THE LARGE-SAMPLE POWER OF TESTS BASED ON PERMUTATIONS OF OBSERVATIONS[1]

By Wassily Hoeffding

University of North Carolina

Summary. The paper investigates the power of a family of nonparametric tests which includes those known as tests based on permutations of observations. Under general conditions the tests are found to be asymptotically (as the sample size tends to ∞) as powerful as certain related standard parametric tests. The results are based on a study of the convergence in probability of certain random distribution functions. A more detailed summary will be found at the end of the Introduction.

1. Introduction. Let X be a random variable whose values are points x in a space $\mathfrak{X}$. The probability distribution of X is characterised by the probability measure $P(A) = \Pr\{X \varepsilon A\}$, defined on an additive class $\mathfrak{A}$ of subsets A of $\mathfrak{X}$. (In the applications to be considered $\mathfrak{X}$ can be taken as a finite-dimensional Euclidean space, $\mathfrak{A}$ as the family of Borel sets.) Let $\mathfrak{G}$ be a finite group of transformations g of $\mathfrak{X}$ onto itself which also map $\mathfrak{A}$ onto itself. Thus, for every g in $\mathfrak{G}$, every x in $\mathfrak{X}$ and every A in $\mathfrak{A}$, the point gx is in $\mathfrak{X}$ and the set gA of points gx, $x \varepsilon A$, is in $\mathfrak{A}$. Let M be the number of elements in $\mathfrak{G}$. Let H be a hypothesis which implies that the distribution of X is invariant under the transformations in $\mathfrak{G}$, so that for every g in $\mathfrak{G}$, gX has the same distribution as X.

For example, let $\mathfrak{X}$ be the n-dimensional Euclidean space, and let H be the hypothesis that the components $X_1, \cdots, X_n$ of X are independent, each X_i being symmetrically distributed about the median 0. Then H implies that the distribution of X is invariant under changes of sign of the X_i. Here $M = 2^n$. Alternatively, if $X_1, \cdots, X_m, \cdots, X_n$ are independent, $X_1, \cdots, X_m$ have a common distribution and $X_{m+1}, \cdots, X_n$ have a common distribution, then the distribution of X is invariant under the $M = m!\,(n-m)!$ permutations which permute the first m or the last $n - m$ components.

All real-valued functions of x to be considered are understood to be measurable ($\mathfrak{A}$). The expected value of a function $f(X)$ when X has distribution P will be denoted by $E_P f(X)$ or $Ef(X)$.

By a test of H we shall mean a function $\phi(x)$, $0 \leqq \phi(x) \leqq 1$, which expresses the probability with which H is rejected when X takes the value x. The power of the test ϕ with respect to P (the unconditional probability of rejecting H when P is the true distribution and test ϕ is used) is equal to $E_P\phi(X)$. If $E_P\phi(X) = \alpha$ whenever H is true, the test ϕ is said to be similar of size α for testing H.

This paper will be mainly concerned with tests of the following type. Let $t(x)$ be a real-valued function on $\mathfrak{X}$. For every $x \varepsilon \mathfrak{X}$ let

$$t^{(1)}(x) \leqq t^{(2)}(x) \leqq \cdots \leqq t^{(M)}(x)$$

[1] Work done under the sponsorship of the Office of Naval Research

be the ordered values $t(gx)$, for all g in $\mathcal{G}$. Given a number α, $0 < \alpha < 1$, let k be defined by

$$k = M - [M\alpha],$$

where $[M\alpha]$ denotes the largest integer less than or equal to $M\alpha$. Let $M^+(x)$ and $M^0(x)$ be the numbers of values $t^{(j)}(x)$, $(j = 1, \cdots, M)$, which are greater than $t^{(k)}(x)$ and equal to $t^{(k)}(x)$, respectively, and let

$$a(x) = \frac{M\alpha - M^+(x)}{M^0(x)}.$$

Since $M^+(x) \leqq M - k \leqq M\alpha$ and $M^+(x) + M^0(x) \geqq M - k + 1 > M\alpha$, we have $0 \leqq a(x) < 1$.

Let the test $\phi(x)$ be defined by

$$\phi(x) = \begin{cases} 1 & \text{if } t(x) > t^{(k)}(x), \\ a(x) & \text{if } t(x) = t^{(k)}(x), \\ 0 & \text{if } t(x) < t^{(k)}(x). \end{cases} \tag{1.1}$$

For every $x \varepsilon \mathcal{X}$ we have

$$\textstyle\sum_g \phi(gx) = M^+(x) + a(x)M^0(x) = M\alpha,$$

where $\sum_g$ stands for summation over all g in $\mathcal{G}$. If the distribution P of X is invariant under all g in $\mathcal{G}$, we have

$$M\alpha = E_P \textstyle\sum_g \phi(gX) = \sum_g E_P\,\phi(X) = ME_P\,\phi(X).$$

Hence the test ϕ is similar of size α for testing H.

Tests which are essentially of the form (1.1) have been considered by R. A. Fisher [3], Pitman [11], Welch [14]. Lehmann and Stein [8] have shown that tests of this type, with suitable functions $t(x)$, are most powerful (or most powerful similar, etc.) for testing certain nonparametric hypotheses H against specified alternatives.

A test of the form (1.1) differs from a conventional test mainly in that the "critical value," $t^{(k)}(X)$, is a random variable. This circumstance makes the exact evaluation of its power function difficult. It will, however, be shown that under certain conditions $t^{(k)}(X)$ is close to a constant with high probability. Then the power of the test can be approximated in terms of the distribution function of $t(X)$.

More precisely, suppose that the objects so far considered, $\mathcal{X} = \mathcal{X}_n$, $\mathcal{G} = \mathcal{G}_n$, $t(x) = t_n(x)$, etc., are defined for an infinite sequence of positive integers n. It will be assumed that the size α of the test is fixed and that $M \to \infty$ as $n \to \infty$. Then

$$k/M \to 1 - \alpha \qquad \text{as } n \to \infty.$$

Suppose that for a given sequence $\{P_n\}$ of distributions of $X = X^{(n)}$ the following two conditions are satisfied:

CONDITION A. *There exists a constant* λ *such that* $t_n^{(k)}(X) \to \lambda$ *in probability.*

CONDITION B. *There exists a function* $H(y)$, *continuous at* $y = \lambda$, *such that for every* y *at which* $H(y)$ *is continuous*

$$\Pr\{t_n(X) \leqq y\} \to H(y).$$

From (1.1) we have

$$\Pr\{t_n(X) > t_n^{(k)}(X)\} \leqq E_{P_n}\phi_n(X) \leqq \Pr\{t_n(X) \geqq t_n^{(k)}(X)\}. \tag{1.2}$$

Hence it follows that Conditions A, B imply

$$E_{P_n}\phi_n(X) \to 1 - H(\lambda). \tag{1.3}$$

It should be noted that the function $t(x)$ in the definition (1.1) of $\phi(x)$ can be replaced by any function $t'(x)$ such that for every x in $\mathfrak{X}$ and every two elements g, g' of $\mathcal{G}$ the difference $t'(gx) - t'(g'x)$ has the same sign as $t(gx) - t(g'x)$. For example, this is true for $t'(x) = c(x)f(t(x)) + d(x)$, where $f(y)$ is an increasing function, $c(x) > 0$, and $c(x)$, $d(x)$ are invariant under $\mathcal{G}$ (cf. Lehmann and Stein [8]). Thus if Conditions A, B are not satisfied, they may possibly be satisfied after $t_n(x)$ has been replaced by a suitable function $t'_n(x)$.

In general λ and $H(y)$ will depend on the sequence $\{P_n\}$. It will, however, be seen that the dependence of λ on $\{P_n\}$ is much less pronounced than that of $H(y)$, in the sense that for a class C of sequences $\{P_n\}$ the value λ is the same while $1 - H(\lambda)$ ranges from α to 1.

For every x in $\mathfrak{X}$ let $MF_n(y, x)$ be the number of elements g in $\mathcal{G}$ for which $t_n(gx) \leqq y$. For x fixed, $F_n(y, x)$ is a distribution function. Suppose that for some sequence $\{P_n\}$ the following condition is satisfied:

CONDITION A′. $F_n(y, X) \to F(y)$ *in probability for every* y *at which* $F(y)$ *is continuous, where* $F(y)$ *is a distribution function, the equation* $F(y) = 1 - \alpha$ *has a unique solution* $y = \lambda$, *and* $F(y)$ *is continuous at* $y = \lambda$.

It will be shown in Section 3 that A′ implies that $t_n^{(k)}(X) \to \lambda$ in probability, so that A is satisfied with λ as defined in A′; furthermore, if H is true for every P_n of the sequence, $t_n(X)$ has the limiting distribution function $F(y)$.

Let ϕ_n^* be a test of the conventional form $\phi_n^*(x) = 1$, a_n^*, or 0 according as $t_n(x) > \lambda_n$, $= \lambda_n$, or $< \lambda_n$, where $0 \leqq a_n^* \leqq 1$ and λ_n is a constant. Suppose that λ_n and a_n^* are so chosen that the test ϕ_n^* has size α for testing that $P_n = P_n^*$, a distribution for which H is true. It follows from the preceding paragraph that if A′ is satisfied for $\{P_n^*\}$, then $\lambda_n \to \lambda$. Moreover, if B holds,

$$E_{P_n}\phi_n^*(X) \to 1 - H(\lambda). \tag{1.4}$$

Hence if $C(\lambda)$ denotes the class of all sequences $\{P_n\}$ for which A′, with λ fixed, and B, with some $H(y)$, are satisfied, and if $C(\lambda)$ contains $\{P_n^*\}$, then the powers of the tests ϕ_n and ϕ_n^* tend to the same limit for every $\{P_n\}$ in $C(\lambda)$. The nonparametric test ϕ_n can be said to be asymptotically as powerful with respect to $C(\lambda)$ as ϕ_n^*. This result will be of particular interest when ϕ_n^* is a most powerful, or otherwise "optimum," parametric test, as in the examples of this paper.

