$$E(r_n) = 2^{-n+1}. \tag{2.9}$$

By use of similar devices the variances and covariances of the r_i may also be obtained. We omit the details of those calculations and also the presentation of the covariances, since the latter are not necessary for the proof of Theorem 2. The results are:

$$\sigma^2(r_i) = n\left(\frac{1}{2^{i+1}} + \frac{3 - 2i}{2^{2i+2}}\right) + \left(\frac{3 - i}{2^{i+1}} + \frac{3i^2 - 12i + 5}{2^{2i+2}}\right), \qquad (i < \tfrac{1}{2}n). \tag{2.10}$$

The limitation on the value of i is necessary because the processes for summing binomial coefficients with the aid of the device described above are no longer applicable. The matter is easily settled, however, for if X is a stochastic variable which can take only the values 0 or 1, then

$$\sigma^2(X) = E(X) - [E(X)]^2.$$

The r_i for $i > \frac{1}{2}n$ are such variables, so that

$$\sigma^2(r_i) = \frac{n - i + 3}{2^{i+1}} - \frac{(n - i + 3)^2}{2^{2i+2}}, \qquad (n > i > \tfrac{1}{2}n), \tag{2.11}$$

$$\sigma^2(r_n) = \frac{(2^{n-1} - 1)}{2^{2n-2}}. \tag{2.12}$$

Also without difficulty we have

$$\sigma^2(r_{\frac{1}{2}n}) = \frac{n + 6}{2^{\frac{1}{2}(n+4)}} - \frac{(n + 6)^2}{2^{n+4}} + \frac{1}{2^{n-2}}, \tag{2.13}$$

when n is even and > 2, and

$$E(r) = \tfrac{1}{2}(n + 1), \tag{2.14}$$

$$\sigma^2(r) = \tfrac{1}{4}(n - 1). \tag{2.15}$$

Finally,

$$E(r'_{(k+1)}) = (n - k + 1)2^{-k-1}. \tag{2.16}$$

The next results we shall need may be expressed in the following:

THEOREM 1: *As n approaches infinity, the joint distribution of the stochastic* *variables* $\bar{r}_1, \cdots, \bar{r}_k, \bar{r}'_{(k+1)}$ (*k any fixed positive integer*), *approaches the multivariate normal distribution.*

This theorem is proved as follows: Make the substitutions

$$x_i = \frac{r_i - n \cdot 2^{-i-1}}{\sqrt{n}}, \qquad (i = 1, 2, \cdots, k),$$

$$x'_{(k+1)} = \frac{r'_{(k+1)} - n \cdot 2^{-k-1}}{\sqrt{n}}$$

in the expression

$$2^{-n+1} \frac{\left(\sum_{i=1}^{k} r_i + r'_{(k+1)}\right)!}{(r'_{(k+1)})! \prod_{i=1}^{k} (r_i)!} \binom{n - t - kr'_{(k+1)} - 1}{r'_{(k+1)} - 1},$$

which comes from (2.7), and regard t as equal to $\sum_{i=1}^{k} ir_i$. Replace the various factorials by their asymptotic approximations as given by Stirling's formula and simplify the resulting expression. The subsequent procedure is simple but laborious and we omit the details, which are like those of the classical proof of De Moivre's theorem as given, for example, in Frechet [1], p. 89.

We now prove the following theorem on additive partition functions:

THEOREM 2: *Let $f(x)$ be a function defined for all positive integral values of x which fulfills the following conditions*:

(a). *There exists a pair of positive integers, a and b, such that*

$$\frac{f(a)}{f(b)} \neq \frac{a}{b}, \tag{2.17}$$

(b). *the series*

$$\sum_{i=1}^{\infty} | f(i) | 2^{-\frac{1}{2}i}, \tag{2.18}$$

converges. Let $F(A)$, a function of the stochastic sequence A, be defined as follows:

$$F(A) = \sum_{i=1}^{s} f(a_i). \tag{2.19}$$

Then for any real y the probability of the inequality $\bar{F}(A) < y$, approaches

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-\frac{1}{2}y^2}\, dy,$$

as $n \to \infty$.

We restate this theorem without use of probabilistic terms:

Let A be any sequence of positive integers whose sum is a given integer n. Consider two sequences A to be different if they contain different elements or the same elements arranged in a different order. Let $f(x)$ and $F(A)$ be defined as above, with the aforementioned restrictions. Then there exist, for every positive integer n, two numbers E_n and σ_n, such that 2^{-n+1} multiplied by the number of sequences A for which the inequality

$$F(A) - E_n < y\sigma_n,$$

holds, approaches

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-\frac{1}{2}y^2}\, dy,$$

as $n \to \infty$.

For convenience, the proof will be divided into a number of lemmas.

If $\varphi(y)$ is any continuous distribution function, then it is well known that $\varphi(y)$ is uniformly continuous and that consequently, for any arbitrarily small, positive ϵ, there exist two positive numbers, h and D, with the following properties:

(a). If y_1 and y_2 are any real numbers such that $| y_1 - y_2 | < h$, then $| \varphi(y_1) - \varphi(y_2) | < \epsilon$,

(b). If y is such that $| y | > D$, then $\varphi(| y |) > 1 - \epsilon$, and $\varphi(-| y |) < \epsilon$.

We now first prove

LEMMA 1: *Let X and Y be two stochastic variables, both of which possess finite means and variances. Suppose that there exists a continuous distribution function $\varphi(y)$ and two small positive numbers ϵ and δ* (say $\epsilon < 1/10$, $\delta < 1/10$), *such that*

$$| P\{X < y\} - \varphi(y) | < \epsilon, \tag{2.20}$$

for all y, and

$$\frac{\sigma(Y)}{\sigma(X)} = \delta. \tag{2.21}$$

Let h and D be chosen as above for $\varphi(y)$, with the additional proviso that $h < \frac{1}{2}$ and $D > 1$ Suppose further that

$$\delta < \min\left(\frac{h}{4D}, \frac{h\epsilon}{8}\right). \tag{2.22}$$

Then

$$| P\{(\overline{X + Y}) < y\} - \varphi(y) | < 3\epsilon, \tag{2.23}$$

for all y.

PROOF: We have

$$\sigma^2(X + Y) = \sigma^2(X) + 2\sigma(XY) + \sigma^2(Y).$$

Since, as is well known,

$$| \sigma(XY) | \leq \sigma(X)\sigma(Y),$$

it follows from (2.21) that

$$\sigma(X + Y) = (1 + \delta')\sigma(X), \tag{2.24}$$

where $|\delta'| \leq \delta$ Hence

$$\sigma\left(\frac{Y - E(Y)}{\sigma(X + Y)}\right) < 2\delta. \tag{2.25}$$

From Tchebycheff's inequality and (2.21) it then follows that, if $d = h/4$,

$$P\left\{\left|\frac{Y - E(Y)}{\sigma(X + Y)}\right| > d\right\} < 4\frac{\delta^2}{d^2}, \tag{2.26}$$

and

$$\frac{4\delta^2}{d^2} < \epsilon^2 < \epsilon. \tag{2.27}$$

Now

$$\begin{aligned} P\left\{\frac{X - E(X)}{\sigma(X + Y)} < y - d\right\} &= P\left\{\frac{X - E(X)}{\sigma(X + Y)} < y - d; \left|\frac{Y - E(Y)}{\sigma(X + Y)}\right| \leq d\right\} \\ &\quad + P\left\{\frac{X - E(X)}{\sigma(X + Y)} < y - d; \left|\frac{Y - E(Y)}{\sigma(X + Y)}\right| > d\right\} \\ &< P\{(\overline{X + Y}) < y\} + \epsilon \\ &= P\left\{(\overline{X + Y}) < y; \left|\frac{Y - E(Y)}{\sigma(X + Y)}\right| \leq d\right\} \\ &\quad + P\left\{(\overline{X + Y}) < y; \left|\frac{Y - E(Y)}{\sigma(X + Y)}\right| > d\right\} + \epsilon \\ &< P\left\{\frac{X - E(X)}{\sigma(X + Y)} < y + d\right\} + 2\epsilon. \end{aligned} \tag{2.28}$$

Hence, from (2.24)

$$\begin{aligned} &P\{\bar{X} < (y - d)(1 + \delta')\} - \epsilon \\ &\quad < P\{(\overline{X + Y}) < y\} < P\{\bar{X} < (y + d)(1 + \delta')\} + \epsilon \end{aligned} \tag{2.29}$$

and consequently, from (2.20)

$$\begin{aligned} &\varphi(y - d + y\delta' - d\delta') - 2\epsilon \\ &\quad < P\{(\overline{X + Y}) < y\} < \varphi(y + d + y\delta' + d\delta') + 2\epsilon. \end{aligned} \tag{2.30}$$

Now if $|y| \leq 2D$, then from (2.22)

$$d + |y\delta'| + d|\delta'| < \frac{h}{4} + \frac{h}{2} + \frac{h}{4} = h,$$

and if $|y| > 2D$, then also from (2.22)

$$|y| - d - |y\delta'| - d|\delta'| > |y|(1-\delta) - \frac{h}{2} > \frac{3}{4}|y| > \frac{3}{2}D.$$

Recalling the definitions of h and D, it follows from (2.30) that, for all y,

$$\varphi(y) - 3\epsilon < P\{(\overline{X+Y}) < y\} < \varphi(y) + 3\epsilon. \tag{2.31}$$

This proves Lemma 1.

LEMMA 2: *For any fixed pair a, b, of positive integers such that $a < b$,*

$$\lim_{n\to\infty} \frac{[E(r)]^{b-a}\cdot[E(r_b)]^a}{[E(r_a)]^b} = 1 \tag{2.32}$$

PROOF: From (2.8), for fixed i

$$\frac{1}{n}E(r_i) \to 2^{-i-1},$$

and from (2.14) $\frac{1}{n}E(r) \to \frac{1}{2}$ as $n \to \infty$. The required result follows easily.

