A CLASS OF STATISTICS WITH ASYMPTOTICALLY NORMAL DISTRIBUTION[1]

By Wassily Hoeffding

Institute of Statistics, University of North Carolina

1. Summary. Let $X_1, \cdots, X_n$ be n independent random vectors, $X_\nu = (X_\nu^{(1)}, \cdots, X_\nu^{(r)})$, and $\Phi(x_1, \cdots, x_m)$ a function of $m(\leq n)$ vectors $x_\nu = (x_\nu^{(1)}, \cdots, x_\nu^{(r)})$. A statistic of the form $U = \Sigma''\Phi(X_{\alpha_1}, \cdots, X_{\alpha_m})/n(n-1) \cdots (n-m+1)$, where the sum Σ'' is extended over all permutations $(\alpha_1, \cdots, \alpha_m)$ of m different integers, $1 \leq \alpha_i \leq n$, is called a U-statistic. If $X_1, \cdots, X_n$ have the same (cumulative) distribution function (d.f.) $F(x)$, U is an unbiased estimate of the population characteristic $\theta(F) = \int \cdots \int \Phi(x_1, \cdots, x_m) dF(x_1) \cdots dF(x_m)$. $\theta(F)$ is called a regular functional of the d.f. $F(x)$. Certain optimal properties of U-statistics as unbiased estimates of regular functionals have been established by Halmos [9] (cf. Section 4).

The variance of a U-statistic as a function of the sample size n and of certain population characteristics is studied in Section 5.

It is shown that if $X_1, \cdots, X_n$ have the same distribution and $\Phi(x_1, \cdots, x_m)$ is independent of n, the d.f. of $\sqrt{n}(U-\theta)$ tends to a normal d.f. as $n \to \infty$ under the sole condition of the existence of $E\Phi^2(X_1, \cdots, X_m)$. Similar results hold for the joint distribution of several U-statistics (Theorems 7.1 and 7.2), for statistics U' which, in a certain sense, are asymptotically equivalent to U (Theorems 7.3 and 7.4), for certain functions of statistics U or U' (Theorem 7.5) and, under certain additional assumptions, for the case of the X_ν's having different distributions (Theorems 8.1 and 8.2). Results of a similar character, though under different assumptions, are contained in a recent paper by von Mises [18] (cf. Section 7).

Examples of statistics of the form U or U' are the moments, Fisher's k-statistics, Gini's mean difference, and several rank correlation statistics such as Spearman's rank correlation and the difference sign correlation (cf. Section 9). Asymptotic power functions for the non-parametric tests of independence based on these rank statistics are obtained. They show that these tests are not unbiased in the limit (Section 9f). The asymptotic distribution of the coefficient of partial difference sign correlation which has been suggested by Kendall also is obtained (Section 9h).

2. Functionals of distribution functions. Let $F(x) = F(x^{(1)}, \cdots, x^{(r)})$ be an r-variate d.f. If to any F belonging to a subset $\mathfrak{D}$ of the set of all d.f.'s in the r-dimensional Euclidean space is assigned a quantity $\theta(F)$, then $\theta(F)$ is called a

[1] Research under a contract with the Office of Naval Research for development of multivariate statistical theory.

functional of F, defined on $\mathfrak{D}$. In this paper the word functional will always mean functional of a d.f.

An infinite population may be considered as completely determined by its d.f., and any numerical characteristic of an infinite population with d.f. F that is used in statistics is a functional of F. A finite population, or sample, of size n is determined by its d.f., $S(x)$ say, and its size n. n itself is not a functional of S since two samples of different size may have the same d.f.

If $S(x^{(1)}, \cdots, x^{(r)})$ is the d.f. of a finite population, or a sample, consisting of n elements

$$x_\alpha = (x_\alpha^{(1)}, \cdots, x_\alpha^{(r)}), \qquad (\alpha = 1, \cdots, n), \tag{2.1}$$

then $nS(x^{(1)}, \cdots, x^{(r)})$ is the number of elements x_α such that

$$x_\alpha^{(1)} \leq x^{(1)}, \cdots, x_\alpha^{(r)} \leq x^{(r)}.$$

Since $S(x^{(1)}, \cdots, x^{(r)})$ is symmetric in $x_1, \cdots, x_n$, and retains its value for a sample formed from the sample (2.1) by adding one or more identical samples, the same two properties hold true for a sample functional $\theta(S)$. Most statistics in current use are functions of n and of functionals of the sample d.f.

A random sample $\{X_1, \cdots, X_n\}$ is a set of n independent random vectors

$$X_\alpha = (X_\alpha^{(1)}, \cdots, X_\alpha^{(r)}), \qquad (\alpha = 1, \cdots, n). \tag{2.2}$$

For any fixed values $x^{(1)}, \cdots, x^{(r)}$, the d.f. $S(x^{(1)}, \cdots, x^{(r)})$ of a random sample is a random variable. The functional $\theta(S)$, where S is the d.f. of the random sample, is itself a random variable, and may be called a random functional.

A remarkable application of the theory of functionals to functionals of d.f.'s has been made by von Mises [18] who considers the asymptotic distributions of certain functionals of sample d.f.'s. (Cf. also Section 7.)

3. Unbiased estimation and regular functionals. Consider a functional $\theta = \theta(F)$ of the r-variate d.f. $F(x) = F(x^{(1)}, \cdots, x^{(r)})$, and suppose that for some sample size n, θ admits an unbiased estimate for any d.f. F in $\mathfrak{D}$. That is, if $X_1, \cdots, X_n$ are n independent random vectors with the same d.f. F, there exists a function $\varphi(x_1, \cdots, x_n)$ of n vector arguments (2.1) such that the expected value of $\varphi(X_1, \cdots, X_n)$ is equal to $\theta(F)$, or

$$\int \cdots \int \varphi(x_1, \cdots, x_n)\, dF(x_1) \cdots dF(x_n) = \theta(F) \tag{3.1}$$

for every F in $\mathfrak{D}$. Here and in the sequel, when no integration limits are indicated, the integral is extended over the entire space of $x_1, \cdots, x_n$. The integral is understood in the sense of Stieltjes-Lebesgue.

The estimate $\varphi(x_1, \cdots, x_n)$ of $\theta(F)$ is called unbiased over $\mathfrak{D}$.

A functional $\theta(F)$ of the form (3.1) will be referred to as *regular over* $\mathfrak{D}$.[2]

[1] This is an adaptation to functionals of d.f.'s of the term "regular functional" used by Volterra [21].

Thus, the functionals regular over $\mathfrak{D}$ are those admitting an unbiased estimate over $\mathfrak{D}$.

If $\theta(F)$ is regular over $\mathfrak{D}$, let $m(\leq n)$ be the smallest sample size for which there exists an unbiased estimate $\Phi(x_1, \cdots, x_m)$ of θ over $\mathfrak{D}$:

$$\theta(F) = \int \cdots \int \Phi(x_1, \cdots, x_m)\, dF(x_1) \cdots dF(x_m) \tag{3.2}$$

for any F in $\mathfrak{D}$. Then m will be called *the degree over* $\mathfrak{D}$ of the regular functional $\theta(F)$.