It also can happen that for different sequences $\{P_n\}$, $t^{(k)}(X)$ converges to different values λ, but in every case the test ϕ_n is asymptotically as powerful as the most powerful test for a parametric family of distributions to which P_n belongs. This point will be illustrated in Section 7.

In most applications to be considered, $H(y)$ is either a (cumulative) distribution function, or $H(y) \equiv 0$. In the latter case the relations (1.3) and (1.4) merely imply that both tests are consistent (have limiting power 1). The case $0 < H(\lambda) < 1$ will usually occur when P_n approaches, in a certain sense, the null hypothesis. For example, let P_n be the distribution of two independent random samples of m and $n - m$ observations from two normal distributions with means $\mu_1 \leqq \mu_2$ and common variance σ^2. Let $\mathcal{G}$ consist of the $M = n!$ permutations of the n observations. Let $t_n(x)$ be the standard t-statistic for two samples. The results of Section 6 imply that Condition A′ is satisfied with $F(y) = \Phi(y)$, where

$$\Phi(y) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{y} e^{-\frac{1}{2}t^2}\, dt. \tag{1.5}$$

Condition B is satisfied with $H(y) = \Phi(y - c)$ if $(\mu_2 - \mu_1)\,\sigma^{-1}\cdot\{m(n - m)/n\}^{\frac{1}{2}}$ tends to a finite limit c. This will not be the case if, as is frequently assumed, $m/n \to p$, $0 < p < 1$, and $\delta = (\mu_2 - \mu_1)\sigma^{-1}$ is independent of n. In this case one can, however, conclude that if δ is sufficiently small, the number N of observations required to achieve the power $1 - \Phi(\lambda - c)$ is approximately given by $\delta\{p(1 - p)N\}^{\frac{1}{2}} = c$, and this is true for either test. In this sense the asymptotic relative efficiency of the two tests is arbitrarily close to one for δ sufficiently small.

The main object of this paper is to indicate several methods for ascertaining that Condition A is satisfied. By way of illustration the methods are applied to a number of tests which have been considered in the literature. In Section 2 bounds for $t^{(k)}(x)$ are obtained which provide a simple criterion for consistency. Sufficient conditions for the convergence to zero of the variance of the random variable $F_n(y, X)$ are given (Section 3) and used to obtain the large-sample power of several tests (Sections 4–7). The remaining Sections 8–10 show how a theorem can be applied which gives sufficient conditions for the convergence of $F_n(y, x^{(n)})$, for a sequence of fixed values $x^{(n)}$. The fulfilment of these conditions in probability for a sequence of random variables $X^{(n)}$ is found to be sufficient for the convergence in probability of $F_n(y, X^{(n)})$. An extension to random distributions of the second limit theorem of probability theory (Section 10) generalizes a recent result of Ghosh [6].

2. Bounds for $t^k(x)$; consistency. In this section it will be shown that, given a test $\phi(x)$ of the form (1.1), the function $t(x)$ can always be so chosen that one or two moments of the distribution function $F_n(y, x)$ are (essentially) fixed for all x, and the critical value $t^{(k)}(x)$ is confined to a finite interval which depends only on α.

Let G be a random variable whose values are the M elements g of $\mathcal{G}$, each element having the same probability M^{-1}. Then $F_n(y, x)$, as defined in Section 1, is the distribution function of the random variable $t(Gx)$.

Let $m(x)$ and $v(x)$ denote the mean and the variance of $t(Gx)$, so that

$$m(x) = M^{-1} \textstyle\sum_g t(gx), \qquad v(x) = M^{-1} \textstyle\sum_g [t(gx) - m(x)]^2.$$

Let $t'(x) = v(x)^{-\frac{1}{2}}[t(x) - m(x)]$ if $v(x) > 0$, $t'(x) = 0$ if $v(x) = 0$. Then the test $\phi(x)$ in (1.1) is not changed if $t(x)$ is replaced by $t'(x)$. Thus we may always assume that the distribution function $F_n(y, x)$ has mean 0 and variance less than or equal to 1. If a probability limit $F(y)$ of $F_n(y, X)$ exists for all y, then $F(y)$ is a distribution function with the same properties. If, moreover, the probability of $t(gX) = t(X)$ for all g in $\mathcal{G}$ tends to 0 as $n \to \infty$, then the probability of $v(X) = 0$ tends to 0, and $F(y)$ has variance 1. In a similar way, if $t(x) \geqq 0$, we may, for instance, replace $t(x)$ by a function $t'(x)$ such that $t'(x) \geqq 0$ and $Et'(Gx) = c$, an arbitrary positive constant.

THEOREM 2.1. *If* $t(x) \geqq 0$, $Et(Gx) = c > 0$, *then*

$$t^{(k)}(x) < \frac{c}{\alpha}. \tag{2.1}$$

If $Et(Gx) = 0$, $Et(Gx)^2 \leqq 1$, *then*

$$-\left(\frac{\alpha}{1-\alpha}\right)^{\frac{1}{2}} \leqq t^{(k)}(x) < \left(\frac{1-\alpha}{\alpha}\right)^{\frac{1}{2}}. \tag{2.2}$$

PROOF. We have

$$MF(t^{(k)}(x) - 0, x) \leqq k - 1 < M - M\alpha \leqq k \leqq MF_n(t^{(k)}(x), x),$$

so that

$$F_n(t^{(k)}(x) - 0, x) < 1 - \alpha \leqq F_n(t^{(k)}(x), x).$$

If $t(x) \geqq 0$, $Et(Gx) = c$, then for every $z > 0$

$$1 - F_n(z - 0, x) = \Pr\{t(Gx) \geqq z\} \leqq \frac{c}{z}.$$

Hence (2.1).

If $Et(Gx) = 0$, $Et(Gx)^2 = c^2 \leqq 1$, relation (2.2) follows in a similar way by using the inequalities of Tchebycheff-Cantelli (see, e.g., [4], p. 126 or [12], p. 198)

$$F_n(y, x) \leqq \frac{1}{1 + c^{-2} y^2} \qquad \text{if } y < 0,$$

$$F_n(y - 0, x) \geqq 1 - \frac{1}{1 + c^{-2} y^2} \qquad \text{if } y > 0.$$

Apart from providing, via (1.2), crude bounds for the power of ϕ, Theorem 2.1 permits us to draw the following conclusion. If $t_n(x)$ satisfies either of the conditions of the theorem and, for some sequence $\{P_n\}$ of distributions, $H(y) = \lim \Pr\{t_n(X) \leqq y\} = 0$ for all real y, which is a sufficient condition for consistency of the tests ϕ_n^*, then the tests ϕ_n are also consistent. This result is independent of whether $t_n^{(k)}(X)$ converges in probability to a constant.

3. Sufficient conditions for the convergence in probability of $t_n^{(k)}(X)$.

THEOREM 3.1. *Suppose that for a sequence $\{P_n\}$ of distributions of $X = X^{(n)}$, $F_n(y, X)$ tends in probability to $F(y)$ for every y at which $F(y)$ is continuous, where $F(y)$ is a distribution function and the equation $F(y) = 1 - \alpha$ has a unique solution $y = \lambda$. Then $t_n^{(k)}(X) \to \lambda$ in probability.*

PROOF. By the definitions of $t_n^{(k)}(x)$ and $F_n(y, x)$,

$$\Pr\{t_n^{(k)}(X) \leqq y\} = \Pr\{F_n(y, X) \geqq k/M\} \tag{3.1}$$

for every real y. Let y be a point of continuity of $F(y)$. Since, by assumption, $k/M \to 1 - \alpha = F(\lambda)$, and $y < \lambda$ implies $F(y) < F(\lambda)$, the right-hand side of (3.1) tends to 0 if $y < \lambda$. Similarly it tends to 1 if $y > \lambda$. Hence $t_n^{(k)}(X) \to \lambda$ in probability.