For any n we now define

$$B(k;n) = \sum_{i=1}^{k} r_i[f(i)],$$

and

$$C(k;n) = \sum_{i=k+1}^{n} r_i[f(i)].$$

Then

$$F(A) = B(k;n) + C(k;n).$$

LEMMA 3: *For any real y and any fixed positive integral k the probability that the stochastic variable $B(k;n)$ shall fulfill the inequality $\bar{B}(k;n) < y$ approaches $\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y} e^{-\frac{1}{2}y^2}\,dy$, as $n \to \infty$.*

PROOF: By Theorem 1, the stochastic variables $\bar{r}_1, \bar{r}_2, \cdots, \bar{r}_k, \bar{r}'_{(k+1)'}$ are asymptotically jointly normally distributed. As an immediate consequence so are the variables $\bar{r}_1, \bar{r}_2, \cdots, \bar{r}_k$, and hence $B(k;n)$, which is a linear function with constant coefficients $f(1), f(2), \cdots f(k)$, of $r_1, r_2, \cdots, r_k$, is asymptotically normally distributed.

LEMMA 4. *There exists a constant $c > 0$, such that, for all n sufficiently large,*

$$\sigma^2(F(A)) > cn. \tag{2.33}$$

PROOF: For any sufficiently large, arbitrary, but fixed n, we will construct two sets, S_1 and S_2, of sequences A, with the following properties: S_1 and S_2 have the same probability p, with p always greater than β, a fixed positive constant which does not depend on n. Since the probabilities of S_1 and S_2 are equal, each possesses the same number of sequences A. Between the member sequences of the sets S_1 and S_2 we will establish a one-to-one correspondence such that, if A_1 is a member of S_1 and A_2 is its corresponding sequence in S_2, then

$$|F(A_1) - F(A_2)| > 2d\sqrt{n}, \tag{2.34}$$

where d is a fixed positive constant which does not depend on n.

It is easy to see that such a construction would prove the lemma. The probability of any sequence A is 2^{-n+1}. Hence the contribution of a corresponding pair A_1 and A_2 to the variance of $F(A)$ is by (2.34) not less than $2^{-n+2}d^2n$ and the contribution of the sets S_1 and S_2 is not less than $2\beta d^2 n$.

It remains then to carry out the construction of S_1 and S_2. For the sake of simplicity in notation, we shall carry out the construction with the assumption that the integers a and b of (2.17) are 1 and 2. It will be readily apparent, however, that the proof is perfectly general and with trivial changes holds for any pair a, b. This lemma is the only place where the hypothesis (2.17) is used. The latter condition is necessary because, if for every pair of positive integers i and j,

$$\frac{f(i)}{f(j)} = \frac{i}{j},$$

then $F(A)$ is a constant multiple of n, for $n = \sum_i ir_i$ and then

$$F(A) = \sum_i f(a_i) = \sum_i r_i f(i) = f(1) \sum_i ir_i = nf(1).$$

Each sequence A uniquely determines the "coördinate" complex

$$\{r_1, r_2, \cdots, r_n\}$$

which we prefer to write as the pair $L = (l, l')$:

$$l = \{r_1, r_2\},$$

$$l' = \{r_3, r_4, \cdots, r_n\}.$$

To each pair (l, l') there correspond in general many sequences A whose exact number may be explicitly given in terms of factorials. The totality of all A whose L have the same second member l' will be called the group determined by l', or just the group l'. The subset of a group l' all of whose A have the same r_1 will be called the family (l', r_1). All the A in the same family have the same L. For l' and r_1 determine r_2 through the equation $\sum_i ir_i = n$.

According to Theorem 1 for $k = 2$, r_1, r_2, r_3' are asymptotically jointly normally distributed. Let

$$\sigma_1 = \lim_{n\to\infty} \frac{\sigma(r_1)}{\sqrt{n}}$$

The limiting variances of r_2 and r_3' are constant multiples of $n\sigma_1^2$. Therefore the set H of all A whose L satisfy the constraints

$$\frac{n}{4} < r_1 < \frac{n}{4} + \sqrt{n}\,\sigma_1$$

$$(2.35) \qquad \frac{n}{8} < r_2 < \frac{n}{8} + \sqrt{n}\,\sigma_1$$

$$\frac{n}{8} < r_3' < \frac{n}{8} + \sqrt{n}\,\sigma_1$$

has, by virtue of the fact that the limiting correlation coefficients of the variables r_1, r_2, r_3' are all less than 1 in absolute value, a positive probability, which exceeds a fixed positive constant γ for sufficiently large n. If any member sequence A of a family is in H, the entire family is obviously in H. Any sequence A belongs to one and only one family. Hence the set H may be decomposed in a disjunct way into entire families. Let $\left(l', \frac{n}{4} + h_1\right)$ be *any* family in H, where of course $0 < h_1 < \sqrt{n}\,\sigma_1$. Consider the (second) family $\left(l', \frac{n}{4} + 2\sqrt{n}\,\sigma_1 + h_1\right)$. This family is not in H. We now wish to show that the probability of the second family exceeds c' times the probability of the first family, where c' is a fixed positive constant which does not depend on either n or the particular families in question.

For the first family, let

$$r_1 = \frac{n}{4} + h_1, \qquad r_3' = \frac{n}{8} + h_3,$$

$$r_2 = \frac{n}{8} + h_2, \qquad r = \frac{n}{2} + h_1 + h_2 + h_3.$$

Hence

$$(2.36) \qquad 0 < h_i < \sqrt{n}\sigma_1 \qquad (i = 1, 2, 3).$$

For the second family we therefore have, since both families are in the same group,

$$r_1 = \frac{n}{4} + 2\sqrt{n}\,\sigma_1 + h_1,$$

$$r_2 = \frac{n}{8} - \sqrt{n}\,\sigma_1 + h_2,$$

$$r_3' = \frac{n}{8} + h_3,$$

$$r = \frac{n}{2} + \sqrt{n}\,\sigma_1 + h_1 + h_2 + h_3.$$

The ratio of the probability of the second family to that of the first family equals the ratio of the number of sequences A in the second family to the number of sequences A in the first family. By elementary combinatorics, since both families are in the same group, the latter ratio is

$$\frac{\left(\frac{n}{2} + \sqrt{n}\,\sigma_1 + h_1 + h_2 + h_3\right)!}{\left(\frac{n}{4} + 2\sqrt{n}\,\sigma_1 + h_1\right)!\left(\frac{n}{8} - \sqrt{n}\,\sigma_1 + h_2\right)!} \frac{\left(\frac{n}{4} + h_1\right)!\left(\frac{n}{8} + h_2\right)!}{\left(\frac{n}{2} + h_1 + h_2 + h_3\right)!} \tag{2.37}$$

and hence exceeds

$$\left(\frac{n}{2} + h_1 + h_2 + h_3\right)^{\sqrt{n}\,\sigma_1} \times \left(\frac{n}{4} + 2\sqrt{n}\,\sigma_1 + h_1\right)^{-2\sqrt{n}\,\sigma_1}\left(\frac{n}{8} - \sqrt{n}\,\sigma_1 + h_2\right)^{\sqrt{n}\,\sigma_1}. \tag{2.38}$$

At this point, if we had been using the numbers a and b of (2.17), we would make use of Lemma 2. In the present case the result of that lemma is trivial. It is easy to see, therefore, that (2.38) equals

$$\left(1 + \frac{2h_1 + 2h_2 + 2h_3}{n}\right)^{\sqrt{n}\,\sigma_1} \times \left(1 + \frac{8\sqrt{n}\,\sigma_1 + 4h_1}{n}\right)^{-2\sqrt{n}\,\sigma_1}\left(1 - \frac{8\sqrt{n}\,\sigma_1 - 8h_2}{n}\right)^{\sqrt{n}\,\sigma_1}, \tag{2.39}$$

which, in view of (2.36), exceeds

$$\left(1 + \frac{12\sigma_1}{\sqrt{n}}\right)^{-2\sqrt{n}\,\sigma_1} \cdot \left(1 - \frac{8\sigma_1}{\sqrt{n}}\right)^{\sqrt{n}\,\sigma_1} \tag{2.40}$$

which, in turn, for sufficiently large n, exceeds

$$\tfrac{1}{2} \cdot e^{-24\sigma_1^2 - 8\sigma_1^2} = \tfrac{1}{2} e^{-32\sigma_1^2} = c'. \tag{2.41}$$

We are now ready to construct S_1 and S_2. Let

$$f_1 = (l', r_1)$$

be any family in H and consider the family

$$f_2 = (l', r_1 + 2\sqrt{n}\,\sigma_1).$$

Select in any manner whatsoever $c'\nu$ of the sequences A in f_1, where ν is the total number of sequences in f_1. Call this set of sequences f^*. Select in any manner whatsoever $c'\nu$ sequences from f_2 and call this set f^{**}. That there exist at least $c'\nu$ sequences in f_2 is assured by equation (2.41). In any manner whatsoever establish a one-to-one correspondence between the sequences of f^* and f^{**}. Suppose A_1 and A_2 are corresponding sequences. Since f^* and f^{**} belong to the same group, and since $f(2) \neq 2f(1)$, we have

$$(2.42)\qquad \begin{aligned} |F(A_1) - F(A_2)| &= |f(2)\sqrt{n}\,\sigma_1 - 2f(1)\sqrt{n}\,\sigma_1| \\ &= |f(2) - 2f(1)|\sqrt{n}\,\sigma_1, \end{aligned}$$

so that (2.34) holds with

$$(2.43)\qquad d = \tfrac{1}{4}\,|f(2) - 2f(1)|\,\sigma_1 .$$

Now proceed in this manner for all the families f_1 in H. The union of all the sets f^{**} is the set S_1 and the union of all the sets f^{**} is the set S_2. It is clear that, since the probability of H exceeds γ, the probability p of S_1 exceeds $\beta = c'\gamma$. This proves Lemma 4.