If the expected value of $\varphi(X_1, \cdots, X_n)$ is equal to $\theta(F)$ whenever it exists, $\varphi(x_1, \cdots, x_n)$ will be called a *distribution-free unbiased estimate* (d-f. u.e.) of $\theta(F)$. The degree of $\theta(F)$ over the set $\mathfrak{D}_0$ of d.f.'s F for which the right hand side of (3.1) exists will be simply termed the *degree* of $\theta(F)$.

A regular functional of degree 1 over $\mathfrak{D}$ is called a linear regular functional over $\mathfrak{D}$. If $\theta(F)$ has the same value for all F in $\mathfrak{D}$, $\theta(F)$ may be termed a regular functional of degree zero over $\mathfrak{D}$.

Any function $\Phi(x_1, \cdots, x_m)$ satisfying (3.2) will be referred to as a *kernel* of the regular functional $\theta(F)$.

For any regular functional $\theta(F)$ there exists a kernel $\Phi_0(x_1, \cdots, x_m)$ symmetric in $x_1, \cdots, x_m$. For if $\Phi(x_1, \cdots, x_m)$ is a kernel of $\theta(F)$,

$$\Phi_0(x_1, \cdots, x_m) = \frac{1}{m!} \Sigma \Phi(x_{\alpha_1}, \cdots, x_{\alpha_m}), \tag{3.3}$$

where the sum is taken over all permutations $(\alpha_1, \cdots, \alpha_m)$ of $(1, \cdots, m)$, is a symmetric kernel of $\theta(F)$.

If $\theta_1(F)$ and $\theta_2(F)$ are two regular functionals of degrees m_1 and m_2 over $\mathfrak{D}$, then the sum $\theta_1(F) + \theta_2(F)$ and the product $\theta_1(F)\theta_2(F)$ are regular functionals of degrees $\leq m = \text{Max}\ (m_1, m_2)$ and $\leq m_1 + m_2$, respectively, over $\mathfrak{D}$. For if $\Phi_i(x_1, \cdots, x_{m_i})$ is a kernel of $\theta_i(F)$, $(i = 1, 2)$, then

$$\theta_1(F) + \theta_2(F) = \int \cdots \int \{\Phi_1(x_1, \cdots, x_{m_1}) + \Phi_2(x_1, \cdots, x_{m_2})\} \, dF(x_1) \cdots dF(x_m)$$

and

$$\theta_1(F)\theta_2(F) = \int \cdots \int \Phi_1(x_1, \cdots, x_{m_1})\Phi_2(x_{m_1+1}, \cdots, x_{m_1+m_2}) \, dF(x_1) \cdots dF(x_{m_1+m_2}).$$

More generally, *a polynomial in regular functionals is itself a regular functional.*

Examples of linear regular functionals are the moments about the origin,

$$\mu'_{\nu_1, \cdots, \nu_r} = \int \cdots \int (x^{(1)})^{\nu_1} \cdots (x^{(r)})^{\nu_r} \, dF(x^{(1)}, \cdots, x^{(r)}).$$

A moment about the mean is a polynomial in moments μ' about 0, and hence a regular functional over the set $\mathcal{D}_0$ of d.f.'s for which it exists (cf. Halmos [9]). For instance, the variance of $X^{(1)}$,

$$\sigma^2 = \int\int ((x_1^{(1)})^2 - x_1^{(1)} x_2^{(1)})\, dF(x_1^{(1)})\, dF(x_2^{(1)})$$

is a regular functional of degree 2. A symmetrical kernel of σ^2 is $(x^{(1)} - x^{(2)})^2/2$. If $\mathcal{D}$ is the set of univariate d.f.'s with mean μ and existing second moment, σ^2 is a linear regular functional of F over $\mathcal{D}$, since then we have

$$\sigma^2 = \int (x_1^{(1)} - \mu)^2\, dF(x_1^{(1)}).$$

The function

$$v = \frac{1}{n(n-1)} \sum_{\alpha \neq \beta} \frac{1}{2} (x_\alpha^{(1)} - x_\beta^{(1)})^2 = \frac{1}{n-1} \sum_\alpha \left(x_\alpha^{(1)} - \frac{1}{n} \sum_\beta x_\beta^{(1)} \right)^2$$

is a distribution-free unbiased estimate of σ^2. The function

$$\Gamma\left(\frac{n-1}{2}\right) \sqrt{\frac{n-1}{2}}\, \sqrt{v} / \Gamma\left(\frac{n}{2}\right)$$

is known to be an unbiased estimate of σ over the set of univariate normal d.f.'s, but it is not a d.-f. u.e.

4. *U*-statistics. Let $x_1, \cdots, x_n$ be a sample of n vectors (2.1) and $\Phi(x_1, \cdots, x_m)$ a function of $m (\leq n)$ vector arguments. Consider the function of the sample,

$$(4.1) \quad U = U(x_1, \cdots, x_n) = \frac{1}{n(n-1) \cdots (n-m+1)} \Sigma'' \Phi(x_{\alpha_1}, \cdots, x_{\alpha_m}),$$

where Σ'' stands for summation over all permutations $(\alpha_1, \cdots, \alpha_m)$ of m integers such that

$$(4.2) \qquad 1 \leq \alpha_i \leq n, \qquad \alpha_i \neq \alpha_j \text{ if } i \neq j, \qquad (i, j = 1, \cdots, m).$$

U is the average of the values of Φ in the set of ordered subsets of m members of the sample (2.1). U is symmetric in $x_1, \cdots, x_n$.

Any statistic of the form (4.1) will be called a *U-statistic*. Any function $\Phi(x_1, \cdots, x_m)$ satisfying (4.1) will be referred to as a *kernel* of the statistic U.

If $\Phi(x_1, \cdots, x_m)$ is a kernel of a regular functional $\theta(F)$ defined on a set $\mathcal{D}$, then U is an unbiased estimate of $\theta(F)$ over $\mathcal{D}$:

$$(4.3) \qquad \theta(F) = \int \cdots \int U(x_1, \cdots, x_n)\, dF(x_1) \cdots dF(x_n)$$

for every F in $\mathcal{D}$.

For $n = m$, U reduces to the symmetric kernel (3.3) of $\theta(F)$.

From a recent paper by Halmos [9] it follows for the case of univariate d.f.'s ($r = 1$):

If $\theta(F)$ is a regular functional of degree m over a set $\mathfrak{D}$ containing all purely discontinuous d.f.'s, U is the only unbiased estimate over $\mathfrak{D}$ which is symmetric in $x_1, \cdots, x_n$, and U has the least variance among all unbiased estimates over $\mathfrak{D}$.

These results and the proofs given by Halmos can easily be extended to the multivariate case ($r > 1$).