A sufficient condition for a sequence of random variables to converge in probability to a constant c is that their means and variances converge, respectively, to c and 0. If the random variables are uniformly bounded, the condition is also necessary. Hence $F_n(y, X) \to F(y)$ in probability if and only if

$$EF_n(y, X) \to F(y), \qquad EF_n(y, X)^2 \to F(y)^2. \tag{3.2}$$

We can write

$$F_n(y, x) = M^{-1} \sum_g C(gx),$$

where $C(x) = 1$ or 0 according as $t_n(x) \leqq y$ or $> y$. Hence

$$EF_n(y, X) = M^{-1} \sum_g \Pr\{t_n(gX) \leqq y\}, \tag{3.3}$$

$$EF_n(y, X)^2 = M^{-2} \sum_g \sum_{g'} \Pr\{t_n(gX) \leqq y, t_n(g'X) \leqq y\}. \tag{3.4}$$

Let G be the random transformation defined in Section 2, let G' have the same distribution as G, and let G, G' and X be mutually independent. Then equations (3.3), (3.4) can be written as

$$EF_n(y, X) = \Pr\{t_n(GX) \leqq y\}, \tag{3.5}$$

$$EF_n(y, X)^2 = \Pr\{t_n(GX) \leqq y, t_n(G'X) \leqq y\}. \tag{3.6}$$

Note that $t_n(GX)$ and $t_n(G'X)$ are identically distributed, but not independent (except in the trivial case when the random variable $F_n(y, X)$ has variance 0). Equations (3.5) and (3.6) imply that (3.2) is satisfied if $t_n(GX)$ has the limiting distribution function $F(y)$, and $t_n(GX)$ and $t_n(G'X)$ are independent in the limit. Making use of Theorem 3.1, we can state

THEOREM 3.2. *Suppose that, for some sequence $\{P_n\}$ of distributions, $t_n(GX)$ and $t_n(G'X)$ have the limiting joint distribution function $F(y)F(y')$. Then for every y at which $F(y)$ is continuous*

$$F_n(y, X) \to F(y) \textit{ in probability},$$

and if the equation $F(y) = 1 - \alpha$ has a unique solution $y = \lambda$,

$$t_n^{(k)}(X) \to \lambda \textit{ in probability}.$$

We also observe the following. If H is true, $t_n(GX)$ and $t_n(X)$ have the same distribution. Thus if $F_n(y, X) \to F(y)$ in probability for a sequence of distributions invariant under $\mathcal{G}$, then $t_n(X)$ has the limiting distribution $F(y)$. An implication concerning the test ϕ_n^* was pointed out in the Introduction.

The next theorem, 3.3, gives conditions under which two functions $t_n(x)$ and $t'_n(x)$ are, in a certain sense, asymptotically equivalent.

THEOREM 3.3. *Let* $t'_n(x) = c_n(x)t_n(x) + d_n(x)$, *where*

$$c_n(GX) \to 1 \text{ and } d_n(GX) \to 0 \text{ in probability,} \tag{3.7}$$

and let $F'_n(y, x) = \Pr\{t'_n(Gx) \leqq y\}$. *Then*

$$F_n(y, X) \to F(y) \text{ in probability} \tag{3.8}$$

if and only if

$$F'_n(y, X) \to F(y) \text{ in probability.} \tag{3.9}$$

PROOF. It is sufficient to show that (3.8) implies (3.9). As has been seen, (3.8) is equivalent to $\Pr\{t_n(GX) \leqq y\} \to F(y)$, $\Pr\{t_n(GX) \leqq y, t_n(G'X) \leqq y\} \to F(y)^2$. Due to assumption (3.7) these relations remain true if $t_n(x)$ is replaced by $t'_n(x)$. This implies (3.9).

The fulfilment of the conditions of Thorem 3.2 can frequently be demonstrated with the aid of the central limit theorem for vectors. One version of this theorem, which will be of particular use in Section 6, is stated below as Theorem 3A. It easily follows from Uspensky's proof [12] of the central limit theorem for vectors.

THEOREM 3A. *Let* $(Y_1, Y'_1), (Y_2, Y'_2), \cdots, (Y_n, Y'_n)$ *be n independent random vectors,* $EY_i = EY'_i = 0$, $E|Y_i|^3 < \infty$, $E|Y'_i|^3 < \infty$. *Let*

$$\tilde{Y} = \sum_1^n Y_i\left(\sum_1^n EY_i^2\right)^{-\frac{1}{2}}, \qquad \tilde{Y}' = \sum_1^n Y'_i\left(\sum_1^n EY_i'^2\right)^{-\frac{1}{2}},$$

$$\rho = E\tilde{Y}\tilde{Y}',$$

$$\omega = \sum_1^n E|Y_i|^3\left(\sum_1^n EY_i^2\right)^{-3/2}, \qquad \omega' = \sum_1^n E|Y'_i|^3\left(\sum_1^n EY_i'^2\right)^{-3/2}.$$

Then for any two real numbers y, y'

$$|\Pr\{\tilde{Y} \leqq y, \tilde{Y}' \leqq y'\} - \Phi(y)\Phi(y')| \leqq f(\rho, \omega, \omega'),$$

where $\Phi(y)$ *is defined by* (1.5) *and the function* $f(u, v, w)$ *is independent of* n, y, y' *and of the distribution of the* Y_i, Y'_i, *and* $f(u, v, w) \to 0$ *as* $u \to 0, v \to 0, w \to 0$.

4. Test for the median of a symmetrical distribution. Let $\mathfrak{X}$ be the Euclidean n-dimensional space and H the hypothesis that the components $X_1, \cdots, X_n$ of the random vector X are independent and each X_i is symmetrically distributed about the median 0. H implies that the distribution of X is invariant under the $M = 2^n$ transformations $gX = ((-1)^{j_1}X_1, \cdots, (-1)^{j_n}X_n)$, $j_i = 0$ or $1, i = 1$,

$\cdots, n$. The random transformation Gx of x can be written $Gx = (G_1x_1, \cdots, G_nx_n)$, where $G_1, \cdots, G_n$ are independent, $G_i = -1$ or 1 with probabilities $\frac{1}{2}, \frac{1}{2}$. Let $\phi(x)$ be the test (1.1) with

$$t(x) = \sum_1^n x_i \left(\sum_1^n x_i^2\right)^{-\frac{1}{2}},$$

or $t(x) = 0$ if $\sum_1^n x_i^2 = 0$. The factor $(\sum x_i^2)^{-\frac{1}{2}}$ is invariant under the transformations g and is so chosen that $t(Gx)$ has mean 0 and variance 1 (unless $x_1 = \cdots = x_n = 0$). Bounds for $t^{(k)}(x)$ can be obtained from Theorem 2.1.

It follows from the results of Lehmann and Stein [8] that the test ϕ is most powerful similar for testing H against the alternative that $X_1, \cdots, X_n$ are independent with a common normal distribution whose mean is positive; the test ϕ with $t(x)$ replaced by $|t(x)|$ is most stringent similar for testing H against the alternative of a common normal distribution with nonzero mean. It will suffice to consider the former, "one-sided" test. The results will be easily applicable to the "two-sided" case.

Let $Y_i = G_iX_i$, $Y_i' = G_i'X_i$, where all G_i, G_i' are independent, identically distributed, and independent of the X_i. Then $Y_i^2 = Y_i'^2 = X_i^2$,

$$t(GX) = n^{-\frac{1}{2}} \sum_{i=1}^n Y_i \left(n^{-1} \sum_{i=1}^n X_i^2\right)^{-\frac{1}{2}},$$

$$t(G'X) = n^{-\frac{1}{2}} \sum_{i=1}^n Y_i' \left(n^{-1} \sum_{i=1}^n X_i^2\right)^{-\frac{1}{2}}.$$

Suppose that $X_1, \cdots, X_n$ are independent and identically distributed with mean μ and positive variance σ^2. By Khintchine's theorem,

$$n^{-1} \sum_1^n X_i^2 \rightarrow \sigma^2 + \mu^2$$

in probability. Hence $(t(GX), t(G'X))$ has the same limiting distribution (if any) as

$$\left((\sigma^2 + \mu^2)^{-\frac{1}{2}} n^{-\frac{1}{2}} \sum_1^n Y_i, \quad (\sigma^2 + \mu^2)^{-\frac{1}{2}} n^{-\frac{1}{2}} \sum_1^n Y_i'\right). \tag{4.1}$$

The vectors $(Y_1, Y_1'), \cdots, (Y_n, Y_n')$ are independent and identically distributed, with

$$EY_i = EY_i' = 0, EY_i^2 = EY_i'^2 = \sigma^2 + \mu^2, EY_iY_i' = EG_iG_i'X_i^2 = EG_iEG_i'EX_i^2 = 0.$$

By the central limit theorem for identically distributed vectors (see, e.g., Cramér [2], p. 286), the random vector (4.1) has the limiting distribution function $\Phi(y)\Phi(y')$. The same is true of $(t(GX), t(G'X))$. By Theorem 3.2, $t^{(k)}(X) \rightarrow \lambda$ in probability, where $\Phi(\lambda) = 1 - \alpha$.

Under the same conditions we have for every fixed y

$$\lim_{n\rightarrow\infty} \Pr\{t(X) \leqq (y + n^{\frac{1}{2}}\mu/\sigma)(1 + (\mu/\sigma)^2)^{-\frac{1}{2}}\} = \Phi(y).$$

Hence if μ/σ is independent of n (as is implied in the assumptions) and positive, the function $H(y)$ of Section 1 is $\equiv 0$, and the power of the test tends to 1. It follows from the Lyapunov form of the central limit theorem and its extension to vectors (for example, Theorem 3A) that all results remain true if the common distribution of $X_1, \cdots, X_n$ depends on n, provided $E \mid X_1 \mid^3 \sigma^{-3} = o(n^{\frac{1}{2}})$. If $(\mu/\sigma)n^{\frac{1}{2}}$ converges to a constant c, then $H(y) = \Phi(y - c)$. An alternative interpretation of this result, with μ/σ fixed but small, is indicated in the Introduction.

The function $t(x)$ is an increasing function of Student's statistic for testing whether the mean of n independent random variables with a common normal distribution is zero. Thus the test ϕ_n^* of Section 1, with suitably chosen λ_n, is equivalent to Student's (one-sided) test whose size (for testing the normal hypothesis) is equal to the size α of the test ϕ. The two tests have the same limiting power under the alternatives considered.