LEMMA 5. *For any arbitrarily small positive number ξ there exists a positive integer $\mu(\xi)$, such that for any $k > \mu(\xi)$ and all n greater than a fixed lower bound,*

$$(2.44)\qquad \sigma^2[C(k;n)] < \xi n.$$

PROOF: Since

$$C(k;n) = \sum_{i=k+1}^{n} r_i f(i),$$

and, as is well known,

$$|\sigma(XY)| \leq \sigma(X)\sigma(Y)$$

we have

$$(2.45)\qquad \sigma^2[C(k;n)] \leq \left[\sum_{i=k+1}^{n} |f(i)|\,\sigma(r_i)\right]^2$$

From (2.10) it follows readily that

$$(2.46)\qquad \sigma^2(r_i) < \frac{n}{2^i} + \frac{5}{2^{i+1}} + \left(\frac{-i}{2^{i+1}} + \frac{3i^2}{2^{2i+2}}\right),$$

and the quantity in parentheses in the right member of (2.46) is easily seen to be negative, so that, for $i < \frac{1}{2}n$ and $n \geq 3$,

$$(2.47)\qquad \sigma(r_i) < \sqrt{2n}\,2^{-\frac{1}{2}i}$$

From (2.11) and the definition of r_i, it follows easily that (2.47) holds also when $i > \frac{1}{2}n$ and $n \geq 3$.

Hence, in view of (2.12), (2.13), and the convergence of the series in (2.18), the desired result follows from (2.45).

LEMMA 6. *Let the ξ of Lemma 5 be $< \frac{1}{4}c$, where c is as in Lemma 4. Then for $k > \mu(\xi)$ and n larger than a fixed lower bound*

$$(2.48)\qquad \sigma^2(B(k;n)) > \tfrac{1}{4}cn.$$

PROOF: Since

$$F(A) = B(k;n) + C(k;n),$$

we have

$$\sigma^2(F(A)) = \sigma^2(B(k;n)) + \sigma^2(C(k;n)) + 2\sigma(BC)$$
$$\leq \sigma^2(B) + \sigma^2(C) + 2\sigma(B)\sigma(C) = (\sigma(B) + \sigma(C))^2.$$

Hence from (2.33) and (2.44) $\sqrt{cn} < \sigma(B) + \frac{1}{2}\sqrt{cn}$ and the required result follows.

Proof of the Theorem: Let ϵ be an arbitrarily small positive number. For all n sufficiently large we have, by Lemma 3,

$$\left| P\{\bar{B}(k;n) < y\} - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-\frac{1}{2}y^2}\, dy \right| < \epsilon,$$

for all y. For a small ξ to be chosen later and large enough k and n we have, by Lemmas 5 and 6,

$$\frac{\sigma(C(k;n))}{\sigma(B(k;n))} = \delta < \frac{4\xi}{c}. \tag{2.49}$$

Now let the $\varphi(y)$ of Lemma 1 be defined as

$$\varphi(y) \equiv \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-\frac{1}{2}y^2}\, dy,$$

and choose h and D as in Lemma 1 for our present ϵ. Since c is fixed and ξ still at our disposal, choose ξ sufficiently small so that the δ of (2.49) satisfies (2.22). Since the hypothesis of Lemma 1 is satisfied, we have, from (2.23) and Lemma 3, for all n sufficiently large,

$$| P\{\bar{F}(A) < y\} - \varphi(y) | < 3\epsilon$$

for all y. This is the required result.

3. Partitions of two integers. Let n_1 and n_2 be positive integers, $n_1 + n_2 = n$. $\frac{n_1}{n} = e_1$, $\frac{n_2}{n} = e_2$, and $e = \max(e_1, e_2)$. Let $V = (v_1, v_2, \cdots, v_s)$ be any sequence of positive integers v_i $(i = 1, 2, \cdots, s)$ where $a_1 + a_3 + a_5 + \cdots$ equals either one of n_1 and n_2, while $a_2 + a_4 + a_6 + \cdots$ equals the other. Such sequences are of statistical importance (cf. Wald and Wolfowitz [2]). As before, sequences V with different elements or with the same elements in different order will be considered different and to each sequence V will be assigned the same probability, which is therefore easily seen to be $\frac{n_1!\, n_2!}{n!}$.

Let r_{1i} be the number of elements equal to i in that one of the two sequences $(a_1, a_3, a_5, \cdots)$ and $(a_2, a_4, a_6, \cdots)$ the sum of whose elements is n_1 and let r_{2i} be the corresponding number for the other sequence. Let

$$s_i = r_{1i} + r_{2i},$$
$$r_1 = \sum_i r_{1i}, \qquad r_2 = \sum_i r_{2i},$$
$$s = r_1 + r_2, \qquad r'_{1(k+1)} = \sum_{i=k+1}^{n_1}{}^{*} r_{1i}$$
$$r'_{2(k+1)} = \sum_{i=k+1}^{n_2} r_{2i}.$$

The necessary computations such as are given in the beginning of the previous section have been performed by Mood [3] and we summarize them as follows:

THEOREM 3 (Mood): *As n approaches infinity while e_1 and e_2 remain constant, the joint distribution of the stochastic variables*

$$\bar{r}_{11}, \bar{r}_{12}, \cdots, \bar{r}_{1k}, \bar{r}'_{1(k+1)}, \bar{r}_{21}, \bar{r}_{22}, \cdots \bar{r}_{2k}$$

(where k is any fixed positive integer), approaches the multivariate normal distribution.

Mood (loc. cit.) gives the following parameters, with the convention that

(3.1) $$x^{(i)} = x(x-1)(x-2)\cdots(x-i+1):$$

(3.2) $$E(r_{1i}) = \frac{(n_2+1)^{(2)} n_1^{(i)}}{n^{(i+1)}},$$

(3.3) $$\lim_{n\to\infty} \frac{E(r_{1i})}{n} = e_1^i e_2^2,$$

(3.4) $$\lim_{n\to\infty} \frac{E(r'_{1(k+1)})}{n} = e_1^{k+1} e_2,$$

(3.5) $$\sigma^2(r_{1i}) = \frac{n_2^{(2)}(n_2+1)^{(2)} n_1^{(2i)}}{n^{(2i+2)}} + \frac{(n_2+1)^{(2)} n_1^{(i)}}{n^{(i+1)}}\left(1 - \frac{(n_2+1)^{(2)} n_1^{(i)}}{n^{(i+1)}}\right),$$

(3.6) $$\lim_{n\to\infty} \frac{\sigma^2(r_{1i})}{n} = e_1^{2i-1} e_2^3[(i+1)^2 e_1 e_2 - i^2 e_2 - 2e_1] + e_1^i e_2^2.$$

The corresponding parameters for r_{2i} may be obtained from the above by interchange of n_1 and n_2. Also

(3.7) $$\lim_{n\to\infty} \frac{E(r_1)}{n} = \lim_{n\to\infty} \frac{E(r_2)}{n} = e_1 e_2.$$

For additive partition functions we have the following theorem:

THEOREM 4. *Let $f(x)$ be a function defined for all positive integral values of x which fulfills the following conditions:*

a) *There exists a pair of positive integers, a and b, such that*

(3.8) $$\frac{f(a)}{f(b)} \neq \frac{a}{b};$$

b) *the series*

$$\sum_{i=1}^{\infty} | f(i) | e^{i/2} \tag{3.9}$$

converges. Let $F(V)$, *a function of the stochastic sequence* V, *be defined as follows:*

$$F(V) = \sum_{i=1}^{s} f(v_i) \tag{3.10}$$

Then for any real y *the probability of the inequality* $\bar{F}(V) < y$ *approaches*

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-\frac{1}{2}y^2}\, dy,$$

as $n \to \infty$, *while* e_1 *and* e_2 *remain constant.*

The basic idea of the proof of this theorem is the same as that of the proof of Theorem 2. We omit all the steps which can be written without difficulty by analogy to those in Theorem 2 and present only those where some major change is necessary. The numbering of the lemmas will correspond to that of Theorem 2.

LEMMA 2. *For any fixed pair,* a *and* b, *of positive integers such that* $a < b$,

$$[E(r_1)]^{b-a} \cdot [E(r_2)]^{b-a} \cdot [E(r_{1b})]^{a} \cdot [E(r_{2b})]^{a} \cdot [E(r_{1a})]^{-b} \cdot [E(r_{2a})]^{-b} \to 1, \tag{3.11}$$

as $n \to \infty$.

The proof is the same as before.

The following are the definitions corresponding to those of Theorem 2:

$$B(k;n) = \sum_{i=1}^{k} s_i f(i),$$

$$C(k;n) = \sum_{i=k+1}^{n} s_i f(i).$$

Then as before

$$F(V) = B(k;n) + C(k;n).$$

LEMMA 4. Statement is the same as that for Theorem 2. The following important changes must be made in the proof:

Each sequence V determines the coordinate complex

$$\begin{Bmatrix} r_{11}, r_{12}, \cdots, r_{1n} \\ r_{21}, r_{22}, \cdots, r_{2n} \end{Bmatrix}$$

also

$$l = \begin{Bmatrix} r_{11}, r_{12} \\ r_{21}, r_{22} \end{Bmatrix},$$

and

$$l' = \begin{Bmatrix} r_{13}, & \cdots, & r_{1n} \\ r_{23}, & \cdots, & r_{2n} \end{Bmatrix}.$$