Combining (3.3) and (4.1) we may write a U-statistic in the form

$$U(x_1, \cdots, x_n) = \binom{n}{m}^{-1} \Sigma' \Phi_0(x_{\alpha_1}, \cdots, x_{\alpha_m}), \tag{4.4}$$

where the kernel Φ_0 is symmetric in its m vector arguments and the sum Σ' is extended over all subscripts α such that

$$1 \leq \alpha_1 < \alpha_2 < \cdots < \alpha_m \leq n.$$

Another statistic frequently used for estimating $\theta(F)$ is $\theta(S)$, where $S = S(x)$ is the d.f. of the sample (2.1). If S is substituted for F in (3.2), we have

$$\theta(S) = \frac{1}{n^m} \sum_{\alpha_1=1}^{n} \cdots \sum_{\alpha_m=1}^{n} \Phi(x_{\alpha_1}, \cdots, x_{\alpha_m}). \tag{4.5}$$

In particular, the sample moments have this form; their kernel Φ is obtained by the method described in section 3.

If $m = 1$, $\theta(S) = U$. If $m = 2$,

$$\theta(S) = \frac{n-1}{n} U + \frac{1}{n} \left\{ \frac{1}{n} \sum_{\alpha=1}^{n} \Phi(x_\alpha, x_\alpha) \right\},$$

and $\theta(S)$ is a linear function of U-statistics with coefficients depending on n. This is easily seen to be true for any m. In general $\theta(S)$ is not an unbiased estimate of $\theta(F)$. If, however, the expected value of $\theta(S)$ exists for every F in $\mathfrak{D}$, we have

$$E\{\theta(S)\} = \theta(F) + O(n^{-1}),$$

and the estimate $\theta(S)$ of $\theta(F)$ may be termed unbiased in the limit over $\mathfrak{D}$.

Numerous statistics in current use have the form of, or can be expressed in terms of U-statistics. From what was said above about moments as regular functionals, it is easy to obtain U-statistics which are d.-f. u.e.'s of the moments about the mean of any order (cf. Halmos [9]). Fisher's k-statistics are U-statistics, as follows from their definition as unbiased estimates of the cumulants, symmetric in the sample values. Another example is Gini's mean difference

$$\frac{1}{n(n-1)} \sum_{\alpha \neq \beta} | x_\alpha^{(1)} - x_\beta^{(1)} |.$$

More examples, in particular of rank correlation statistics, will be given in section 9.

5. The variance of a U-statistic. Let $X_1, \cdots, X_n$ be n independent random vectors with the same d.f. $F(x) = F(x^{(1)}, \cdots, x^{(r)})$, and let

$$U = U(X_1, \cdots, X_n) = \binom{n}{m}^{-1} \Sigma' \Phi(X_{\alpha_1}, \cdots, X_{\alpha_m}), \tag{5.1}$$

where $\Phi(x_1, \cdots, x_m)$ is symmetric in $x_1, \cdots, x_m$ and Σ' has the same meaning as in (4.4). Suppose that the function Φ does not involve n.

If $\theta = \theta(F)$ is defined by (3.2), we have

$$E\{U\} = E\{\Phi(X_1, \cdots, X_m)\} = \theta.$$

Let

$$\Phi_c(x_1, \cdots, x_c) = E\{\Phi(x_1, \cdots, x_c, X_{c+1}, \cdots, X_m)\}, \quad (c = 1, \cdots, m), \tag{5.2}$$

where $x_1, \cdots, x_c$ are arbitrary fixed vectors and the expected value is taken with respect to the random vectors $X_{c+1}, \cdots, X_m$. Then

$$\Phi_{c-1}(x_1, \cdots, x_{c-1}) = E\{\Phi_c(x_1, \cdots, x_{c-1}, X_c)\}, \tag{5.3}$$

and

$$E\{\Phi_c(X_1, \cdots, X_c)\} = \theta, \qquad (c = 1, \cdots, m). \tag{5.4}$$

Define

$$\Psi(x_1, \cdots, x_m) = \Phi(x_1, \cdots, x_m) - \theta, \tag{5.5}$$

$$\Psi_c(x_1, \cdots, x_c) = \Phi_c(x_1, \cdots, x_c) - \theta, \quad (c = 1, \cdots, m). \tag{5.6}$$

We have

$$\Psi_{c-1}(x_1, \cdots, x_{c-1}) = E\{\Psi_c(x_1, \cdots, x_{c-1}, X_c)\}, \tag{5.7}$$

$$E\{\Psi_c(X_1, \cdots, X_c)\} = E\{\Psi(X_1, \cdots, X_m)\} = 0, \quad (c = 1, \cdots, m). \tag{5.8}$$

Suppose that the variance of $\Psi_c(X_1, \cdots, X_c)$ exists, and let

$$\zeta_0 = 0, \qquad \zeta_c = E\{\Psi_c^2(X_1, \cdots, X_c)\}, \qquad (c = 1, \cdots, m). \tag{5.9}$$

We have

$$\zeta_c = E\{\Phi_c^2(X_1, \cdots, X_c)\} - \theta^2. \tag{5.10}$$

$\zeta_c = \zeta_c(F)$ is a polynomial in regular functionals of F, and hence itself a regular functional of F (of degree $\leq 2m$).

If, for some parent distribution $F = F_0$ and some integer d, we have $\zeta_d(F_0) = 0$, this means that $\Psi_d(X_1, \cdots, X_d) = 0$ with probability 1. By (5.7) and (5.9), $\zeta_d = 0$ implies $\zeta_1 = \cdots = \zeta_{d-1} = 0$.

If $\zeta_1(F_0) = 0$, we shall say that the regular functional $\theta(F)$ is *stationary*[3] for $F = F_0$. If

$$\zeta_1(F_0) = \cdots = \zeta_d(F_0) = 0, \qquad \zeta_{d+1}(F_0) > 0, \qquad (1 \leq d \leq m), \tag{5.11}$$

$\theta(F)$ will be called *stationary of order* d for $F = F_0$.

If $(\alpha_1, \cdots, \alpha_m)$ and $(\beta_1, \cdots, \beta_m)$ are two sets of m different integers, $1 \leq \alpha_i, \beta_i \leq n$, and c is the number of integers common to the two sets, we have, by the symmetry of Ψ,

$$E\{\Psi(X_{\alpha_1}, \cdots, X_{\alpha_m})\Psi(X_{\beta_1}, \cdots, X_{\beta_m})\} = \zeta_c. \tag{5.12}$$

If the variance of U exists, it is equal to

$$\begin{aligned}\sigma^2(U) &= \binom{n}{m}^{-2} E\{\Sigma'\Psi(X_{\alpha_1}, \cdots, X_{\alpha_m})\}^2 \\ &= \binom{n}{m}^{-2} \sum_{c=0}^{m} \Sigma^{(c)} E\{\Psi(X_{\alpha_1}, \cdots, X_{\alpha_m})\Psi(X_{\beta_1}, \cdots, X_{\beta_m})\},\end{aligned}$$

where $\Sigma^{(c)}$ stands for summation over all subscripts such that

$$1 \leq \alpha_1 < \alpha_2 < \cdots < \alpha_m \leq n, \qquad 1 \leq \beta_1 < \beta_2 < \cdots < \beta_m \leq n,$$

and exactly c equations

$$\alpha_i = \beta_j$$

are satisfied. By (5.12), each term in $\Sigma^{(c)}$ is equal to ζ_c. The number of terms in $\Sigma^{(c)}$ is easily seen to be

$$\frac{n(n-1)\cdots(n-2m+c+1)}{c!(m-c)!(m-c)!} = \binom{m}{c}\binom{n-m}{m-c}\binom{n}{m},$$

and hence, since $\zeta_0 = 0$,

$$\sigma^2(U) = \binom{n}{m}^{-1} \sum_{c=1}^{m} \binom{m}{c}\binom{n-m}{m-c}\zeta_c. \tag{5.13}$$