Similar results can be obtained for more general alternatives, for instance when the X_i are not identically distributed, provided only the central limit theorem can be applied.

5. An analysis of variance test. Let $\mathfrak{X}$ be a Euclidean space of np dimensions. Let $X = (X_1, \cdots, X_n)$ where $X_i = (X_{i1}, \cdots, X_{ip})$, $i = 1, \cdots, n$, are n independent random vectors of $p \geqq 2$ components, and let H be a hypothesis which implies that the distribution of each X_i is invariant under the $p!$ permutations of its components. Then the distribution of X is invariant under a group $\mathcal{G}$ of $M = (p!)^n$ permutations. For example, if in an agricultural experiment p treatments are randomly assigned to the p plots in each of n blocks, and X_{ij} is the yield of the plot in the ith block which has received the jth treatment, hypothesis H may be assumed to hold when there is no difference in the treatment effects.

Let the test $\phi(x)$ be defined by (1.1) with

$$t(x) = \frac{\sum_{j=1}^{p} \left(\sum_{i=1}^{n} (x_{ij} - x_{i.}) \right)^2}{\sum_{i=1}^{n} (p-1)^{-1} \sum_{j=1}^{p} (x_{ij} - x_{i.})^2},$$

where $x_{i.} = p^{-1} \sum_{j=1}^{p} x_{ij}$. If the denominator vanishes, define $t(x) = p - 1$ (say). The denominator, which is invariant under permutations in $\mathcal{G}$, is so chosen that $Et(Gx) = p - 1$ for all x.

In the traditional analysis of variance one assumes that the X_{ij} are independent normal with common variance and means $EX_{ij} = b_i + t_j$. The equivalent of hypothesis H is that $t_1 = \cdots = t_p$. The usual F- (or z-) statistic for testing this hypothesis is an increasing function of $t(X)$.

A nonparametric test essentially equivalent to $\phi(x)$ was considered by Fisher [3] in the case $p = 2$, by Welch [14] and Pitman [11] in the general case.

Extending the customary alternative, suppose that

$$X_{ij} = Y_{ij} + b_i + t_j, \qquad i = 1, \cdots, n; \qquad j = 1, \cdots, p,$$

where the Y_{ij} are mutually independent and identically distributed,

$$EY_{ij} = 0, \qquad \text{var } Y_{ij} = \sigma^2 > 0,$$

and the b_i and t_j are constants. It will be assumed that p is fixed and $n \to \infty$.

We can write

$$t(x) = \frac{\sum_{j=1}^{p} u_j(x)^2}{n^{-1} \sum_{i=1}^{n} (p-1)^{-1} \sum_{j=1}^{p} (x_{ij} - x_{i.})^2},$$

where

$$u_j(x) = n^{-\frac{1}{2}} \sum_{i=1}^{n} (x_{ij} - x_{i.}), \qquad j = 1, \cdots, p.$$

Since

$$X_{ij} - X_{i.} = Y_{ij} - Y_{i.} + t_j - \bar{t},$$

where $\bar{t} = p^{-1} \sum_1^p t_j$, has a distribution independent of i, the random variables

$$(p-1)^{-1} \sum_{j=1}^{p} (X_{ij} - X_{i.})^2, \qquad i = 1, \cdots, n,$$

are independent and identically distributed with mean $\sigma^2(1 + \delta^2)$, where

$$\delta^2 = \sigma^{-2}(p-1)^{-1} \sum_{k=1}^{p} (t_j - \bar{t})^2.$$

It follows that

$$n^{-1} \sum_{i=1}^{n} (p-1)^{-1} \sum_{j=1}^{p} (X_{ij} - X_{i.})^2 \to \sigma^2(1 + \delta^2) \text{ in probability.}$$

The expression on the left is invariant under the permutations in $\mathcal{G}$. Hence if we let

$$t'(x) = \sigma^{-2}(1 + \delta^2)^{-1} \sum_{j=1}^{p} u_j(x)^2,$$

then $(t(GX), t(G'X))$ has the same limiting distribution as $(t'(GX), t'(G'X))$.

We have

$$u_j(x) = \sum_{k=1}^{p} (\delta_{jk} - p^{-1}) v_k(x), \qquad v_k(x) = n^{-\frac{1}{2}} \sum_{i=1}^{n} (x_{ik} - b_i),$$

where δ_{jk} is Kronecker's delta. Let

$$V_i = v_j(GX), \qquad V_j' = v_j(G'X).$$

Then the random vector $n^{\frac{1}{2}}V = n^{\frac{1}{2}}(V_1 \cdots, V_p, V'_1, \cdots, V'_p)$ is the sum of n independent random vectors, each of which has the distribution of

$$Z^* = (Z_{R_1}, \cdots, Z_{R_p}, Z_{R'_1}, \cdots, Z_{R'_p}),$$

where $Z_1, \cdots, Z_p$ are independent, Z_j has the distribution of $Y_{ij} + t_j$, and $(R_1, \cdots, R_p)$ and $(R'_1, \cdots, R'_p)$ are two independent random vectors, independent of the Z_j, whose values are the $p!$ equally probable permutations of $(1, \cdots, p)$. By the central limit theorem for sums of identically distributed vectors, the limiting distribution of $V - EV$ is $2p$-variate normal with the covariance matrix of Z^*. We have

$$EZ^m_{R_j} = Ep^{-1} \sum_{k=1}^{p} Z^m_k, \quad j = 1, \cdots, p; \qquad m = 1, 2,$$

hence

$$EZ_{R_j} = \bar{t}, \qquad \text{var } Z_{R_j} = \sigma^2(1 + (1 - p^{-1})\delta^2).$$

If $j \neq j'$,

$$EZ_{R_j} Z_{R_{j'}} = Ep^{-1}(p-1)^{-1} \sum\sum_{k \neq k'} Z_k Z_{k'} = p^{-1}(p-1)^{-1} \sum\sum_{k \neq k'} t_k t_{k'}$$
$$= p(p-1)^{-1}\bar{t}^2 - p^{-1}(p-1)^{-1} \sum t_k^2,$$

hence

$$\text{cov }(Z_{R_j}, Z_{R_{j'}}) = -\sigma^2 p^{-1} \delta^2, \qquad j \neq j'.$$

The $Z_{R'_j}$ have the same distribution as the Z_{R_j}, and since $EZ_{R_j} Z_{R'_{j'}} = E(p^{-1}\sum Z_k)^2$ has the same value for all j, j', we have

$$\text{cov }(Z_{R_j}, Z_{R'_{j'}}) = C \text{ (say)}, \qquad j, j' = 1, \cdots, p.$$

Hence

$$EV_j = EV'_j = n^{\frac{1}{2}}\bar{t},$$
$$\text{var }(V_j) = \text{var }(V'_j) = \sigma^2(1 + (1 - p^{-1})\delta^2),$$
$$\text{cov }(V_j, V_{j'}) = \text{cov }(V'_j, V'_{j'}) = -\sigma^2 p^{-1}\delta^2, \qquad j \neq j',$$
$$\text{cov }(V_j, V'_{j'}) = C.$$

Let $\| c_{ij} \|$ be an orthonormal $p \times p$ matrix with $c_{p1} = \cdots = c_{pp}$, and let

$$W_j = \sum_{k=1}^{p} c_{jk} V_k, \qquad W'_j = \sum_{k=1}^{p} c_{jk} V'_k.$$

Then

$$\sum_{j=1}^{p} u_j(GX)^2 = \sum_{j=1}^{p-1} W_j^2, \qquad \sum_{j=1}^{p} u_j(G'X)^2 = \sum_{j=1}^{p-1} W'^2_j.$$

For $j, j' \leqq p - 1$ we obtain

$$EW_j = EW'_j = 0, \qquad EW_jW'_{j'} = 0,$$
$$EW_jW_{j'} = EW'_jW'_{j'} = \delta_{jj'}\sigma^2(1 + \delta^2).$$

Hence the limiting distribution of $(W_1, \cdots, W_{p-1}, W'_1, \cdots, W'_{p-1})$ is that of $2p - 2$ independent normal variables, each with mean 0 and variance $\sigma^2(1 + \delta^2)$. It follows that the limiting distribution of $(t'(GX), t'(G'X))$, and hence of $(t(GX), t(G'X))$, is that of $(\chi^2_{p-1}, \chi'^2_{p-1})$, where χ^2_{p-1} and χ'^2_{p-1} are independent, each having the chi-square distribution with $p - 1$ degrees of freedom. By Theorem 3.2, $t^{(k)}(X) \to \lambda$ in probability, where $\Pr\{\chi^2_{p-1} > \lambda\} = \alpha$. The test is asymptotically as powerful as the conventional analysis of variance test of the same size α.