The set H is the set of all V whose L satisfy the constraints

$$ne_1e_2^2 < r_{11} < ne_1e_2^2 + \sqrt{n}\,\sigma_{11},$$
$$ne_1^2e_2^2 < r_{12} < ne_1^2e_2^2 + \sqrt{n}\,\sigma_{11},$$
$$ne_1^2e_2 < r_{21} < ne_1^2e_2 + \sqrt{n}\,\sigma_{11},$$
$$ne_1^2e_2^2 < r_{22} < ne_1^2e_2^2 + \sqrt{n}\,\sigma_{11},$$
$$ne_1^3e_2 < r_{13}' < ne_1^3e_2 + \sqrt{n}\,\sigma_{11},$$

where

$$\sigma_{11} = \lim_{n\to\infty} \frac{\sigma(r_{11})}{\sqrt{n}}.$$

The representative family for H is characterized by

$$(l', ne_1e_2^2 + h_{11}),$$

and this family is compared with the family

$$(l', ne_1e_2^2 + 2\sqrt{n}\,\sigma_{11} + h_{11}).$$

For the members of the family in H

$$r_{11} = ne_1e_2^2 + h_{11} = nm_{11} + h_{11},$$
$$r_{12} = ne_1^2e_2^2 + h_{12} = nm_{12} + h_{12},$$
$$r_{21} = ne_1^2e_2 + h_{21} = nm_{21} + h_{21},$$
$$r_{22} = ne_1^2e_2^2 + h_{22} = nm_{22} + h_{22},$$
$$r_{13}' = ne_1^3e_2 + h_{13} = nm_{13}' + h_{13},$$
$$r_1 = ne_1e_2 + h' = nm + h,$$
$$|r_2 - r_1| \le 1,$$

where

$$h_{ij} < \sqrt{n}\,\sigma_{11}, \tag{3.12}$$

$$h = h_{11} + h_{12} + h_{13}. \tag{3.13}$$

And for the members of the second family

$$r_{11} = nm_{11} + 2\sqrt{n}\,\sigma_{11} + h_{11},$$
$$r_{12} = nm_{12} - \sqrt{n}\,\sigma_{11} + h_{12},$$

$$\begin{aligned} r_{21} &= nm_{21} + 2\sqrt{n}\,\sigma_{11} + h_{21} + \theta_{21}, \\ r_{22} &= nm_{22} - \sqrt{n}\,\sigma_{11} + h_{22} + \theta_{22}, \\ r'_{13} &= nm'_{13} + h_{13}, \\ r_1 &= nm + \sqrt{n}\,\sigma_{11} + h, \\ &\quad | r_2 - r_1 | \leq 1, \end{aligned}$$

with

$$| \theta_{21} | \leq 1, \qquad | \theta_{22} | \leq 1.$$

To the expression (2.37) corresponds the expression (3.14), with $| \theta | \leq 1$:

$$\begin{aligned} &\frac{(nm_{11} + h_{11})!\,(nm_{12} + h_{12})!}{(nm + h)!} \times \frac{(nm_{21} + h_{21})!\,(nm_{22} + h_{22})!}{(nm + h)!} \\ &\times \frac{(nm + h + \sqrt{n}\,\sigma_{11})!}{(nm_{11} + 2\sqrt{n}\,\sigma_{11} + h_{11})!\,(nm_{12} - \sqrt{n}\,\sigma_{11} + h_{12})!} \\ &\times \frac{(nm + h + \sqrt{n}\,\sigma_{11} + \theta)!}{(nm_{21} + 2\sqrt{n}\,\sigma_{11} + h_{21} + \theta_{21})!\,(nm_{22} - \sqrt{n}\,\sigma_{11} + h_{22} + \theta_{22})!}, \end{aligned} \tag{3.14}$$

which exceeds

$$\begin{aligned} &(nm + h)^{2\sqrt{n}\,\sigma_{11}} \times (nm_{11} + 2\sqrt{n}\,\sigma_{11} + h_{11})^{-2\sqrt{n}\,\sigma_{11}} \\ &\times (nm_{12} - \sqrt{n}\,\sigma_{11} + h_{12})^{\sqrt{n}\,\sigma_{11}} \\ &\times (nm_{21} + 2\sqrt{n}\,\sigma_{11} + h_{21})^{-2\sqrt{n}\,\sigma_{11}} \\ &\times (nm_{22} - \sqrt{n}\,\sigma_{11} + h_{22})^{\sqrt{n}\,\sigma_{11}}. \end{aligned} \tag{3.15}$$

Employing Lemma 2, we find that (3.15) equals

$$\begin{aligned} &\left(1 + \frac{h}{nm}\right)^{2\sqrt{n}\,\sigma_{11}} \times \left(1 + \frac{2\sqrt{n}\,\sigma_{11} + h_{11}}{nm_{11}}\right)^{-2\sqrt{n}\,\sigma_{11}} \\ &\times \left(1 + \frac{-\sqrt{n}\,\sigma_{11} + h_{12}}{nm_{12}}\right)^{\sqrt{n}\,\sigma_{11}} \\ &\times \left(1 + \frac{2\sqrt{n}\,\sigma_{11} + h_{21}}{nm_{21}}\right)^{-2\sqrt{n}\,\sigma_{11}} \\ &\times \left(1 + \frac{-\sqrt{n}\,\sigma_{11} + h_{22}}{nm_{22}}\right)^{\sqrt{n}\,\sigma_{11}}. \end{aligned} \tag{3.16}$$

In view (3.12) and (3.13), (3.16) exceeds

$$\begin{aligned} &\left(1 + \frac{3\sqrt{n}\,\sigma_{11}}{nm_{11}}\right)^{-2\sqrt{n}\,\sigma_{11}} \times \left(1 - \frac{\sqrt{n}\,\sigma_{11}}{nm_{12}}\right)^{\sqrt{n}\,\sigma_{11}} \\ &\times \left(1 + \frac{3\sqrt{n}\,\sigma_{11}}{nm_{21}}\right)^{-2\sqrt{n}\,\sigma_{11}} \times \left(1 - \frac{\sqrt{n}\,\sigma_{11}}{nm_{22}}\right)^{\sqrt{n}\,\sigma_{11}}, \end{aligned} \tag{3.17}$$

which, for sufficiently large n, in turn exceeds

$$\tag{3.18} \tfrac{1}{2}\cdot e^{-\sigma_{11}^2}\left(\frac{6}{m_{11}}+\frac{1}{m_{12}}+\frac{6}{m_{21}}+\frac{1}{m_{22}}\right)=c'.$$

LEMMA 5. Statement is the same as for Theorem 2. The proof then proceeds as follows:

We have

$$\tag{3.19} \sigma^2(C(k;n)) \le \left(\sum_{i=1}^{2}\sum_{j=k+1}^{n} |f(j)|\,\sigma(r_{ij})\right)^2.$$

From an examination of (3.5) and (3.6) we may see without any difficulty that the second of the three terms of the right member of (3.5) (after removal of parentheses) is asymptotically equal to n times the last term of the right member of (3.6) and hence that the other two terms of the right member of (3.5) are asymptotically equal to n times the right member of (3.6) without its last term. Now when

$$\frac{i+1}{i}\sqrt{e_1}<1$$

which will always occur when i is equal to or greater than a sufficiently large fixed integer μ, that part of the right member of (3.6) which is in square brackets is easily seen to be negative. Hence from the definition of asymptotic equivalence it follows that, for all n sufficiently large,

$$\tag{3.20} \frac{n_2^{(2)}(n_2+1)^{(2)}n_1^{(2\mu)}}{n^{(2\mu+2)}} < \frac{(n_2+1)^{(2)}(n_2+1)^{(2)}n_1^{(\mu)}n_1^{(\mu)}}{n^{(\mu+1)}n^{(\mu+1)}},$$

and

$$\tag{3.21} \frac{(n_2+1)^{(2)}n_1^{(\mu)}}{n^{(\mu+1)}} < 2ne^{\mu+2} < 2ne^{\mu}.$$

Hence, for all n sufficiently large,

$$\tag{3.22} \sigma^2(r_{1\mu}) < 2ne^{\mu}.$$

Now consider the expression (3.5) for $i=\mu$ and $i=\mu+1$. Passage from μ to $\mu+1$ multiplies the first term of the right member of (3.5) by

$$\tag{3.23} \frac{(n_1-2\mu)(n_1-2\mu-1)}{(n-2\mu-2)(n-2\mu-3)},$$

and the third term of the right member by

$$\tag{3.24} \frac{(n_1-\mu)^2}{(n-\mu-1)^2}.$$

It is easy to see that for large but fixed μ and all n greater than a lower bound which is a function of μ only, the expression (3.23) is less than the expression (3.24). Hence, in view of (3.20), the sum of the first and third terms of the right member of (3.5) for $i = \mu + 1$ is negative. Now consider what happens to the second term of the right member of (3.5) when i goes from μ to $\mu + 1$. It is multiplied by

$$\frac{(n_1 - \mu)}{(n - \mu - 1)}, \tag{3.25}$$

which, also for large but fixed μ and all n larger than a lower bound which is a function of μ only, is easily seen to be less than e. Consequently

$$\sigma^2(r_{1(\mu+1)}) < 2ne^{\mu+1}. \tag{3.26}$$

It can be seen without difficulty that such a passage of (3.5) to the next higher index is always accompanied by multiplication by expressions similar to (3.23), (3.24), and (3.25), for which similar inequalities hold and that consequently

$$0 \leq \sigma^2(r_{1i}) < 2ne^i, \tag{3.27}$$

and for similar reasons

$$0 \leq \sigma^2(r_{2i}) < 2ne^i,$$

for all i not less than μ and for all n greater than a lower bound which is a function of μ only (although it may be necessary to increase the original μ so that *both* the last two equations hold). The required result follows from (3.19) and the convergence of the series (3.9).

The proof of Theorem 4 follows along the same lines as that of Theorem 2. When $f(x) \equiv 1$, $F(V) \equiv U(V)$, the statistic discussed in [2]. Other such results follow from specialization of $f(x)$. Theorem 4 may also be generalized so that the elements v_i which add up to n_1 are operated on by a function f_1, while the elements v_i which add up to n_2 are operated on by another function f_2, but this is easy to see and we do not go into the details.