When the distributions of $X_1, \cdots, X_n$ are different, $F_\nu(x)$ being the d.f. of X_ν, let

$$\theta_{\alpha_1,\cdots,\alpha_m} = E\{\Phi(X_{\alpha_1}, \cdots, X_{\alpha_m})\}, \tag{5.14}$$

$$\begin{aligned}&\Psi_{c(\alpha_1,\cdots,\alpha_c)\beta_1,\cdots,\beta_{m-c}}(x_1, \cdots, x_c) \\ &= E\{\Phi(x_1, \cdots, x_c, X_{\beta_1} \ldots, X_{\beta_{m-c}})\} - \theta_{\alpha_1,\cdots,\alpha_c,\beta_1,\cdots,\beta_{m-c}},\end{aligned} \tag{5.15}$$

$$(c = 1, \cdots, m),$$

[3] According to the definition of the derivative of a functional (cf. Volterra [21]; for functionals of d.f.'s cf. von Mises [18]), the function $m(m-1)\ldots(m-d+1)\,\Psi_d(x_1 \ldots x_d)$, which is a functional of F, is a d-th derivative of $\theta(F)$ with respect to F at the "point" F of the space of d.f.'s.

$$\zeta_{c(\alpha_1,\cdots,\alpha_c)\beta_1,\cdots,\beta_{m-c};\gamma_1,\cdots,\gamma_{m-c}}$$

$$= E\{\Psi_{c(\alpha_1,\cdots,\alpha_c)\beta_1,\cdots,\beta_{m-c}}(X_{\alpha_1},\cdots,X_{\alpha_c})\Psi_{c(\alpha_1,\cdots,\alpha_c)\gamma_1,\cdots,\gamma_{m-c}}(X_{\alpha_1},\cdots,X_{\alpha_c})\} \tag{5.16}$$

$$\zeta_{c,n} = \frac{c!(m-c)!(m-c)!}{n(n-1)\cdots(n-2m+c+1)} \Sigma\zeta_{c(\alpha_1,\cdots,\alpha_c)\beta_1,\cdots,\beta_{m-c};\gamma_1,\cdots,\gamma_{m-c}} \tag{5.17}$$

where the sum is extended over all subscripts α, β, γ such that

$$1 \le \alpha_1 < \cdots < \alpha_c \le n, \quad 1 \le \beta_1 < \cdots < \beta_{m-c} \le n, \quad 1 \le \gamma_1 < \cdots \gamma_{m-c} \le n,$$

$$\alpha_i \ne \beta_j, \qquad \alpha_i \ne \gamma_j, \qquad \beta_i \ne \gamma_j.$$

Then the variance of U is equal to

$$\sigma^2(U) = \binom{n}{m}^{-1} \sum_{c=1}^{m} \binom{m}{c}\binom{n-m}{m-c} \zeta_{c,n}. \tag{5.18}$$

Returning to the case of identically distributed X's, we shall now prove some inequalities satisfied by $\zeta_1, \cdots, \zeta_m$ and $\sigma^2(U)$ which are contained in the following theorems:

THEOREM 5.1 *The quantities* $\zeta_1, \cdots, \zeta_m$ *as defined by* (5.9) *satisfy the inequalities*

$$0 \le \frac{\zeta_c}{c} \le \frac{\zeta_d}{d} \qquad \text{if } 1 \le c < d \le m. \tag{5.19}$$

THEOREM 5.2 *The variance* $\sigma^2(U_n)$ *of a U-statistic* $U_n = U(X_1, \cdots, X_n)$, *where* $X_1, \cdots, X_n$ *are independent and identically distributed, satisfies the inequalities*

$$\frac{m^2}{n}\zeta_1 \le \sigma^2(U_n) \le \frac{m}{n}\zeta_m. \tag{5.20}$$

$n\sigma^2(U_n)$ *is a decreasing function of* n,

$$(n+1)\sigma^2(U_{n+1}) \le n\sigma^2(U_n), \tag{5.21}$$

which takes on its upper bound $m\zeta_m$ *for* $n = m$ *and tends to its lower bound* $m^2\zeta_1$ *as* n *increases:*

$$\sigma^2(U_m) = \zeta_m, \tag{5.22}$$

$$\lim_{n\to\infty} n\sigma^2(U_n) = m^2\zeta_1. \tag{5.23}$$

If $E\{U_n\} = \theta(F)$ *is stationary of order* $\ge d-1$ *for the d.f. of* X_α, (5.20) *may be replaced by*

$$\frac{m}{d} K_n(m,d)\zeta_d \le \sigma^2(U_n) \le K_n(m,d)\zeta_m, \tag{5.24}$$

where

$$K_n(m, d) = \binom{n}{m}^{-1} \sum_{c=d}^{m} \binom{m-1}{c-1}\binom{n-m}{m-c}. \tag{5.25}$$

We postpone the proofs of Theorems 5.1 and 5.2.

(5.13) and (5.19) imply that a necessary and sufficient condition for the existence of $\sigma^2(U)$ is the existence of

$$\zeta_m = E\{\Phi^2(X_1, \cdots, X_m)\} - \theta^2 \tag{5.26}$$

or that of $E\{\Phi^2(X_1, \cdots, X_m)\}$.

If $\zeta_1 > 0$, $\sigma^2(U)$ is of order n^{-1}.

If $\theta(F)$ is stationary of order d for $F = F_0$, that is, if (5.11) is satisfied, $\sigma^2(U)$ is of order n^{-d-1}. Only if, for some $F = F_0$, $\theta(F)$ is stationary of order m, where m is the degree of $\theta(F)$, we have $\sigma^2(U) = 0$, and U is equal to a constant with probability 1.

For instance, if $\theta(F_0) = 0$, the functional $\theta^2(F)$ is stationary for $F = F_0$. Other examples of stationary "points" of a functional will be found in section 9d.

For proving Theorem 5.1 we shall require the following:

LEMMA 5.1. *If*

$$\delta_d = \zeta_d - \binom{d}{1}\zeta_{d-1} + \binom{d}{2}\zeta_{d-2} \cdots + (-1)^{d-1}\binom{d}{d-1}\zeta_1, \tag{5.27}$$

we have

$$\delta_d \geq 0, \qquad (d = 1, \cdots, m) \tag{5.28}$$

and

$$\zeta_d = \delta_d + \binom{d}{1}\delta_{d-1} + \cdots + \binom{d}{d-1}\delta_1. \tag{5.29}$$

PROOF. (5.29) follows from (5.27) by induction.

For proving (5.28) let

$$\eta_0 = \theta^2, \qquad \eta_c = E\{\Phi_c^2(X_1, \cdots, X_c)\}, \qquad (c = 1, \cdots, m).$$

Then, by (5.10),

$$\zeta_c = \eta_c - \eta_0,$$

and on substituting this in (5.27) we have

$$\delta_d = \sum_{c=0}^{d} (-1)^{d-c}\binom{d}{c}\eta_c.$$

From (5.9) it is seen that (5.28) is true for $d = 1$. Suppose that (5.28) holds for $1, \cdots, d-1$. Then (5.28) will be shown to hold for d.