6. Two-sample test; tests of randomness. Let $\mathfrak{X}$ be the n-dimensional Euclidean space, and let H be the hypothesis that the n components of $X = (X_1, \cdots, X_n)$ are independent and identically distributed. Then the distribution of X is invariant under all $M = n!$ permutations of its components. Let $\phi(x)$ be the test (1.1) with

$$t(x) = \frac{\sum_1^n (a_i - \bar{a})x_i}{\left\{\sum_1^n (a_i - \bar{a})^2(n - 1)^{-1} \sum_1^n (x_i - \bar{x})^2\right\}^{\frac{1}{2}}}, \tag{6.1}$$

where $a_1, \cdots, a_n$ are given numbers, not all equal, $\bar{a} = n^{-1} \sum_1^n a_i$, $\bar{x} = n^{-1} \sum_1^n x_i$. The numbers $a_i = a_{ni}$ may depend on n. If the denominator vanishes, that is, if $x_1 = \cdots = x_n$, define $t(x) = 0$. The denominator is invariant under all permutations, and is so chosen that $t(GX)$ has mean 0 and variance 1 (unless $x_1 = \cdots = x_n$).

If X has the probability density

$$(2\pi\sigma^2)^{-\frac{1}{2}n} \exp\left\{-(2\sigma^2)^{-1} \sum_1^n (x_i - a_i\xi - \eta)^2\right\} \tag{6.2}$$

and $T(x)$ denotes the standard t-statistic for testing $\xi = 0$, then

$$T(x) = (n - 2)^{\frac{1}{2}}t(x)(n - 1 - t(x)^2)^{-\frac{1}{2}},$$

so that $T(x)$ is an increasing function of $t(x)$.

Lehmann and Stein [8] have shown that the test $\phi(x)$ is most powerful similar for testing H against the alternative (6.2) with $\xi > 0$, and that the test based on $|t(x)|$ is most stringent similar against (6.2) with $\xi \neq 0$. If

$$a_i = 1 \text{ for } i = 1, \cdots, m; \qquad a_i = 0 \text{ for } i = m + 1, \cdots, n, \tag{6.3}$$

then (6.2) is the probability density of two independent random samples from two normal distributions with common variance and means $\xi + \eta$ and η, and the numerator of $t(x)$ is, apart from a constant factor, the difference of the two sample means. Essentially this test was proposed by Pitman [11].

We first consider a case where H is true.

THEOREM 6.1. *Let $t(x)$ be defined by (6.1), let $Z_1, \cdots, Z_n, \cdots$, be independent and identically distributed with $E \mid Z_1 \mid^3 < \infty$ and* var $Z_1 > 0$, *and let $Z = Z^{(n)} = (Z_1, \cdots, Z_n)$. Then in order that for every real y*

$$F_n(y, Z) \to \Phi(y) \text{ in probability} \tag{6.4}$$

it is necessary and sufficient that either Z_1 be normally distributed or

$$\frac{\max_{1 \leq i \leq n} (a_i - \bar{a})^2}{\sum_{i=1}^{n} (a_i - \bar{a})^2} \to 0. \tag{6.5}$$

Results similar to Theorem 6.1 were obtained by Wald and Wolfowitz [13] and Noether [9], who gave sufficient conditions (stronger than those of Theorem 6.1) for $F_n(y, Z) \to \Phi(y)$ with probability one, which, of course, implies (6.4). An argument analogous to that employed by Wald, Wolfowitz, and Noether will be used in Sections 8–10 below to obtain alternative sufficient conditions for (6.4).

PROOF OF THEOREM 6.1. We may and shall assume that

$$\bar{a} = 0, \qquad \sum_1^n a_i^2 = 1, \qquad EZ_1 = 0, \qquad EZ_1^2 = 1. \tag{6.6}$$

Then

$$t(x) = \sum_1^n a_i x_i \left\{ (n-1)^{-1} \sum_1^n (x_i - \bar{x})^2 \right\}^{-\frac{1}{2}}.$$

Since $\bar{Z} \to 0$ and $n^{-1} \sum_1^n Z_i^2 \to 1$ in probability we have

$$(n-1)^{-1} \sum_1^n (Z_i - \bar{Z})^2 \to 1$$

in probability. Hence $(t(GZ), t(G'Z))$ has the same limiting distribution (if any) as $(u(GZ), u(G'Z))$, where

$$u(x) = \sum_1^n a_i x_i .$$

Let $gx = x_r = (x_{r_1}, \cdots, x_{r_n})$, where $r = (r_1, \cdots, r_n)$ is a permutation of $(1, \cdots, n)$. If R and R' are two independent random vectors, independent of Z, such that $\Pr\{R = r\} = \Pr\{R' = r\} = M^{-1}$ for all r, we can write $(GZ, G'Z) = (Z_R, Z_{R'})$. For any two permutations r, r' we have

$$u(Z_r) = \sum_1^n a_i Z_{r_i} = \sum_1^n a_{s_i} Z_i,$$

$$u(Z_{r'}) = \sum_1^n a_i Z_{r'_i} = \sum_1^n a_{s'_i} Z_i,$$

where s_i and s'_i are defined by

$$r_{s_i} = i, \qquad r'_{s'_i} = i, \qquad\qquad i = 1, \cdots, n.$$

First suppose that Z_1 is normally distributed. Then $u(Z_r)$ and $u(Z_{r'})$ have a bivariate normal distribution with means 0, variances 1, and correlation coefficient

$$\rho_{rr'} = Eu(Z_r)u(Z_{r'}) = \sum_1^n a_{s_i} a_{s'_i}.$$

Thus

$$\Pr\{u(Z_r) \leqq y,\ u(Z_{r'}) \leqq y'\} = \Phi(y, y', \rho_{rr'}), \tag{6.7}$$

where

$$\Phi(y, y', \rho) = \int_{-\infty}^{y} \int_{-\infty}^{y'} (2\pi)^{-1}(1-\rho^2)^{-\frac{1}{2}} \exp\left\{-\frac{u^2 - 2\rho uv + v^2}{2(1-\rho^2)}\right\} du\, dv.$$

If both sides of (6.7) are summed over all r, r' and divided by M^2, we obtain

$$\Pr\{u(Z_R) \leqq y,\ u(Z_{R'}) \leqq y'\} = E\Phi(y, y', \rho_{RR'}). \tag{6.8}$$

The random variable $\rho_{RR'}$ has the same distribution as $\sum_1^n a_i a_{R_i}$, and we get $E\rho_{RR'}^2 = (n-1)^{-1}$. Hence $\rho_{RR'} \to 0$ in probability.

Since $\Phi(y, y', \rho) \to \Phi(y)\Phi(y')$ as $\rho \to 0$, we have $\Phi(y, y', \rho_{RR'}) \to \Phi(y)\Phi(y')$ in probability. And since $\Phi(y, y', \rho)$ is a bounded function, this implies

$$E\Phi(y, y', \rho_{RR'}) \to \Phi(y)\Phi(y').$$

Hence $\Phi(y)\Phi(y')$ is the limiting distribution function of $(u(Z_R), u(Z_{R'}))$, and also that of $(t(GZ), t(G'Z))$. Relation (6.4) follows from Theorem 3.2.

Next suppose that (6.5) is satisfied. By assumption (6.6) this condition is equivalent to

$$\max_{1 \leqq i \leqq n} |a_i| \to 0. \tag{6.9}$$

If we let $Y_i = a_{s_i} Z_i$, $Y'_i = a_{s'_i} Z_i$, the conditions of Theorem 3A are fulfilled, and we have $\bar{Y} = u(Z_r)$, $Y' = u(Z_{r'})$, $\rho = \rho_{rr'}$, $\omega = \omega' = E|Z_1|^3 c_n$, where

$$c_n = \sum_1^n |a_i|^3.$$

Hence

$$|\Pr\{u(Z_r) \leqq y,\ u(Z_{r'}) \leqq y'\} - \Phi(y)\Phi(y')| \leqq g(\rho_{rr'}, c_n), \tag{6.10}$$

where the function $g(u, v)$ is independent of n and of the distribution of the Y_i, Y'_i (in particular, independent of r, r'), and $g(u, v) \to 0$ as $u \to 0$, $v \to 0$. Clearly $g(u, v)$ can be so defined that $g(u, v) \leqq 1$ for all u, v.

From (6.10) we obtain in a similar way as before

$$|\Pr\{u(Z_R) \leqq y,\ u(Z_{R'}) \leqq y'\} - \Phi(y)\Phi(y')| \leqq Eg(\rho_{RR'}, c_n). \tag{6.11}$$

Since $c_n \leq \max | a_i | \sum_1^n a_i^2 = \max | a_i |$, condition (6.9) implies that $c_n \to 0$. Since $\rho_{RR'} \to 0$ in probability, $g(\rho_{RR'}, c_n) \to 0$ in probability; and since $g(u, v)$ is bounded, $Eg(\rho_{RR'}, c_n) \to 0$. Relation (6.4) now follows from (6.11) by Theorem 3.2.

Now suppose that Z_1 is not normal and (6.5) is not satisfied, the remaining assumptions of the theorem being fulfilled. Still assuming that the a_i satisfy (6.6), denote by A_n an a_j for which $| a_j | = \max (| a_1 |, \cdots, | a_n |)$. Then infinitely many $| A_n |$ are greater than a positive constant, and since the A_n are bounded, a subsequence $\{A_{n_m}\}$ of $\{A_n\}$ converges to a constant $A \neq 0$. We can write $u(Z) = u_n(Z) = V_n + W_n$, where V_n has the distribution of $A_n Z_1$ and is independent of W_n. As $m \to \infty$, V_{n_m} has the limiting distribution of AZ_1, which is not normal.