4. Tests of hypotheses in the non-parametric case. The great advances that have been made in mathematical statistics in recent years have been in two directions. On the one hand, the foundations of statistics, the theory of estimation and of testing hypotheses have been put on a rigorous basis of probability theory, and on the other, powerful methods for obtaining critical regions and confidence intervals and criteria for appraising their efficacy have been developed. Most of these developments have this feature in common, that the distribution functions of the various stochastic variables which enter into their problems are assumed to be of known functional form, and the theories of estimation and of testing hypotheses are theories of estimation of and of testing hypotheses about, one or more parameters, finite in number, the knowledge of which would completely determine the various distribution functions involved. We shall refer to this situation for brevity as the parametric case, and denote the opposite situation, where the functional forms of the distributions are unknown, as the non-parametric case.

The literature of theoretical statistics, therefore, deals principally with the parametric case. The reasons for this are perhaps partly historic, and partly the fact that interesting results could more readily be expected to follow from the assumption of normality. Another reason is that, while the parametric case was for long developed on an intuitive basis, progress in the non-parametric case requires the use of modern notions. However, the needs of theoretical completeness and of practical research require the development of the theory of the non-parametric case. The purpose of the following section is to contribute to this theory.

Brief mention of some of the literature may be made here. The problem of parametric estimation by confidence intervals, was put on a rigorous foundation by Neyman [4] and extended to the estimation of distribution functions in the non-parametric case by means of confidence belts by Wald and Wolfowitz [5]. Problems of testing non-parametric hypotheses have been treated in various places. The rank correlation coefficient has been used for a long time to test the independence of two variates. Its distribution was shown to be asymptotically normal by Hotelling and Pabst [6] and its small sample distribution was discussed by Olds [7]. The problem of two samples has been discussed, among others, by Thompson [8], Dixon [9] and Wald and Wolfowitz [2]. In 1937, Friedman [10] posed the non-parametric analogue of the problem in the analysis of variance and proposed a very ingenious solution.

All these proposed solutions have this in common, that there exists no general principle which can be applied in each particular case to obtain a critical region, a role which is performed in the parametric case by Fisher's principle of maximum likelihood and the likelihood ratio criterion (Neyman and Pearson, [11]), whose validity, at least for large samples, has been established by Wald ([12], [13]). In each problem the solutions proposed have been intuitive and usually based on an analogy to the corresponding problem in the parametric case. Thus the principal justification for the use of the rank correlation coefficient is that its distribution is independent of the unknown distribution function (under the null hypothesis) and that its structure resembles that of the ordinary correlation coefficient. But any function of the order relations among the variates (cf. [2], p. 148) has a distribution which is independent of the unknown population distribution under the null hypothesis The same objection may be made to papers [8], [9], [10], [2], except that in [2], although the solution there proposed is an intuitive one, the criterion of consistency is extended from the parametric case to the non-parametric one. The fulfilment of this condition is a minimal requirement of a good test and on this basis the solution proposed in one of the previous papers cannot be considered a good one.

In the following section we shall show that the likelihood ratio criterion may be extended to the non-parametric case where the test must be made on the order relations among the observations and that for a certain class of these problems which fulfill the same requirement as that for the application of the likelihood ratio criterion in the parametric case it would thus appear to furnish a general method by which statistics may be obtained for a specific problem. We shall show this by applying it to the problem of two samples. This will serve to explain the method. Another problem will be discussed later. The ultimate justification of any statistic must be its power function, which ought therefore to constitute the next subject of investigation for these problems. Since for problems in the non-parametric case it is almost certain that uniformly most powerful tests do not exist, the question of determining the alternatives with respect to which proposed tests are powerful is particularly important.

5. The problem of two samples. Let X and Y be two stochastic variables with the distribution functions $f(x)$ and $g(x)$, respectively. (The term distribution function will always denote the cumulative distribution function. The letter P followed by an expression in braces will stand for the probability of the relation in braces. Hence $P\{X < x\} = f(x)$ for all x.) $f(x)$ and $g(x)$ are assumed continuous. The n_1 observations $x_1, x_2 \cdots, x_{n_1}$ and n_2 observations $y_1, y_2, \cdots, y_{n_2}$ are made on X and Y respectively. The (null) hypothesis to be tested is that $f(x) \equiv g(x)$. The admissible alternatives are all continuous distribution functions $f(x)$ and $g(x)$ such that $f(x) \not\equiv g(x)$. The $n_1 + n_2 = n$ observations are arranged in ascending order of size, thus: $Z = z_1, \cdots, z_n$ where $z_1 < z_2 < \cdots < z_n$ (the probability that $z_i = z_{i+1}$ is 0). Let $V = v_1, v_2, \cdots, v_n$ be a sequence defined as follows: $v_i = 0$ if z_i is a member of the set $x_1, x_2, \cdots, x_{n_1}$ and $v_i = 1$ if z_i is a member of the set $y_1, y_2, \cdots, y_{n_2}$. Then any statistic used to test the null hypothesis must be a function only of V ([2], p. 148).

We now apply the method of Neyman and Pearson [11] as follows: Ω is the totality of all couples $(d_1(x), d_2(x))$ of continuous distribution functions. The set ω, a subset of Ω, is the totality of all couples of distribution functions for which $d_1 \equiv d_2$. The sample space is the totality of all sequences V. The null hypothesis states that (f, g) is a member of ω. The admissible alternatives are that (f, g) is a member of Ω not in ω. The distribution of any function of V is the same for all members of ω. Hence this essential requirement on the statistic to be selected for the application of the likelihood ratio criterion (cf. [11]) is satisfied by any statistic which is a function of V alone. Furthermore, all sequences V have the same probability if the null hypothesis is true ([2], p. 149). The numerator of the likelihood ratio is therefore a function only of n_1 and n_2, is the same for all V, and is therefore of no further interest. Hence $T'(V)$, a function of V which is a monotonic function of the likelihood ratio for this problem, may be defined as the denominator of the likelihood ratio, as follows: Let $P\{V; (d_1, d_2)\}$ be the probability of V when $f \equiv d_1$, and $g \equiv d_2$. Then

$$T'(V) = \max_{\Omega} P\{V; (d_1, d_2)\}.$$

The critical values of $T'(V)$ are the large values. However, we may use instead of $T'(V)$ a convenient monotonic function of $T'(V)$.

As an approximation to $T'(V)$ we propose $T(V)$, a statistic which is obtained on the assumption that for a given V a couple (d_1^*, d_2^*) which is essentially the same as that of the two sample distribution functions corresponding to the particular V approximates a couple which maximizes the right member of (5.1). (We say "*a*" couple because it cannot be unique.) This assumption seems a reasonable one, particularly for large samples. Only the form of (d_1^*, d_2^*) is assumed and the missing parameters are obtained in accordance with (5.1). Before describing the matter precisely, it must be stressed that this is offered only as a plausible approximation. For certain extreme V, for example, like those where zeros and ones nearly alternate, this is definitely not the maximizing couple. In spite of this the statistic $T(V)$ assigns to these V values which are furthest removed from the critical region for any level of significance, as indeed any good statistic should.

We first define a "*run*" as in [2], p. 149. A subsequence $v_{(t+1)}, v_{(t+2)}, \cdots, v_{(t+r)}$ of V (where r may also be 1) is called a "*run*" if $v_{(t+1)} = v_{(t+2)} = \cdots = v_{(t+r)}$ and if $v_t \neq v_{(t+1)}$ when $t > 0$ and if $v_{(t+r)} \neq v_{(t+r+1)}$ when $t + r < n$. Let l_{1j} be the number of elements in the j^{th} *run* of elements 0, and l_{2j} the number of elements in the j^{th} *run* of elements 1. Suppose for a moment that the first element in V is a 0. Consider the following situation: There is an interval $[a_1, a_2]$, $a_1 < a_2$, on the line $-\infty < x < +\infty$ such that

$$P\{a_1 \leq X \leq a_2\} > 0, \qquad P\{a_1 \leq Y \leq a_2\} = 0,$$
$$P\{X < a_1\} = P\{Y < a_1\} = 0.$$

This is followed by an interval $[b_1, b_2]$, $b_1 = a_2$, such that $P\{b_1 \leq X \leq b_2\} = 0$, $P\{b_1 \leq Y \leq b_2\} > 0$. This is in turn followed by an interval $[a_3, a_4]$, $a_3 = b_2$, such that $P\{a_3 \leq X \leq a_4\} > 0$, $P\{a_3 \leq Y \leq a_4\} = 0$, etc. It is clear that the lengths and location of the intervals described are immaterial, provided only that they do not overlap. Also the distributions of X and Y within each interval are immaterial, provided only that they are continuous. All that matters for finding $P\{V; (d_1^*, d_2^*)\}$ is that the number and the order of the disjunct intervals shall be the same as those of the runs in V, (i.e., intervals of positive probability for X must alternate with intervals of positive probability for Y, the number of intervals of positive probability for X and for Y must equal respectively the number of runs of the element 0 and the number of runs of the element 1, and the probability of the first interval on the left shall be positive for X or for Y according as the first run in V is of elements 0 or of elements 1, with the same relation obtaining between the last interval on the right and the last run in V) and the probability of these intervals. Let P_{1j} be the sought for probability of the interval which corresponds to the jth run of elements 0 and P_{2j} the probability of the interval which corresponds to the jth run of elements 1. In order to obtain V, it is necessary that the elements constituting each run shall fall into its corresponding interval. Then clearly by the multinomial theorem

$$P\{V; (d_1^*, d_2^*)\} = \prod_i n_i! \left(\prod_j (l_{ij}!)^{-1} P_{ij}^{l_{ij}}\right) \tag{5.2}$$

where $i = 1, 2$ and where, when i is fixed, the product with respect to j is taken over all runs of the corresponding element. The right member of (5.2) is to be maximized with respect to the P_{ij}, subject of course to the constraints

$$\sum_j P_{ij} = 1 \qquad (i = 1, 2). \tag{5.3}$$

Then it may easily be verified that the maximum occurs when

$$P_{ij} = \frac{l_{ij}}{n_i} \qquad (i = 1, 2) \tag{5.4}$$

For, after multiplying by a constant and taking the logarithm we introduce two Lagrange multipliers μ_1 and μ_2 so that the maximizing P_{ij} are given by the equations (5.3) and those obtained by equating to zero all the partial derivatives of

$$\sum_i \sum_j (l_{ij} \log P_{ij} - \mu_i P_{ij}).$$

The latter are therefore

$$\mu_i = \frac{l_{ij}}{P_{ij}} \qquad (i = 1, 2),$$

for all j, whence (5.4) follows. It is easy to see that the extremum thus obtained is a maximum and also an absolute maximum. The sought-for statistic $T(V)$ is then the right member of (5.2) after the results (5.4) have been inserted. It may be simplified by removing all factors which are functions only of n_1 and n_2 (since these will then be the same for all V) and recalling that

$$\sum_j l_{ij} = n_i \qquad (i = 1, 2). \tag{5.5}$$

It will be convenient to take the logarithm of the resulting expression, so that with a slight change of notation we finally have

$$T(V) = \sum_i \sum_j \bar{l}_{ij} \tag{5.6}$$

where

$$\bar{l}_{ij} = \log \left(\frac{l_{ij}^{l_{ij}}}{l_{ij}!} \right). \tag{5.7}$$

This result is immediately extensible to the problem of k samples and by way of summary we recapitulate it as follows:

Let there be given k stochastic variables $X_1, \cdots, X_k$ with the respective distribution functions $f_1(x), \cdots, f_k(x)$, about which nothing is known except that they are continuous. Random independent observations, n_i in number, are made on X_i $(i = 1, \cdots, k)$. It is desired to test the hypothesis that $f_1 \equiv f_2 \equiv \cdots \equiv f_k$, the admissible alternatives being all k-tuples of continuous distribution functions. The sequence V is obtained from the sequence Z by replacing an observation on X_i by the element i. Let l_{ij} be the number of elements in the jth run of elements i. Then the corresponding statistic for testing the null hypothesis is $T_k(V)$ or any monotonic function of it, where

$$T_k(V) = \sum_{i=1}^{k} \sum_{j} \bar{l}_{ij}$$

and $\bar{l}_{ij}$ is given by (5.7). The large values of $T_k(V)$ are the critical values.

Let r_{ij} denote the number of runs of length j in the elements i. Let $\sum_j r_{ij} = r_i$. Of course $\sum_j jr_{ij} = n_i$. Also let $s_j = \sum_i r_{ij}$. Then

$$T_k(V) = \sum_i \sum_j \bar{j} r_{ij} \tag{5.8}$$

and

$$T_k(V) = \sum_j \bar{j} s_j . \tag{5.9}$$

If a table were constructed of the numbers (5.7) from 1 to 50, say, or from 1 to 100, this would cover most of the cases arising in practice. The calculation of $T_k(V)$ by means of (5.9) would then be so simple that it could be performed very expeditiously by an ordinary clerk and with very much less labor than is required for most statistics in common use, like the correlation coefficient, for example. As a matter of interest we note that

$$\bar{1} = 0$$

$$\bar{2} = .693$$

$$\bar{3} = 1.50$$

$$\bar{4} = 2.37$$

$$\bar{5} = 3.26$$

and that

$$\bar{p} < p \tag{5.10}$$

where p is any integer ≥ 1. (5.10) follows from the fact that

$$p! > (\sqrt{2\pi p} - 1)p^p e^{-p}.$$

The distribution of $T(V)$ may be found for small samples by enumerating the sequences V, all of which have the same probability under the null hypothesis, and assigning to each V its $T(V)$. The critical region consists of the V's for which $T(V)$ takes the largest values, taken in sufficient number to make the critical region of proper size. It will not be necessary to enumerate all the V's, since it is readily apparent that certain V's can never belong to a critical region of any reasonable size. (Roughly speaking, a V with a large number of runs of short length will yield a small $T(V)$ and vice versa.) For large samples, the result of Section 3 is available, with $f(x) = \bar{x}$. From (5.10) it follows easily that the corresponding series (3.9) is convergent, so that $\overline{T}(V)$ is asymptotically normally distributed. It must be remembered when using tables of the normal distribution that the critical region of $\overline{T}(V)$ lies in only one "tail" of the normal curve. The greatest difficulty will occur for samples of moderate size. Methods like those of Olds [7] will probably help there. It is highly unlikely that any practicable formula which would give the exact distribution of $T(V)$ exists.

A few brief remarks may be made here on a related problem. Suppose we have observations from two bivariate populations about the distributions of both of which nothing is known except that they are continuous and it is sought to test whether the two populations have the same distribution functions. Suppose further that it were required that the statistic used for this purpose be invariant under any topologic transformation of the whole plane into itself. At this point we quote the following topologic theorem, the proof of which was communicated to the author by Dr. Herbert Robbins: *Let* $x_1, y_1, x_2, y_2, \cdots, x_p, y_p$ *be any* $2p$ *distinct points in the plane. There exists a topologic transformation of the whole plane into itself which takes* x_i *into* y_i $(i = 1, 2, \cdots, p)$. As a consequence of this theorem we get the absurd result that the required statistic must be a constant. Hence this statistical problem can have no solution.

As a matter of interest this statistical problem would have no solution even if it were not for the topologic theorem. The fact is that a continuous distribution on a line remains continuous under a topologic transformation of the whole line into itself, but a continuous distribution in a k-dimensional (Euclidean) space $(k > 1)$ may become discontinuous under a topologic transformation of the whole space into itself. (The probability distribution in the first space always determines a probability distribution in the transformed space, for probability functions are defined over all Borel sets of the space (cf. [15], p. 7) and a topologic transformation carries Borel sets into Borel sets (cf. [16], p. 195, Theorem II)). Consider the following example in the plane: A bivariate distribution function assigns probability 1 to a line L oblique to the coordinate axes, while any interval which contains no segment of the line L has probability 0. On the line L the (one-dimensional) probability distribution may be arbitrary, provided it is continuous. The bivariate distribution function is without difficulty seen to be continuous. Now rotate the coördinate axes until one of them is parallel to L. It is easy to see that after the rotation the bivariate distribution function is discontinuous.

The question of whether a useful statistical problem could be obtained by properly delimiting the class of transformations which are to leave the statistic invariant and the solution of such a problem remain to be investigated.

6. The problem of the independence of several variates. This is an important practical problem and one of the earliest discussed in the literature (cf., for example, [6]). Let X_1 and X_2 be stochastic variables with the joint (cumulative) distribution function $F(x_1, x_2)$ which is known to be continuous in both variables jointly (i.e., $F(x_1, x_2) = P\{X_1 < x_1 ; X_2 < x_2\}$, where the right member is the probability of the occurrence of *both* the relations in braces). The marginal distributions $f_1(x_1)$ and $f_2(x_2)$ of X_1 and X_2 respectively are defined as follows:

$$f_1(x_1) = P\{X_1 < x_1\} = \lim_{x_2 \to +\infty} F(x_1, x_2),$$

$$f_2(x_2) = P\{X_2 < x_2\} = \lim_{x_1 \to +\infty} F(x_1, x_2).$$

(It is easy to see that the continuity of $F(x_1, x_2)$ implies the continuity of $f_1(x_1)$ and $f_2(x_2)$.)

The n random, independent pairs of observations $x_{11}, x_{21}, \cdots x_{1n}, x_{2n}$ are made on X_1 and X_2. The null hypothesis states that

$$F(x_1, x_2) \equiv f_1(x_1) \cdot f_2(x_2) \tag{6.1}$$

i.e., that X_1 and X_2 are independent. The alternative hypotheses are that $F(x_1, x_2)$ does not satisfy (6.1).[1]

Let the set $x_{11}, x_{12}, x_{13}, \cdots, x_{1n}$ be arranged in order of ascending size, thus: $Z = z_1, z_2, z_3, \cdots, z_n$ where $z_1 < z_2 < \cdots < z_n$. The jth member of this sequence will be said to have the rank j. In the same manner ranks are assigned to the x_{2j} $(j = 1, \cdots, n)$. (It is easy to see that, since $f_1(x_1)$ and $f_2(x_2)$ are continuous, the probability that $z_j = z_{j+1}$ is 0 etc.) In the sequence Z the element z_j $(j = 1, \cdots, n)$ is replaced by the rank of its associated observation on X_2. We obtain a permutation of the integers $1, 2, \cdots, n$ which we denote by R. If in the procedure for obtaining R, we had reversed the roles of the x_{1j} and x_{2j}, we would have obtained the permutation R'. It is easy to see that any statistic, say M'', used to test the null hypothesis, must be a function only of R, with the added proviso that $M''(R) = M''(R')$. (The rank correlation coefficient is such a statistic.) Under the null hypothesis all the R have the same probability $\left(= \frac{1}{n!}\right)$.

The procedure of applying the likelihood ratio principle to this problem would then be as follows: Ω is the totality of all bivariate distribution functions $H(x_1, x_2)$ which are continuous in both variables jointly. The respective marginal distributions corresponding to $H(x_1, x_2)$ will be denotedby $h_1(x_1)$ and $h_2(x_2)$. ω is a subset of Ω which consists of all $H(x_1, x_2)$ for which $H(x_1, x_2) \equiv h_1(x_1) \cdot h_2(x_2)$. The sample space is the totality of all sequences R. The null hypothesis states that $F(x_1, x_2)$ is a member of ω. The admissible alternatives are that $F(x_1, x_2)$ is a member of Ω not in ω. The distribution of any function of R is the same for all members of ω. Thus the essential requirement for the applicability of the likelihood ratio criterion is fulfilled. All sequences R have the same probability for all members of ω; hence the numerator of the likelihood ratio is a func-

[1] It is easy to see that the independence or dependence of two stochastic variables is not a property which will remain invariant under a topologic transformation of the plane into itself. We therefore require of the statistic only that it be invariant under topologic transformation of *each* variable into itself, separately.

tion only of n which may therefore be ignored. We may then define $M'(R)$, a monotonic function of the likelihood ratio as the denominator of the likelihood ratio, thus:

(6.2) $$M'(R) = \max_{\Omega} P\{R; H(x_1, x_2)\}$$

where $P\{R; H(x_1, x_2)\}$ is the probability of R when $H(x_1, x_2)$ is the joint distribution function of X_1 and X_2. The critical values of $M'(R)$ are the large values.