Let

$$\bar{\Phi}_0(x_1) = \Phi_1(x_1) - \theta, \qquad \bar{\Phi}_c(x_1, x_2, \cdots, x_{c+1})$$
$$= \Phi_{c+1}(x_1, \cdots, x_{c+1}) - \Phi_c(x_2, \cdots, x_{c+1}), \qquad (c = 1, \cdots, d-1).$$

For an arbitrary fixed x_1, let

$$\bar{\eta}_c(x_1) = E\{\bar{\Phi}_c^2(x_1, X_2, \cdots, X_{c+1})\}, \qquad (c = 0, \cdots, d-1).$$

Then, by induction hypothesis,

$$\bar{\delta}_{d-1}(x_1) = \sum_{c=0}^{d-1} (-1)^{d-1-c} \binom{d-1}{c} \bar{\eta}_c(x_1) \geq 0$$

for any fixed x_1.

Now,

$$E\{\bar{\eta}_c(X_1)\} = \eta_{c+1} - \eta_c,$$

and hence

$$E\{\bar{\delta}_{d-1}(X_1)\} = \sum_{c=0}^{d-1} (-1)^{d-1-c} \binom{d-1}{c} (\eta_{c+1} - \eta_c) = \sum_{c=0}^{d} (-1)^{d-c} \binom{d}{c} \eta_c = \delta_d.$$

The proof of Lemma 5.1 is complete.

PROOF OF THEOREM 5.1. By (5.29) we have for $c < d$

$$c\zeta_d - d\zeta_c = c \sum_{a=1}^{d} \binom{d}{a} \delta_a - d \sum_{a=1}^{c} \binom{c}{a} \delta_a$$

$$= \sum_{a=1}^{c} \left[c\binom{d}{a} - d\binom{c}{a} \right] \delta_a + c \sum_{a=c+1}^{d} \binom{d}{a} \delta_a. \tag{5.30}$$

From (5.28), and since $c\binom{d}{a} - d\binom{c}{a} \geq 0$ if $1 \leq a \leq c \leq d$, it follows that each term in the two sums of (5.30) is not negative. This, in connection with (5.9) proves Theorem 5.1.

PROOF OF THEOREM 5.2. From (5.19) we have

$$c\zeta_1 \leq \zeta_c \leq \frac{c}{m} \zeta_m, \qquad (c = 1, \cdots, m).$$

Applying these inequalities to each term in (5.13) and using the identity

$$\binom{n}{m}^{-1} \sum_{c=1}^{m} c \binom{m}{c} \binom{n-m}{m-c} = \frac{m^2}{n}, \tag{5.31}$$

we obtain (5.20).

(5.22) and (5.23) follow immediately from (5.13).

For (5.21) we may write

$$D_n \geq 0, \tag{5.32}$$

where

$$D_n = n\sigma^2(U_n) - (n+1)\sigma^2(U_{n+1}).$$

Let

$$D_n = \sum_{c=1}^{m} d_{n,c}\zeta_c.$$

Then we have from (5.13)

$$d_{n,c} = n\binom{m}{c}\binom{n-m}{m-c}\binom{n}{m}^{-1} - (n+1)\binom{m}{c}\binom{n+1-m}{m-c}\binom{n+1}{m}^{-1}, \tag{5.33}$$

or

$$d_{n,c} = \binom{m}{c}\binom{n-m+1}{m-c}(n-m+1)^{-1}\binom{n}{m}^{-1}\{(c-1)n - (m-1)^2\}, \qquad (1 \le c \le m \le n).$$

Putting

$$c_0 = 1 + \left[\frac{(m-1)^2}{n}\right],$$

where $[u]$ denotes the largest integer $\le u$, we have

$$d_{n,c} \le 0 \qquad \text{if } c \le c_0,$$
$$d_{n,c} > 0 \qquad \text{if } c > c_0.$$

Hence, by (5.19),

$$d_{n,c}\zeta_c \ge \frac{1}{c_0}\zeta_{c_0} c d_{n,c}, \qquad (c = 1, \cdots, m),$$

and

$$D_n \ge \frac{1}{c_0}\zeta_{c_0}\sum_{c=1}^{m} c d_{n,c}.$$

By (5.33) and (5.31), the latter sum vanishes. This proves (5.32).

For the stationary case $\zeta_1 = \cdots = \zeta_{d-1} = 0$, (5.24) is a direct consequence of (5.13) and (5.19). The proof of Theorem 5.2 is complete.

6. The covariance of two U-statistics. Consider a set of g U-statistics,

$$U^{(\gamma)} = \binom{n}{m(\gamma)}^{-1}\Sigma' \Phi^{(\gamma)}(X_{\alpha_1}, \cdots, X_{\alpha_{m(\gamma)}}), \qquad (\gamma = 1, \cdots, g),$$

each $U^{(\gamma)}$ being a function of the same n independent, identically distributed random vectors $X_1, \cdots, X_n$. The function $\Phi^{(\gamma)}$ is assumed to be symmetric in its $m(\gamma)$ arguments ($\gamma = 1, \cdots, g$).

Let

$$E\{U^{(\gamma)}\} = E\{\Phi^{(\gamma)}(X_1, \cdots, X_{m(\gamma)})\} = \theta^{(\gamma)}, \qquad (\gamma = 1, \cdots, g);$$

$$\Psi^{(\gamma)}(x_1, \cdots, x_{m(\gamma)}) = \Phi^{(\gamma)}(x_1, \cdots, x_{m(\gamma)}) - \theta^{(\gamma)}, \qquad (\gamma = 1, \cdots, g); \tag{6.1}$$

$$\Psi_c^{(\gamma)}(x_1, \cdots, x_c) = E\{\Psi^{(\gamma)}(x_1, \cdots, x_c, X_{c+1}, \cdots, X_{m(\gamma)})\}, \tag{6.2}$$

$$(c = 1, \cdots, m(\gamma); \gamma = 1, \cdots, g);$$

$$\zeta_c^{(\gamma,\delta)} = E\{\Psi_c^{(\gamma)}(X_1, \cdots, X_c)\Psi_c^{(\delta)}(X_1, \cdots, X_c)\}, \tag{6.3}$$

$$(\gamma, \delta = 1, \cdots, g).$$

If, in particular, $\gamma = \delta$, we shall write

$$\zeta_c^{(\gamma)} = \zeta_c^{(\gamma,\gamma)} = E\{\Psi_c^{(\gamma)}(X_1, \cdots, X_c)\}^2. \tag{6.4}$$

Let

$$\sigma(U^{(\gamma)}, U^{(\delta)}) = E\{(U^{(\gamma)} - \theta^{(\gamma)})(U^{(\delta)} - \theta^{(\delta)})\}$$

be the covariance of $U^{(\gamma)}$ and $U^{(\delta)}$.