Suppose (6.4) were true. Then $t(Z_R)$, and hence $u_n(Z_R)$, would have the limiting distribution $\Phi(y)$. But Z_R has the same distribution as Z. It would follow that $u_{n_m}(Z)$ tends in distribution to a normal random variable which is the sum of two independent, nonnormal random variables. By a theorem of Cramér ([1], p. 52) this is impossible. The proof is complete.

In the sequel an extension by the author [7] of a theorem of Wald and Wolfowitz [13] will be required which, for purposes of reference, is stated below as Theorem 6A. For every positive integer n let $a = (a_1, \cdots, a_n)$, $b = (b_1, \cdots, b_n)$ be two vectors whose components a_i, b_i are real numbers and may depend on n. Suppose that the a_i are not all equal and the b_i are not all equal. Let the random vector $R = (R_1, \cdots, R_n)$ be defined as in the proof of Theorem 6.1, and let

$$F_n(y, a, b) = \Pr\left\{ \frac{(n-1)^{\frac{1}{2}} \sum_1^n (a_i - \bar{a}) b_{R_i}}{\left[\sum_1^n (a_i - \bar{a})^2 \sum_1^n (b_i - \bar{b})^2 \right]^{\frac{1}{2}}} \leq y \right\},$$

where $\bar{a} = n^{-1} \sum_1^n a_i$, $\bar{b} = n^{-1} \sum_1^n b_i$.

THEOREM 6A. *A sufficient condition for*

$$F_n(y, a, b) \to \Phi(y) \tag{6.12}$$

as $n \to \infty$ *is that*

$$n^{\frac{1}{2}p-1} \frac{\sum_1^n (a_i - \bar{a})^p}{\left[\sum_1^n (a_i - \bar{a})^2 \right]^{\frac{1}{2}p}} \frac{\sum_1^n (b_i - \bar{b})^p}{\left[\sum_1^n (b_i - \bar{b})^2 \right]^{\frac{1}{2}p}} \to 0, \quad p = 3, 4, \cdots. \tag{6.13}$$

Condition (6.13) *is satisfied if*

$$n \frac{\max_{1 \leq i \leq n} (a_i - \bar{a})^2}{\sum_1^n (a_i - \bar{a})^2} \frac{\max_{1 \leq i \leq n} (b_i - \bar{b})^2}{\sum_1^n (b_i - \bar{b})^2} \to 0. \tag{6.14}$$

The next theorem is concerned with the behavior of $t^{(k)}(X)$ under an alternative which generalizes (6.2).

THEOREM 6.2. *Let* $t(x)$ *be defined by* (6.1), *and suppose that*

$$X_i = Z_i + d_i, \qquad i = 1, \cdots, n,$$

where $Z_1, \cdots, Z_n, \cdots$ *are independent and identically distributed with*

$$E\,|\,Z_1\,|^3 < \infty$$

and $\operatorname{var} Z_1 > 0$, *and* $d_1, \cdots, d_n$ *are constants (which may depend on* n*). Then*

$$t^{(k)}(X) \to \lambda \text{ in probability,} \tag{6.15}$$

where $\Phi(\lambda) = 1 - \alpha$, *if*

$$Z_1 \text{ is normal or } \frac{\max_{1 \le i \le n} (a_i - \bar{a})^2}{\sum_1^n (a_i - \bar{a})^2} \to 0 \tag{6.16}$$

and

$$n^{\frac{1}{2}p - 1} \frac{\sum_1^n (a_i - \bar{a})^p}{\left[\sum_1^n (a_i - \bar{a})^2\right]^{\frac{1}{2}p}} \frac{\sum_1^n (d_i - \bar{d})^p}{\left[\sum_1^n (d_i - \bar{d})^2\right]^{\frac{1}{2}p}} \to 0, \qquad p = 3, 4, \cdots, \tag{6.17}$$

the latter condition being satisfied if

$$n \frac{\max_{1 \le i \le n} (a_i - \bar{a})^2}{\sum_1^n (a_i - \bar{a})^2} \frac{\max_{1 \le i \le n} (d_i - \bar{d})^2}{\sum_1^n (d_i - \bar{d})^2} \to 0. \tag{6.18}$$

Relation (6.15) *also holds if* (6.16) *is satisfied and*

$$n^{-1} \sum_1^n (d_i - \bar{d})^2 \to 0 \tag{6.19}$$

or if (6.17) *is satisfied and*

$$n^{-1} \sum_1^n (d_i - \bar{d})^2 \to \infty. \tag{6.20}$$

PROOF. We again make the simplifying assumptions (6.6). In addition we may set

$$\bar{d} = 0.$$

We then have

$$X_i - \bar{X} = Z_i - \bar{Z} + d_i,$$

$$n^{-1} \sum_1^n (X_i - \bar{X})^2 = n^{-1} \sum_1^n (Z_i - \bar{Z})^2 + D_n^2 + 2D_n s_n,$$

where

$$D_n = \left(n^{-1} \sum_1^n d_i^2\right)^{\frac{1}{2}}, \qquad s_n = \sum_1^n d_i Z_i \left(n \sum_1^n d_i^2\right)^{-\frac{1}{2}}.$$

We have $n^{-1} \sum_1^n (Z_i - \bar{Z})^2 \to 1$ in probability. Since $Es_n^2 = n^{-1}$, $s_n \to 0$ in probability. Also $0 \leqq 2D_n \leqq 1 + D_n^2$. Hence

$$\frac{n^{-1} \sum_1^n (X_i - \bar{X})^2}{1 + D_n^2} \to 1 \text{ in probability.}$$

Thus if we let

$$t'(x) = (1 + D_n^2)^{-\frac{1}{2}} \sum_1^n a_i x_i = (1 + D_n^2)^{-\frac{1}{2}} u(x),$$

then $(t(GX), t(G'X)) = (t(X_R), t(X_{R'}))$ has the same limiting distribution (if any) as $(t'(X_R), t'(X_{R'}))$.

Let

$$v(r) = \sum_1^n a_i d_{r_i} \left(n^{-1} \sum_1^n d_i^2\right)^{-\frac{1}{2}}.$$

Then

$$(6.21) \qquad t'(X_R) = \frac{u(Z_R) + D_n v(R)}{(1 + D_n^2)^{\frac{1}{2}}}, \qquad t'(X_{R'}) = \frac{u(Z_{R'}) + D_n v(R')}{(1 + D_n^2)^{\frac{1}{2}}}.$$

Suppose that conditions (6.16) and (6.17) are satisfied, and consider the joint distribution of $u(Z_R)$, $v(R)$, $u(Z_{R'})$, $v(R')$ as $n \to \infty$. It is seen from the proof of Theorem 6.1 that if (6.16) holds true and $p_n M^2$ denotes the number of pairs of permutations (r, r') for which $| \Pr \{u(Z_r) \leqq y,\ u(Z_{r'}) \leqq y'\} - \Phi(y)\Phi(y') |$ is less than a positive constant, then $p_n \to 1$ as $n \to \infty$. By the continuity theorem for the Fourier transform an analogous relation holds for the difference of the characteristic functions,

$$E \exp(itu(Z_r) + it'u(Z_{r'})) - \exp(-\tfrac{1}{2}t^2 - \tfrac{1}{2}t'^2).$$

Hence it follows that the characteristic function of $(v(R), v(R'), u(Z_R), u(Z_{R'}))$,

$$M^{-2} \sum_r \sum_{r'} \exp(i\tau v(r) + i\tau' v(r')) E \exp(itu(Z_r) + it'u(Z_{r'})),$$

differs arbitrarily little from

$$Ee^{i\tau v(R)} Ee^{i\tau' v(R')} e^{-\frac{1}{2}t^2 - \frac{1}{2}t'^2}$$

if n is sufficiently large. By Theorem 6A, condition (6.17) implies that $v(R)$ and $v(R')$ have the standard normal limiting distribution. Hence the limiting joint distribution of $v(R)$, $v(R')$, $u(Z_R)$, $u(Z_{R'})$ is that of four independent standard normal random variables. By (6.21) this implies that $(t'(X_R), t'(X_{R'}))$, and hence $(t(GX), t(G'X))$, has the limiting distribution function $\Phi(y)\Phi(y')$.

If (6.19) is satisfied, then $D_n \to 0$. Since $Ev(R)^2 = n(n-1)^{-1}$ is bounded, this implies that $D_n v(R) \to 0$ in probability, and $(t'(X_R), t'(X_{R'}))$ has the same limiting distribution as $(u(Z_R), u(Z_{R'}))$. When (6.16) holds, we can apply Theorem 6.1.

Similarly, if (6.20) is satisfied, $(t'(X_R), t'(X_{R'})$ has the limiting distribution of $(v(R), v(R'))$, which, under condition (6.17), is given by Theorem 6A. In every case the limiting distribution of $(t(GX), t(G'X))$ is $\Phi(y)\Phi(y')$, and relation (6.15) follows from Theorem 3.2. That condition (6.18) is sufficient for (6.17) is stated in Theorem 6A. This completes the proof.