We now propose an approximation to $M'(R)$ which we shall call $M(R)$. We do this by describing a distribution function $H^*(x_1, x_2)$ for each R which seems a plausible approximation to a maximizing distribution function. It may be derived from certain assumptions about the nature of the maximizing distribution function which we omit. The remarks made in the preceding section about the character of the approximation apply here as well. As before we specify only the form of the function and leave certain parameters, finite in number, to be determined in accordance with (6.2). (If the construction of $H^*(x_1, x_2)$ should appear somewhat involved, this is due only to the analytic description. A sketch will show the essential simplicity of the situation.) We then have

$$M(R) = P\{R; H^*(x_1, x_2)\}.$$

Let $R = a_1, a_2, \cdots, a_n$ be a given permutation of the integers 1 to n. A sub-sequence $a_{(i+1)}, a_{(i+2)}, \cdots, a_{(i+l)}$ will be called a run of length l if the following conditions are fulfilled:

(6.3) The indices of the a's are consecutive,

(6.4) If l' is any integer such that $1 \leq l' < l$, then

$$| a_{(i+l')} - a_{(i+l'+1)} | = 1,$$

(6.5) if $i > 0$, $| a_i - a_{(i+1)} | > 1$,

(6.6) if $i + l < n$, $| a_{(i+l)} - a_{(i+l+1)} | > 1$.

The run will be called an ascending run or a descending run according as $a_{(i+1)} - a_{(i+2)} = -1$ or $+1$. A run of length 1 is of either type, at pleasure. For example, let

$$R = 5, 6, 1, 4, 3, 2.$$

The first run is 5, 6, the second 1, the last 4, 3, 2. 5, 6 is an ascending run of length two, 4, 3, 2 a descending run of length three, and 1 a run of length one.

$H^*(x_1, x_2)$ is a degenerate distribution function such that the relation between X_1 and X_2 is functional (this is a special case of stochastic relationship). That is to say, $X_2 = \varphi(X_1)$, where $\varphi(X_1)$ is a single-valued function defined for all the possible values of X_1, with a single-valued inverse $\varphi^{-1}(X_2)$ defined for all possible values of X_2. Hence $H^*(x_1, x_2)$ is completely specified when the function $X_2 = \varphi(X_1)$ and $h_1^*(x_1)$ the marginal distribution function of X_1, are given ($h_1^*(x_1)$ must of course be continuous).

Consider a system of intervals on the line $-\infty < x_1 < +\infty$ of which $(i - 1, i)$

is the ith, $i = 1, 2, \cdots, n$ and a similar system on the line $-\infty < x_2 < +\infty$. (Actually, as in the previous section, neither the length of the intervals nor their location is material. The intervals need merely be disjunct and in a certain order. We are using these particular intervals to simplify the notation.) Let l_1 be the length of the first run. a_1 is its first element. Then let

$$p_1 = P\{0 \leq X_1 \leq l_1 ; h_1^*(x_1)\}$$

be one of the as yet undetermined parameters. We now partly define $h_1^*(x_1)$ as follows:

$$(6.7) \qquad \begin{aligned} h_1^*(x_1) &= 0, & x_1 &\leq 0 \\ h_1^*(x_1) &= 1, & x_1 &\geq n \\ h_1^*(l_1) &= p_1. \end{aligned}$$

Within the interval $(0, l_1)$, $h_1^*(x_1)$ may be any continuous monotonic increasing function which satisfies (6.7). We partly define $\varphi(X_1)$ as follows:
If the first run is ascending, let

$$(6.8) \qquad \varphi(0) = a_1 - 1$$

$$(6.9) \qquad \varphi(x_1) = a_1 - 1 + x_1, \qquad 0 < x_1 \leq l_1.$$

If the first run is descending, let

$$(6.10) \qquad \varphi(0) = a_1$$

$$(6.11) \qquad \varphi(x_1) = a_1 - x_1, \qquad 0 < x_1 \leq l_1.$$

We proceed in this manner through all the runs of R. Let l_i be the length of the ith run. Let $\lambda_j = \sum_{i<j} l_i$. The first element of the jth run is $a_{(\lambda_j+1)}$. Let

$$p_j = P\{\lambda_j < X_1 \leq \lambda_j + l_j ; h_1^*(x_1)\},$$

be another of the as yet undetermined parameters. We then define $h_1^*(x_1)$ as follows:

$$(6.12) \qquad h_1^*(\lambda_j) = \sum_{i<j} p_i$$

$$(6.13) \qquad h_1^*(\lambda_j + l_j) = \sum_{i \leq j} p_i.$$

Within the interval $(\lambda_j, \lambda_j + l_j)$, $h_1^*(x_1)$ may be any continuous monotonic increasing function which satisfies (6.12) and (6.13). We define $\varphi(X_1)$ as follows:
If the jth run is ascending, let

$$(6.14) \qquad \varphi(x_1) = a_{(\lambda_j+1)} - 1 + x_1 \qquad (\lambda_j < x_1 \leq \lambda_j + l_j).$$

If the jth run is descending, let

$$(6.15) \qquad \varphi(x_1) = a_{(\lambda_j+1)} - x_1 \qquad (\lambda_j < x_1 \leq \lambda_j + l_j).$$

If $l_j = 1$, the run may be considered ascending or descending at pleasure.

In order to obtain R, it is necessary that all the elements of a run shall fall into its corresponding interval. Then it is easy to see that by the multinomial theorem

$$P\{R; H^*(x_1, x_2)\} = n! \prod_i (l_i!)^{-1} p_i^{l_i}. \tag{6.16}$$

The right member of (6.16) is to be maximized with respect to the p_i subject to the constraint

$$\Sigma p_i = 1. \tag{6.17}$$

It is easy to verify that the maximum occurs when

$$p_i = \frac{l_i}{n}. \tag{6.18}$$

$M(R)$ is the right member of (6.16) after the results (6.18) have been inserted. It is convenient to remove all factors which are functions only of n and to take the logarithm of the resulting expression. Then with a slight change of notation we may say that

$$M(R) = \sum_i \bar{l}_i \tag{6.19}$$

where

$$\bar{l}_i = \log \left(\frac{l_i^{l_i}}{l_i!} \right). \tag{6.20}$$

The critical values of $M(R)$ are the large values. One may verify without much difficulty that $M(R) = M(R')$, i.e., that the statistic is symmetric with respect to X_1 and X_2 as indeed it should be.

This result is immediately extensible to the problem of testing whether k stochastic variables $X_1, \cdots, X_k$ are independent. We shall not go into the details, which are similar to those described above, and content ourselves with giving the definition of a run for the case $k = 3$. After the observations on X_1 have been arranged in ascending order, we obtain two sequences R_2 and R_3, the associated ranks of the observations on X_2 and X_3. Let $R_2 = b_1, b_2, \cdots, b_n$ and $R_3 = b'_1, b'_2, \cdots, b'_n$. The ascending sequence of consecutive integers $(i + 1), (i + 2), \cdots, (i + l)$ determines a run of length l if the sequences $b_{(i+1)}, b_{(i+2)}, \cdots, b_{(i+l)}$ and $b'_{(i+1)}, b'_{(i+2)}, \cdots, b'_{(i+l)}$ *both* satisfy (6.4), and if at least one of the sequences satisfies (6.5), and at least one, but not necessarily the same one, satisfies (6.6). The adjectives ascending and descending apply to each sequence separately.

Let r_j be the number of runs of length j in R. Then

$$M(R) = \sum_j \jmath r_j. \tag{6.21}$$

Most of the remarks made in Section 5 about the small sample distribution of $T(V)$ are also applicable to the distribution of $M(R)$. More will be said in the next section about the distribution of $M(R)$ which involves the solution of a combinatorial problem not discussed in the literature.

7. On the distribution of $W(R)$. While most of the remarks made about the small sample distribution of $T(V)$ apply to the question of the distribution of $M(R)$ in small samples, the situation with respect to the distribution of $M(R)$ in samples of medium size and large size is very different and, in certain respects, is more favorable for practical application than is the case with $T(V)$. It would be reasonable to expect, for example, in view of Section 3 and of the structure of the statistic $M(R)$ that the asymptotic distribution of $M(R)$ should be normal. Surprisingly enough, this is not the case. It is not even continuous. In order to clarify the situation, we begin with a few necessary ideas and definitions.

Let the stochastic variable $W(R)$ be defined as the total number, in R, of runs of the sense of Section 6. We shall be interested in the distribution of $W(R)$. The number n of the pairs of observations on X_1 and X_2 (we consider the case of two variates) will be assumed arbitrary but fixed throughout the discussion and will not be exhibited. Let $N(k)$ be the number of sequences R (of the integers 1 to n) which contain exactly k runs.

Consider, for example, for the case $n = 6$, the sequence 2 3 4 6 5 1. We shall say that this sequence contains the "contacts" (2, 3), (3, 4), (6, 5). In general, a contact is defined as the juxtaposition, in the sequence R, of consecutive numbers, whether in ascending or descending order. If k is the number of runs and l the number of contacts in a sequence R, then obviously

$$k + l = n. \tag{7.1}$$

Let R_0 be the sequence $1, 2, \cdots, n$ of the first n integers in ascending order. The $n - 1$ contacts of this sequence may themselves be arranged in a sequence R^* of contacts, thus:

$$(1, 2), (2, 3), \cdots, (n - 1, n).$$

Suppose l of the contacts which constitute the sequence R^* are selected in some manner to form the set O. The remaining $n - 1 - l$ contacts form the complementary set O'. After this selection the sequence R^* may be considered a sequence of the type of the sequences V of Section 5 with the members of O playing the role of the elements 0 and the members of O' playing the role of the elements 1. When R^* is considered in this manner we will write it as $R^*(O)$. The definition of a run of Section 5 as applied to sequences V is now applicable to $R^*(O)$. We will call any such run of the members of O or of O' a group.