In a similar way as for the variance, we find, if $m\ (\gamma) \leq m\ (\delta)$,

$$\sigma(U^{(\gamma)}, U^{(\delta)}) = \binom{n}{m(\gamma)}^{-1} \sum_{c=1}^{m(\gamma)} \binom{m(\delta)}{c} \binom{n - m(\delta)}{m(\gamma) - c} \zeta_c^{(\gamma,\delta)}. \tag{6.5}$$

The right hand side is easily seen to be symmetric in γ, δ.

For $\gamma = \delta$, (6.5) is the variance of $U^{(\gamma)}$ (cf. (5.13)).

We have from (5.23) and (6.5)

$$\lim_{n\to\infty} n\sigma^2(U^{(\gamma)}) = m^2(\gamma)\zeta_1^{(\gamma)},$$

$$\lim_{n\to\infty} n\sigma(U^{(\gamma)}, U^{(\delta)}) = m(\gamma)m(\delta)\zeta_1^{(\gamma,\delta)}.$$

Hence, if $\zeta_1^{(\gamma)} \neq 0$ and $\zeta_1^{(\delta)} \neq 0$, the product moment correlation $\rho(U^{(\gamma)}, U^{(\delta)})$ between $U^{(\gamma)}$ and $U^{(\delta)}$ tends to the limit

$$\lim_{n\to\infty} \rho(U^{(\gamma)}, U^{(\delta)}) = \frac{\zeta_1^{(\gamma,\delta)}}{\sqrt{\zeta_1^{(\gamma)}\zeta_1^{(\delta)}}}. \tag{6.6}$$

7. Limit theorems for the case of identically distributed X_α's. We shall now study the asymptotic distribution of U-statistics and certain related functions. In this section the vectors X_α will be assumed to be identically distributed. An extension to the case of different parent distributions will be given in section 8.

Following Cramér [2, p. 83] we shall say that a sequence of d.f.'s $F_1(x)$, $F_2(x)$, $\cdots$ converges to a d.f. $F(x)$ if $\lim F_n(x) = F(x)$ in every point at which the one-dimensional marginal limiting d.f.'s are continuous.

Let us recall (cf. Cramér [2, p. 312]) that a g-variate normal distribution is called non-singular if the rank r of its covariance matrix is equal to g, and singular if $r < g$.

The following lemma will be used in the proofs.

LEMMA 7.1. *Let $V_1, V_2, \cdots$ be an infinite sequence of random vectors $V_n = (V_n^{(1)}, \cdots, V_n^{(g)})$, and suppose that the d.f. $F_n(v)$ of V_n tends to a d.f. $F(v)$ as $n \to \infty$. Let $V_n^{(\gamma)\prime} = V_n^{(\gamma)} + d_n^{(\gamma)}$, where*

$$\lim_{n\to\infty} E\{d_n^{(\gamma)}\}^2 = 0, \qquad (\gamma = 1, \cdots, g). \tag{7.1}$$

Then the d.f. of $V_n' = (V_n^{(1)\prime}, \cdots, V_n^{(g)\prime})$ tends to $F(v)$.

This is an immediate consequence of the well-known fact that the d.f. of V_n' tends to $F(v)$ if $d_n^{(\gamma)}$ converges in probability to 0 (cf. Cramér [2, p. 299]), since the fulfillment of (7.1) is sufficient for the latter condition.

THEOREM 7.1. *Let $X_1, \cdots, X_n$ be n independent, identically distributed random vectors,*

$$X_\alpha = (X_\alpha^{(1)}, \cdots, X_\alpha^{(r)}), \qquad (\alpha = 1, \cdots, n).$$

Let

$$\Phi^{(\gamma)}(x_1, \cdots, x_{m(\gamma)}), \qquad (\gamma = 1, \cdots, g),$$

be g real-valued functions not involving n, $\Phi^{(\gamma)}$ being symmetric in its $m(\gamma)$ $(\leq n)$ vector arguments $x_\alpha = (x_\alpha^{(1)}, \cdots, x_\alpha^{(r)})$, $(\alpha = 1, \cdots, m(\gamma);\ \gamma = 1, \cdots, g)$. Define

$$U^{(\gamma)} = \binom{n}{m(\gamma)}^{-1} \sum{}' \Phi^{(\gamma)}(X_{\alpha_1}, \cdots, X_{\alpha_{m(\gamma)}}), \qquad (\gamma = 1, \cdots, g), \tag{7.2}$$

where the summation is over all subscripts such that $1 \leq \alpha_1 < \cdots < \alpha_{m(\gamma)} \leq n$. Then, if the expected values

$$\theta^{(\gamma)} = E\{\Phi^{(\gamma)}(X_1, \cdots, X_{m(\gamma)})\}, \qquad (\gamma = 1, \cdots, g), \tag{7.3}$$

and

$$E\{\Phi^{(\gamma)}(X_1, \cdots, X_{m(\gamma)})\}^2, \qquad (\gamma = 1, \cdots, g), \tag{7.4}$$

exist, the joint d.f. of

$$\sqrt{n}(U^{(1)} - \theta^{(1)}), \cdots, \sqrt{n}(U^{(g)} - \theta^{(g)})$$

tends, as $n \to \infty$, to the g-variate normal d.f. with zero means and covariance matrix $(m(\gamma)m(\delta)\zeta_1^{(\gamma,\delta)})$, where $\zeta_1^{(\gamma,\delta)}$ is defined by (6.3). The limiting distribution is non-singular if the determinant $|\zeta_1^{(\gamma,\delta)}|$ is positive.

Before proving Theorem 7.1, a few words may be said about its meaning and its relation to well-known results.

For $g = 1$, Theorem 7.1 states that the distribution of a U-statistic tends, under certain conditions, to the normal form. For $m = 1$, U is the sum of n inde-

pendent random variables, and in this case Theorem 7.1 reduces to the Central Limit Theorem for such sums. For $m > 1$, U is a sum of random variables which, in general, are not independent. Under certain assumptions about the function $\Phi(x_1, \cdots, x_m)$ the asymptotic normality of U can be inferred from the Central Limit Theorem by well-known methods. If, for instance, Φ is a polynomial (as in the case of the k-statistics or the unbiased estimates of moments), U can be expressed as a polynomial in moments about the origin which are sums of independent random variables, and for this case the tendency to normality of U can easily be shown (cf. Cramér [2, p. 365]).

Theorem 7.1 generalizes these results, stating that in the case of independent and identically distributed X_α's the existence of $E\{\Phi^2(X_1, \cdots, X_m)\}$ is sufficient for the asymptotic normality of U. No regularity conditions are imposed on the function Φ. This point is important for some applications (cf. section 9).

Theorem 7.1 and the following theorems of sections 7 and 8 are closely related to recent results of von Mises [18] which were published after this paper was essentially completed. It will be seen below (Theorem 7.4) that the limiting distribution of $\sqrt{n}[U - \theta(F)]$ is the same as that of $\sqrt{n}[\theta(S) - \theta(F)]$ (cf. (4.5)) if the variance of $\theta(S)$ exists. $\theta(S)$ is a differentiable statistical function in the sense of von Mises, and by Theorem I of [18], $\sqrt{n}[\theta(S) - \theta(F)]$ is asymptotically normal if certain conditions are satisfied. It will be found that in certain cases, for instance if the kernel Φ of θ is a polynomial, the conditions of the theorems of sections 7 and 8 are somewhat weaker than those of von Mises' theorem. Though von Mises' paper is concerned with functionals of univariate d.f.'s only, its results can easily be extended to the multivariate case.