If, in particular, X has the normal distribution (6.2), we have $d_i = a_i\xi + \eta$, and the conditions of Theorem 6.2 are fulfilled if either

$$n^{\frac{1}{2}(p-2)} \frac{\sum_1^n (a_i - \bar{a})^p}{\left[\sum_1^n (a_i - \bar{a})^2\right]^{\frac{1}{2}p}} \to 0, \qquad p = 3, 4, \cdots, \tag{6.22}$$

or

$$n^{\frac{1}{2}} \frac{\max_{1 \le i \le n} (a_i - \bar{a})^2}{\sum_1^n (a_i - \bar{a})^2} \to 0, \tag{6.23}$$

(which implies (6.22)), or

$$n^{-1} \sum_1^n (a_i - \bar{a})^2 \to 0. \tag{6.24}$$

In the two-sample case (6.3) the conditions (6.22) and (6.23) are both equivalent to

$$n(m')^{-2} = n^{-1}\left(\frac{m'}{n}\right)^{-2} \to 0,$$

where $m' = \min(m, n - m)$. Condition (6.24) is fulfilled if and only if

$$\frac{m'}{n} \to 0.$$

At least one of the two conditions is satisfied if m/n tends to some limit.

If the conditions of Theorem 6.2 up to and including (6.16) are satisfied, $t(X)$ is asymptotically normally distributed as $n \to \infty$. If the power of Student's (one-sided) test of size α tends to a limit, the power of ϕ tends to the same limit. Theorems 6.1 and 6.2 can be easily extended to the case where $Z_1, \cdots, Z_n$ have a common distribution which depends on n.

7. The two-sample test when one sample is small. It is of some interest to investigate what happens when the necessary and sufficient condition of Theorem 6.1 is not satisfied. In the two-sample case, which will be discussed in this section, this occurs only if m or $n - m$ does not tend to infinity with n.

We first consider a somewhat more general situation. Let $\mathfrak{X}$ be the Euclidean n-dimensional space and $\mathcal{G}$ the group of all $M = n!$ permutations of the n coordinates of a point in $\mathfrak{X}$. Let the components $X_1, \cdots, X_n$ of X be independent. The function $t(x)$ can be arbitrary, subject only to the conditions to be stated.

First assume that $t(x) = u(x_1, \cdots, x_m)$ is a function of $x_1, \cdots, x_m$ only, where m is fixed as $n \to \infty$. The proportion of pairs of permutations r, r' for which the sets $(r_1, \cdots, r_m)$ and $(r'_1, \cdots, r'_m)$ have no elements in common tends to 1 as $n \to \infty$. Hence $t(X_r)$ and $t(X_{r'})$ are independent for a proportion of pairs r, r' which converges to 1. Suppose now that $X_1, \cdots, X_m$ have a common distribution and $X_{m+1}, \cdots, X_n$ have a common distribution. Then for a proportion of permutations r which tends to 1, $t(X_r)$ has the distribution function of $u(X_{m+1}, \cdots, X_{2m})$, which will be denoted by $F(y)$. It follows that $(t(X_R), t(X_{R'}))$ has the limiting distribution function $F(y)F(y')$. If the equation $F(y) = 1 - \alpha$ has a unique solution $y = \lambda$, $t^{(k)}(X) \to \lambda$ in probability by Theorem 3.2.

The same conclusions hold under the more general assumption that $t(x)$ is of the form $c(x)u(x_1, \cdots, x_m) + d(x)$, where $c(X_R) \to 1$ and $d(X_R) \to 0$ in probability, as follows from Theorem 3.3.

Now let $t(x)$ be defined by (6.1) with the a_i given by (6.3). Then

$$t(x) = \left\{\frac{m(n-m)}{n(n-1)} \sum_1^n (x_i - \bar{x})^2\right\}^{-\frac{1}{2}} \left(\sum_1^m x_i - m\bar{x}\right).$$

Suppose that m is fixed, and that the common distribution of $X_{m+1}, \cdots, X_n$ has mean μ and variance σ^2. Then

$$\bar{X} = n^{-1} \sum_1^n X_i \to \mu, \; n^{-1} \sum_1^n (X_i - \bar{X})^2 \to \sigma^2$$

in probability. Hence the preceding results can be applied with

$$u(x_1, \cdots, x_m) = m^{-\frac{1}{2}} \sigma^{-1} \sum_1^m (x_i - \mu).$$

Observe that the probability limit λ of $t^{(k)}(X)$ depends on the distribution of X_{m+1}. Now it follows from [8] that the two-sample test ϕ is most powerful similar for testing H not only against the normal alternative (6.2), (6.3), but also against any alternative with a density of the form

$$\prod_{i=1}^m f(x_i, \theta_1) \cdot \prod_{i=m+1}^n f(x_i, \theta_2), \tag{7.1}$$

where

$$f(y, \theta) = A(\theta)B(y)e^{\theta y}, \; \theta_1 > \theta_2 .$$

On the other hand, the most powerful test of size α for testing that $X_1, \cdots, X_n$ are independent with the common density $f(y, \theta)$, where

$$\theta = (m\theta_1 + (n - m)\theta_2)/n,$$

against (7.1) is of the form $\phi^*(x) = 1$ or 0 according as $m^{-\frac{1}{2}} \sum_1^m (x_i - \bar{x}) > c_n$ or $< c_n$, where $\sigma^{-1} c_n$ converges to the probability limit λ of $t^{(k)}(X)$. In other words, $t^{(k)}(X)$ always tends in probability to the "correct" value λ, so that the test ϕ is asymptotically as powerful as the most powerful parametric test for the case where the function $f(y, \theta)$ is known. This phenomenon is analogous to the relation between, say, the one-sided two-sample t-test for normal distributions with unknown variance σ^2 and the most powerful tests (corresponding to the different values of σ^2) when σ^2 is known.

8. An alternative approach. In the remaining part of the paper an alternative method of proving that $F_n(y, X)$ tends to $F(y)$ in probability will be considered. It is an extension of an argument used by Wald and Wolfowitz [13].

Suppose that the quantities $a_i = a_{ni}$, $b_i = b_{ni}$ in Theorem 6A are random variables which have a joint distribution for all i and all n, and suppose that one of the conditions (6.13), (6.14) is satisfied with probability 1. Then (6.12) holds with probability 1.

For example, let $X = (U_1, V_1, U_2, V_2, \cdots, U_n, V_n)$, where the pairs (U_i, V_i), $i = 1, \cdots, n$, are independent and identically distributed, and let H be the hypothesis that U_i and V_i are independent. When H is true, the distribution of X is invariant under the $M = (n!)^2$ permutations which permute $(U_1, \cdots, U_n)$ or $(V_1, \cdots, V_n)$. Let

$$t(x) = \frac{(n-1)^{\frac{1}{2}} \sum_1^n (u_i - \bar{u})(v_i - \bar{v})}{\left\{ \sum_1^n (u_i - \bar{u})^2 \sum_1^n (v_i - \bar{v})^2 \right\}^{\frac{1}{2}}}.$$

A test equivalent to the corresponding test $\phi(x)$ was considered by Pitman [11].

Since $t(x)$ is invariant under permutations of the pairs (u_i, v_i), the distribution of $t(Gx)$ is the same as the conditional distribution with $u_1, \cdots, u_n$ held in a fixed order and only the v_i permuted in all possible ways. Hence if in Theorem 6A we let $a_i = u_i$, $b_i = v_i$, then $F_n(y, a, b)$ is identical with $F_n(y, x)$. Suppose that U_1 and V_1 have finite moments of any order. Then the strong law of large numbers implies that condition (6.13) of Theorem 6A, with $a_i = U_i$, $b_i = V_i$, is satisfied with probability 1. Hence $F_n(y, X) \to \Phi(y)$ with probability 1, and a fortiori in probability. Theorem 3.1 can now be applied.

That condition (6.13) is satisfied with probability 1 can be shown under weaker assumptions (cf. Noether [9], [10]). Since, however, only weak convergence (in probability) of $F_n(y, X)$ is required for our purposes, a proof of strong convergence seems redundant. In fact, it will be shown in Sections 9 and 10 that if the conditions of Theorem 6A are satisfied as limits in probability, then the conclusion holds as a limit in probability.

9. Ordinary convergence and convergence in probability. Let $f_n(x_n)$, $g_n(x_n)$, $n = 1, 2, \cdots$, be two sequences of real-valued functions of elements x_n in a space $\mathfrak{X}_n$. Let X_n denote a random variable with values in $\mathfrak{X}_n$, $n = 1, 2, \cdots$.

THEOREM 9.1. *If* $f_n(x_n) \to 0$ *implies* $g_n(x_n) \to 0$, *then* $f_n(X_n) \to 0$ *in probability implies* $g_n(X_n) \to 0$ *in probability.*

PROOF. Suppose the theorem were false. Then there exist two sequences of functions $\{f_n\}$, $\{g_n\}$ such that $f_n(x_n) \to 0$ implies $g_n(x_n) \to 0$, and a sequence of random variables $\{X_n\}$ such that $f_n(X_n) \to 0$ in probability but, for some $\delta > 0$ and some $\epsilon > 0$, $\Pr\{| g_n(X_n) | > \delta\} > \epsilon$ for infinitely many n. Let m be an arbitrary positive integer. Consider the events

$$A_n = \{| g_n(X_n) | > \delta\}, \qquad B_n^{(m)} = \{| f_n(X_n) | < m^{-1}\}.$$

We have $\Pr\{A_n\} > \epsilon$ for infinitely many n, and there exists a number N_m such that $\Pr\{B_n^{(m)}\} > 1 - \frac{1}{2}\epsilon$ for $n > N_m$. If $A_n \cdot B_n^{(m)}$ denotes the joint occurrence of A_n and $B_n^{(m)}$,

$$\Pr\{A_n \cdot B_n^{(m)}\} \geqq \Pr\{A_n\} + \Pr\{B_n^{(m)}\} - 1 > \epsilon + 1 - \tfrac{1}{2}\epsilon - 1 > 0$$

for infinitely many n.