We wish first to answer the following question: In how many ways can the set O be selected from among the elements of R^* so that it will contain l members arranged in $R^*(O)$ in i groups? If, for a given O, i' be the number of groups into which O' is divided in $R^*(O)$, it is clear that $i - i'$ can equal only -1, 0, or $+1$. Hence only four situations can arise, as follows:

a) $i' = i + 1$. The first group in $R^*(O)$ is therefore composed of elements of

O'. The number of ways in which l elements can be divided into i runs of the type of Section 2 is the coefficient of x^l in the purely formal expansion of

$$(x + x^2 + x^3 + \cdots)^i = \left(\frac{x}{1 - x}\right)^i$$

and is therefore $\binom{l-1}{i-1}$. Similarly $n - 1 - l$ elements can be divided into $i' = i + 1$ runs in $\binom{n-l-2}{i}$ ways. Hence this situation will arise in $\binom{l-1}{i-1}\binom{n-l-2}{i}$ ways.

b) $i' = i - 1$. By a similar argument as above, this can occur in $\binom{l-1}{i-1}\binom{n-l-2}{i-2}$ ways.

c) $i' = i$ and the first group is made up of elements from O. This will occur in $\binom{l-1}{i-1}\binom{n-l-2}{i-1}$ ways.

d) $i' = i$ and the first group is made up of elements from O'. This will also occur in $\binom{l-1}{i-1}\binom{n-l-2}{i-1}$ ways.

The set O which contains l elements arranged in i groups can therefore be selected in

$$\binom{l-1}{i-1}\left(\binom{n-l-2}{i} + \binom{n-l-2}{i-2} + 2\binom{n-l-2}{i-1}\right) \tag{7.2}$$

ways, and the quantity (7.2) is, by elementary combinatorics, equal to

$$\binom{l-1}{i-1}\binom{n-l}{i}. \tag{7.3}$$

Let any set O of l contacts divided into i groups be selected from R^*. Imagine that each contact in O sets up, in R_0, an unbreakable bond which links the two elements involved in the contact, but no contact in O' creates such a bond. Given these bonds set up by O, we seek the number of different sequences into which the n elements of R_0 can be permuted while respecting these bonds. Since there are l bonds, we can actually manipulate only $n - l$ entities, except that two elements linked by a bond may have their order reversed; for example, if O contains (1, 2), 1 may either precede or follow 2 and the bond would still be respected. However, if one contact in a group is reversed, the group as a whole must be reversed, else a bond would be broken. Hence the number of distinct sequences into which R_0 may be permuted while all the bonds set up by O are respected is $2^i(n - l)!$.

Let us refer to the sequences thus obtained as the family generated by O. All the sequences in a family are distinct. Now let O range over all sets of l

contacts selected from R^*. The various families obtained will not be disjunct, but some will have sequences in common. In spite of this, we seek the total of the number of sequences in all the families. The total of the number of sequences in all the families generated by sets of l contacts divided into i groups is, by (7.3) and the result of the preceding paragraph,

$$2^i \cdot \binom{l-1}{i-1} \binom{n-l}{i} (n-l)!. \tag{7.4}$$

Sets of l contacts may consist of 1, 2, $\cdots$ l groups, so that the total number of sequences in all the families generated by sets of l contacts is

$$A_l = \sum_{i=1}^{l} 2^i \binom{l-1}{i-1} \binom{n-l}{i} (n-l)! \tag{7.5}$$

where l may take the values 1, 2, $\cdots$, $(n-1)$. The conventions on the combinatorial symbols will be:

$$\binom{a}{0} = 1, \qquad a \geq 0,$$

$$\binom{a}{b} = 0, \qquad a < b.$$

Define A_0 as

$$A_0 = n!. \tag{7.6}$$

The following equation is trivial:

$$A_0 = \sum_{i=1}^{n} N(i). \tag{7.7}$$

We now consider all the families generated by sets O which contain exactly l contacts. As was said before, the total of the number of sequences in each is A_l. Let $H(l)$ be the set of all the sequences in all these families, with each sequence in $H(l)$ counted as many times as the number of families in which it occurs. Every sequence in $H(l)$ has the l contacts of the set O which generated it, but after permuting R_0 other contacts may still exist. Hence every sequence in $H(l)$ has at least l contacts and therefore by (7.1), at most $n - l$ runs. Clearly, a sequence which has exactly l contacts occurs exactly once in $H(l)$, since it can appear only in the family generated by the set O of its l contacts and in no other family. A sequence which has exactly $(l + 1)$ contacts will appear exactly $\binom{l+1}{l}$ times in $H(l)$, for it will appear once in each family generated by a set O which consists of one of the $\binom{l+1}{l}$ selections of l contacts from among its $(l + 1)$ contacts, and in no other family. Similarly each sequence which has exactly $(l + 2)$ contacts will appear in $H(l)$ $\binom{l+2}{l}$ times, and so forth. We therefore obtain, in view of (7.1),

$$A_l = \sum_{i=l}^{n-1} \binom{i}{l} N(n-i) \qquad (l = 1, 2, \cdots, (n-1)). \tag{7.8}$$

The system of n linear equations (7.7) and (7.8) completely determines the quantities $N(1), N(2), \cdots, N(n)$. The matrix of these equations has a determinant whose absolute value is 1, so that the quantities $N(1), N(2), \cdots, N(n)$ may readily be expressed in determinantal form. Furthermore the moments of $W(R)$ are readily found from these equations. Thus from (7.8) for $l = 1$ we find

$$E(W(R)) = \frac{n^2 - 2n + 2}{n} \sim n - 2 \tag{7.9}$$

and from (7.8) for $l = 2$ and $l = 1$ we find, after a little obvious manipulation,

$$\sigma^2(W(R)) = \frac{2n^3 - 8n^2 + 6n + 4}{n^3 - n^2} \sim 2. \tag{7.10}$$

Higher moments of $W(R)$ may be found in similar manner.

Since the limiting variance of $W(R)$ is 2 it follows that the asymptotic distribution is not continuous. For n of any size the bulk of the values are concentrated in a short interval ending at n. When $W(R) = n$, $M(R) = 0$, when $W(R) = n - 1$, $M(R) = \log 2$, and when $W(R) = n - 2$, $M(R) = \log 4\frac{1}{2}$ or $\log 4$. It is easy to see that for the values of $W(R)$ which differ very little from n there are only a small number of values of $M(R)$, whose asymptotic distribution is also discontinuous. The statistic $W(R)$ is therefore a good approximation to the statistic $M(R)$ for the purposes of tests of significance (for $M(R)$ the large values are the critical values and for $W(R)$ the small values are critical), and has a few additional practical advantages. It is even easier to compute than $M(R)$; the computation is best performed by counting contacts. Since the limiting variance is a small constant, it follows that many tests of significance can be performed simply by use of Tchebycheff's inequality. For example, suppose a given large sample contains 9 contacts, i.e., $n - 9$ runs (we say a "large" sample in order to use the simple limiting mean and variance; if desired or for a small sample these latter may be computed exactly by (7.9) and (7.10)). Then by Tchebycheff's inequality it follows that the probability of obtaining $n - 9$ or fewer runs is less than .041. Thus the presence of 9 contacts would be sufficient to render a sample of great size significant on a 5% level. For the few numbers of contacts about which doubt will exist as to whether or not they are critical values two procedures are possible. Either the equations (7.7) and (7.8) may be solved exactly for the doubtful values, or several higher moments may be found from (7.8) and the methods of Wald [14] can be applied to delimit the missing probabilities to any accuracy desired. By enumerating the few values of $M(R)$ which correspond to several of the largest values of $W(R)$ the distribution of $M(R)$ may be computed sufficiently to serve the purposes of tests of significance.

REFERENCES

[1] MAURICE FRECHET, *Gênêralitês sur les Probabilitês, Variables Aleatoires*, Paris (1937).
[2] A. WALD and J. WOLFOWITZ, *Annals of Math. Stat.*, Vol. 11 (1940), p. 147.
[3] A. M. MOOD, *Annals of Math. Stat.*, Vol. 11 (1940), p. 367.
[4] J. NEYMAN, *Phil. Trans. Roy. Soc. London*, Vol. 231 (1937), pp. 333–380.
[5] A. WALD and J. WOLFOWITZ, *Annals of Math. Stat.*, Vol. 10 (June, 1939), p. 105.
[6] H. HOTELLING and M. PABST, *Annals of Math. Stat.*, Vol. 7 (1936), p. 29.
[7] E. G. OLDS, *Annals of Math. Stat.*, Vol. 9 (1938), p. 133.
[8] W. R. THOMPSON, *Annals of Math. Stat.*, Vol. 9 (1938), p. 281.
[9] W. J. DIXON, *Annals of Math. Stat.*, Vol. 11 (June, 1940), p. 199.
[10] MILTON FRIEDMAN, *Jour. Amer. Stat. Assoc.*, Vol. 32 (1937), p. 675.
[11] J. NEYMAN and E. PEARSON, *Trans. Royal Soc.*, A., Vol. 231 (1933), p. 295.
[12] A. WALD, *Bull. Amer. Math. Soc.*, Vol. 46 (1940), p. 235.
[13] A. WALD, *Bull. Amer. Math. Soc.*, Vol. 47 (1941), p. 396.
[14] A. WALD, *Trans. Amer. Math. Soc.*, Vol. 46 (1939), p. 280.
[15] HARALD CRAMER, *Random Variables and Probability Distributions*, Cambridge (1937).
[16] F. HAUSDORFF, *Mengenlehre* (Second Edition), Berlin and Leipzig, 1927.