For the particular case of a discrete population (where F is a step function), U and $\theta(S)$ are polynomials in the sample frequencies, and their asymptotic distribution may be inferred from the fact that the joint distribution of the frequencies tends to the normal form (cf. also von Mises [18]).

In Theorem 7.1 the functions $\Phi^{(\gamma)}(x_1, \cdots, x_{m(\gamma)})$ are supposed to be symmetric. Since, as has been seen in section 4, any U-statistic with non-symmetric kernel can be written in the form (4.4) with a symmetric kernel, this restriction is not essential and has been made only for the sake of convenience. Moreover, in the condition of the existence of $E\{\Phi^2(X_1, \cdots, X_m)\}$, the symmetric kernel may be replaced by a non-symmetric one. For, if Φ is non-symmetric, and Φ_0 is the symmetric kernel defined by (3.3), $E\{\Phi_0^2(X_1, \cdots, X_m)\}$ is a linear combination of terms of the form $E\{\Phi(X_{\alpha_1}, \cdots, X_{\alpha_m})\ \Phi\ (X_{\beta_1}, \cdots, X_{\beta_m})\}$, whose existence follows from that of $E\{\Phi^2(X_1, \cdots, X_m)\}$ by Schwarz's inequality.

If the regular functional $\theta(F)$ is stationary for $F = F_0$, that is, if $\zeta_1 = \zeta_1(F_0) = 0$ (cf. section 5), the limiting normal distribution of $\sqrt{n}(U - \theta)$ is, according to Theorem 7.1, singular, that is, its variance is zero. As has been seen in section 5, $\sigma^2(U)$ need not be zero in this case, but may be of some order n^{-c}, $(c = 2, 3, \cdots, m)$, and the distribution of $n^{c/2}(U - \theta)$ may tend to a limiting form which is not normal. According to von Mises [18], it is a limiting distribution of type c, $(c = 2, 3, \cdots)$.

According to Theorem 5.2, $\sigma^2(U)$ exceeds its asymptotic value $m^2\zeta_1/n$ for any finite n. Hence, if we apply Theorem 7.1 for approximating the distribution of U when n is large but finite, we underestimate the variance of U. For many applications this is undesirable, and for such cases the following theorem, which is an immediate consequence of Theorem 7.1, will be more useful.

THEOREM 7.2. *Under the conditions of Theorem* 7.1, *and if*

$$\zeta_1^{(\gamma)} > 0, \qquad (\gamma = 1, \cdots, g),$$

the joint d.f. of

$$(U^{(1)} - \theta^{(1)})/\sigma(U^{(1)}), \cdots, (U^{(g)} - \theta^{(g)})/\sigma(U^{(g)})$$

tends, as $n \to \infty$, *to the g-variate normal d.f. with zero means and covariance matrix* $(\rho^{(\gamma,\delta)})$, *where*

$$\rho^{(\gamma,\delta)} = \lim_{n\to\infty} \frac{\sigma(U^{(\gamma)}, U^{(\delta)})}{\sigma(U^{(\gamma)})\sigma(U^{(\delta)})} = \frac{\zeta_1^{(\gamma,\delta)}}{\sqrt{\zeta_1^{(\gamma)}\zeta_1^{(\delta)}}}, \qquad (\gamma, \delta = 1, \cdots, g).$$

PROOF OF THEOREM 7.1. The existence of (7.4) entails that of

$$\zeta_m^{(\gamma)} = E\{\Phi^{(\gamma)}(X_1, \cdots, X_{m(\gamma)})\}^2 - (\theta^{(\gamma)})^2$$

which, by (5.19), (5.20) and (6.6), is sufficient for the existence of

$$\zeta_1^{(\gamma)}, \cdots, \zeta_{m-1}^{(\gamma)}, \quad \text{of} \quad \sigma^2(U^{(\gamma)}), \quad \text{and of} \quad \zeta_1^{(\gamma,\delta)} \leq \sqrt{\zeta_1^{(\gamma)}\zeta_1^{(\delta)}}.$$

Now, consider the g quantities

$$Y^{(\gamma)} = \frac{m(\gamma)}{\sqrt{n}} \sum_{\alpha=1}^{n} \Psi_1^{(\gamma)}(X_\alpha), \qquad (\gamma = 1, \cdots, g)$$

where $\Psi_1^{(\gamma)}(x)$ is defined by (6.2). $Y^{(1)}, \cdots, Y^{(g)}$ are sums of n independent, random variables with zero means, whose covariance matrix, by virtue of (6.3), is

$$\{\sigma(Y^{(\gamma)}, Y^{(\delta)})\} = \{m(\gamma)m(\delta)\zeta_1^{(\gamma,\delta)}\}. \tag{7.5}$$

By the Central Limit Theorem for vectors (cf. Cramér [1, p. 112]), the joint d.f. of $(Y^{(1)}, \cdots, Y^{(g)})$ tends to the normal g-variate d.f. with the same means and covariances.

Theorem 7.1 will be proved by showing that the g random variables

$$Z^{(\gamma)} = \sqrt{n}(U^{(\gamma)} - \theta^{(\gamma)}), \qquad (\gamma = 1, \cdots, g), \tag{7.6}$$

have the same joint limiting distribution as $Y^{(1)}, \cdots, Y^{(g)}$.

According to Lemma 7.1 it is sufficient to show that

$$\lim_{n\to\infty} E(Z^{(\gamma)} - Y^{(\gamma)})^2 = 0, \qquad (\gamma = 1, \cdots, n). \tag{7.7}$$

For proving (7.7), write

$$E\{Z^{(\gamma)} - Y^{(\gamma)}\}^2 = E\{Z^{(\gamma)}\}^2 + E\{Y^{(\gamma)}\}^2 - 2E\{Z^{(\gamma)}Y^{(\gamma)}\}. \tag{7.8}$$

By (5.13) we have

$$E\{Z^{(\gamma)}\}^2 = n\sigma^2(U^{(\gamma)}) = m^2(\gamma)\zeta_1^{(\gamma)} + O(n^{-1}), \tag{7.9}$$

and from (7.5),

$$E\{Y^{(\gamma)}\}^2 = m^2(\gamma)\zeta_1^{(\gamma)}. \tag{7.10}$$

By (7.2) and (6.1) we may write for (7.6)

$$Z^{(\gamma)} = \sqrt{n}\binom{n}{m(\gamma)}^{-1} \Sigma' \Psi^{(\gamma)}(X_{\alpha_1}, \cdots, X_{\alpha_{m(\gamma)}}),$$

and hence

$$E\{Z^{(\gamma)}Y^{(\gamma)}\} = m(\gamma)\binom{n}{m(\gamma)}^{-1} \sum_{\alpha=1}^{n} \sum' E\{\Psi_1^{(\gamma)}(X_\alpha)\Psi^{(\gamma)}(X_{\alpha_1}, \cdots, X_{\alpha_{m(\gamma)}})\}.$$