Hence for every positive integer m there exists a sequence $\{x_n^{(m)}\}$, $x_n^{(m)} \varepsilon \mathfrak{X}_n$, such that $| f_n(x_n^{(m)}) | < m^{-1}$ for $n > N_m$ and $| g_n(x_n^{(m)}) | > \delta$ for infinitely many n. For every $m = 1, 2, \cdots$ there exists an integer $K_m \geqq N_{m+1}$ such that

$$| g_{K_m}(x_{K_m}^{(m)}) | > \delta$$

and $K_1 < K_2 < \cdots$. Let $K_0 = 0$,

$$x'_n = x_n^{(m)} \text{ for } n = K_{m-1} + 1, \cdots, K_m; \qquad m = 1, 2, \cdots$$

Then $| f_n(x'_n) | < m^{-1}$ for $n > K_m$, hence $f_n(x'_n) \to 0$, and $| g_n(x'_n) | > \delta$ for infinitely many n. But this contradicts the assumption.

Let, in particular, x_n be the vector (a, b) of Theorem 6A, f_n the left-hand side of (6.14) and $g_n = F_n(y, a, b) - \Phi(y)$. Then Theorem 9.1 shows that if a and/or b are replaced by random vectors, the fulfilment of (6.14) in probability implies that (6.12) holds in probability. Theorem 9.1 does not suffice to draw the same conclusion if the infinitely many relations (6.13) are satisfied in probability. That the conclusion is permissible will be shown in Section 10.

We conclude this section by stating, without proof, conditions which imply the fulfilment of (6.14) in probability. It can be shown that

$$\text{(9.1)} \qquad n^{1-(2/h)} \frac{\max_{1 \leqq i \leqq n} (X_i - \bar{X})^2}{\sum_1^n (X_i - \bar{X})^2} \to 0 \text{ in probability}$$

if $X_1, \cdots, X_n, \cdots$ are independent, identically distributed, $E | X_1 |^h < \infty$ for some $h \geqq 2$. Relation (9.1) with $h = 2$ also holds if $X_1, \cdots, X_n$ are independent with common mean and finite second moments and satisfy the Lindeberg condition of the central limit theorem. More generally, (9.1) holds if $EX_i = d_i$, the $X'_i = X_i - d_i$ satisfy one of the previously stated conditions and

$$n^{1-(2/h)} \frac{\max_{1 \leqq i \leqq n} (d_i - \bar{d})^2}{\sum_1^n (d_i - \bar{d})^2} \to 0.$$

Hence one can obtain alternative sufficient conditions for $t_n^{(k)}(X) \to \lambda$ in probability in the examples of Sections 6 and 8. Thus in the case of Section 8 it is sufficient that U_i and V_i have finite moments of order 4.

10. The second limit theorem for random distributions. A generalization by Fréchet and Shohat [5] of Markov's so-called second limit theorem of probability theory states that if the distribution function $F(y)$ is uniquely determined by its moments and $\{F_n(y)\}$ is a sequence of distribution functions whose moments converge to the corresponding moments of $F(y)$, then $F_n(y) \to F(y)$ at every point of continuity of $F(y)$. An extension of this theorem to the case where the $F_n(y)$ are random distribution functions and ordinary convergence is replaced by convergence in probability was given by M. N. Ghosh [6] under certain additional assumptions concerning $F(y)$ and its moments. The following theorem shows that the extension holds with no restrictions.

THEOREM 10.1. *Let $F(y)$ be a distribution function on the real line which is uniquely determined by its moments*

$$\mu_k = \int_{-\infty}^{\infty} y^k \, dF(y), \qquad k = 1, 2, \cdots$$

Let $\{F_n(y)\}$, $n = 1, 2, \cdots$, be a sequence of random distribution functions with moments μ_{nk}, and suppose that

$$\mu_{nk} \to \mu_k \text{ in probability as } n \to \infty, \qquad k = 1, 2, \cdots .$$

Then

$$F_n(y) \to F(y) \text{ in probability}$$

at every point of continuity of $F(y)$.

The proof is based on the following lemma.

LEMMA 10.1. *Let $F(y)$ be a distribution function which is uniquely determined by its moments μ_k, $k = 1, 2, \cdots$. Then for every y' at which $F(y)$ is continuous and for every $\epsilon > 0$ there exist a positive integer $m = m(y', \epsilon)$ and a positive number $\delta = \delta(y', \epsilon)$ such that for every distribution function $G(y)$ whose moments ν_k satisfy the inequalities*

$$|\nu_k - \mu_k| < \delta, \qquad k = 1, \cdots, m,$$

we have

$$|G(y') - F(y')| < \epsilon.$$

PROOF.[2] Assume the lemma to be false. Then for some y' at which $F(y)$ is continuous and for some $\epsilon > 0$ there do not exist positive numbers m, δ for

[2] The author is indebted to H. Robbins for the proof of Lemma 10.1.

which the conclusion of the lemma holds. Hence for every positive integer m there exists a distribution function $G_m(y)$ with moments ν_{mk} such that

$$| \nu_{mk} - \mu_k | < m^{-1}, \qquad k = 1, \cdots, m,$$

and

$$| G_m(y') - F(y') | \geqq \epsilon.$$

But $\{G_m(y)\}$, $m = 1, 2, \cdots$, is a sequence of distribution functions whose moments, ν_{mk}, converge to μ_k for all $k = 1, 2, \cdots$. By the aforementioned theorem of Fréchet and Shohat, $G_m(y') \to F(y')$, which leads to a contradiction.

Proof of Theorem 10.1. Let y' be a point of continuity of $F(y)$. Given $\epsilon > 0$, let $m = m(y', \epsilon)$, $\delta = \delta(y', \epsilon)$ be defined as in Lemma 10.1. Given $\eta > 0$, choose N so that

$$\Pr \{| \mu_{nk} - \mu_k | < \delta, k = 1, \cdots, m\} > 1 - \eta \qquad \text{for } n > N.$$

It follows from Lemma 10.1 that $| F_n(y') - F(y') | < \epsilon$ with probability $> 1 - \eta$ for $n > N$. The proof is complete.

It will now be shown that if the relations (6.13) are satisfied as limits in probability, (6.12) holds in probability. It can be seen from the proof of Theorem 6A in [7] that if (6.13) holds for $p = 3, 4, \cdots k$, then the moments up to order k of the distribution $F_n(y, a, b)$ converge to the corresponding moments of $\Phi(y)$. By Theorem 9.1 this implies that if (6.13) holds in probability for every $p = 3, 4, \cdots$, then every moment of $F_n(y, a, b)$ converges in probability to the corresponding moment of $\Phi(y)$. By Theorem 10.1, $F_n(y, a, b) \to \Phi(y)$ in probability.

Relations (6.13) can be shown to hold in probability under conditions which are slightly weaker than those indicated at the end of Section 9, though the gain does not seem to be considerable.

REFERENCES

[1] H. Cramér, *Random Variables and Probability Distributions, Cambridge Tracts in Mathematics, No. 36*, Cambridge University Press, 1937.
[2] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946.
[3] R. A. Fisher, *The Design of Experiments*, 5th ed., Hafner Publishing Co., New York, 1949.
[4] M. Fréchet, *Généralités sur les probabilités. Variables aléatoires.* Gauthier-Villars, Paris, 1937.
[5] M. Fréchet and J. Shohat, "A proof of the generalized second limit theorem in the theory of probability," *Trans. Am. Math. Soc.*, Vol. 33 (1931), p. 533–543.
[6] M. N. Ghosh, "Convergence of random distribution functions," *Bull. Calcutta Math. Soc.*, Vol. 42 (1950), p. 217–226.
[7] W. Hoeffding, "A combinatorial central limit theorem," *Annals of Math. Stat.*, Vol. 22 (1951), pp. 558–566.
[8] E. L. Lehmann and C. Stein, "On the theory of some non-parametric hypotheses," *Annals of Math. Stat.*, Vol. 20 (1949), pp. 28–45.
[9] G. E. Noether, "On a theorem by Wald and Wolfowitz," *Annals of Math. Stat.*, Vol. 20 (1949), pp. 455–458.

[10] G. E. NOETHER, "Asymptotic properties of the Wald-Wolfowitz test of randomness," *Annals of Math. Stat.*, Vol. 21 (1950), pp. 231–246.

[11] E. J. G. PITMAN, "Significance tests which may be applied to samples from any population," *Jour. Roy. Stat. Soc. Suppl.*, Vol. 4 (1937), pp. 119–130.
"Significance tests which may be applied to samples from any population. II. The correlation coefficient test," *Jour. Roy. Stat. Soc. Suppl.*, Vol. 4 (1937), pp. 225–232.
"Significance tests which may be applied to samples from any population. III. The analysis of variance test," *Biometrika*, Vol. 29 (1938), pp. 322–335.

[12] J. V. USPENSKY, *Introduction to Mathematical Probability*, McGraw-Hill Book Co., 1937.

[13] A. WALD AND J. WOLFOWITZ, "Statistical tests based on permutations of the observations," *Annals of Math. Stat.*, Vol. 15 (1944), pp. 358–372.

[14] B. L. WELCH, "On the z-test in randomized blocks and Latin squares," *Biometrika*, Vol. 29 (1937), pp. 21–52.