The term

$$E\{\Psi_1^{(\gamma)}(X_\alpha)\Psi^{(\gamma)}(X_{\alpha_1}, \cdots, X_{\alpha_{m(\gamma)}})\}$$

is $= \zeta_1^{(\gamma)}$ if

$$\alpha_1 = \alpha \quad \text{or} \quad \alpha_2 = \alpha \cdots \quad \text{or} \quad \alpha_{m(\gamma)} = \alpha \tag{7.11}$$

and 0 otherwise. For a fixed α, the number of sets $\{\alpha_1, \cdots, \alpha_{m(\gamma)}\}$ such that $1 \leq \alpha_1 < \cdots < \alpha_{m(\gamma)} \leq n$ and (7.11) is satisfied, is $\binom{n-1}{m(\gamma)-1}$. Thus,

$$E\{Z^{(\gamma)}Y^{(\gamma)}\} = m(\gamma)\binom{n}{m(\gamma)}^{-1} n\binom{n-1}{m(\gamma)-1}\zeta_1^{(\gamma)} = m^2(\gamma)\zeta_1^{(\gamma)}. \tag{7.12}$$

On inserting (7.9), (7.10), and (7.12) in (7.8), we see that (7.7) is true.

The concluding remark in Theorem 7.1 is a direct consequence of the definition of a non-singular distribution. The proof of Theorem 7.1 is complete.

Theorems 7.1 and 7.2 deal with the asymptotic distribution of $U^{(1)}, \cdots, U^{(g)}$, which are unbiased estimates of $\theta^{(1)}, \cdots, \theta^{(g)}$. The unbiasedness of a statistic is, of course, irrelevant for its asymptotic behavior, and the application of Lemma 7.1 leads immediately to the following extension of Theorem 7.1 to a larger class of statistics.

THEOREM 7.3. *Let*

$$U^{(g)\prime} = U^{(g)} + \frac{b_n^{(\gamma)}}{\sqrt{n}}, \qquad (\gamma = 1, \cdots, g), \tag{7.13}$$

where $U^{(\gamma)}$ is defined by (7.2) *and $b_n^{(\gamma)}$ is a random variable. If the conditions of Theorem* 7.1 *are satisfied, and* $\lim E\{b_n^{(\gamma)}\}^2 = 0$, $(\gamma = 1, \cdots, g)$, *then the joint distribution of*

$$\sqrt{n}(U^{(1)\prime} - \theta^{(1)}), \cdots, \sqrt{n}(U^{(g)\prime} - \theta^{(g)})$$

tends to the normal distribution with zero means and covariance matrix

$$\{m(\gamma)m(\delta)\zeta_1^{(\gamma,\delta)}\}.$$

This theorem applies, in particular, to the regular functionals $\theta(S)$ of the sample d.f.,

$$\theta(S) = \frac{1}{n^m} \sum_{\alpha_1=1}^{n} \cdots \sum_{\alpha_m=1}^{n} \Phi(X_{\alpha_1}, \cdots, X_{\alpha_m}),$$

in the case that the variance of $\theta(S)$ exists. For we may write

$$n^m \theta(S) = \binom{n}{m} U + \Sigma^* \Phi(X_{\alpha_1}, \cdots, X_{\alpha_m}),$$

where the sum Σ^* is extended over all m-tuplets $(\alpha_1, \cdots, \alpha_m)$ in which at least one equality $\alpha i = \alpha_j (i \neq j)$ is satisfied. The number of terms in Σ^* is of order n^{m-1}. Hence

$$\theta(S) - U = \frac{1}{n} D,$$

where the expected value $E\{D^2\}$, whose existence follows from that of $\sigma^2\{\theta(S)\}$, is bounded for $n \to \infty$. Thus, if we put $U^{(\gamma)'} = \theta^{(\gamma)}(S)$, the conditions of Theorem 7.3 are fulfilled. We may summarize this result as follows:

THEOREM 7.4. *Let $X_1, \cdots, X_n$ be a random sample from an r-variate population with d.f. $F(x) = F(x^{(1)}, \cdots, x^{(r)})$, and let*

$$\theta^{(\gamma)}(F) = \int \cdots \int \Phi^{(\gamma)}(x_1, \cdots, x_{m(\gamma)})\, dF(x_1) \cdots dF(x_m{}^{(\gamma)}), \qquad (\gamma = 1, \cdots, g),$$

be g regular functionals of F, where $\Phi^{(\gamma)}(x_1, \cdots, x_{m(\gamma)})$ is symmetric in the vectors $x_1, \cdots, x_{m(\gamma)}$ and does not involve n. If $S(x)$ is the d.f. of the random sample, and if the variance of

$$\theta^{(\gamma)}(S) = \frac{1}{n^m} \sum_{\alpha_1=1}^{n} \cdots \sum_{\alpha_{m(\gamma)}=1}^{n} \Phi^{(\gamma)}(X_{\alpha_1}, \cdots, X_{\alpha_{m(\gamma)}})$$

exists, the joint d.f. of

$$\sqrt{n}\{\theta^{(1)}(S) - \theta^{(1)}(F)\}, \cdots, \sqrt{n}\{\theta^{(g)}(S) - \theta^{(g)}(F)\}$$

tends to the g-variate normal d.f. with zero means and covariance matrix

$$\{m(\gamma)m(\delta)\zeta_1^{(\gamma,\delta)}\}.$$

The following theorem is concerned with the asymptotic distribution of a function of statistics of the form U or U'.

THEOREM 7.5. *Let $(U') = (U^{(1)'}, \cdots, U^{(g)'})$ be a random vector, where $U^{(\gamma)'}$ is defined by* (7.13), *and suppose that the conditions of Theorem* 7.3 *are satisfied. If the function $h(y) = h(y^{(1)}, \cdots, y^{(g)})$ does not involve n and is continuous together with its second order partial derivatives in some neighborhood of the point $(y) = (\theta) = (\theta^{(1)}, \cdots, \theta^{(g)})$, then the distribution of the random variable $\sqrt{n}\{h(U') - h(\theta)\}$ tends to the normal distribution with mean zero and variance*

$$\sum_{\gamma=1}^{g} \sum_{\delta=1}^{g} m(\gamma)m(\delta) \left(\frac{\partial h(y)}{\partial y^{(\gamma)}}\right)_{y=\theta} \left(\frac{\partial h(y)}{\partial y^{(\delta)}}\right)_{y=\theta} \zeta_1^{(\gamma,\delta)}.$$

Theorem 7.5 follows from Theorem 7.3 in exactly the same way as the theorem on the asymptotic distribution of a function of moments follows from the fact of their asymptotic normality; cf. Cramér [2, p. 366]. We shall therefore omit the proof of Theorem 7.5. Since any moment whose variance exists has the form $U' = \theta(S)$ (cf. section 4 and Theorem 7.4), Theorem 7.5 is a generalization of the theorem on a function of moments.

8. Limit theorems for $U(X_1, \cdots, X_n)$ when the X_α's have different distributions. The limit theorems of the preceding section can be extended to the case when the X_α's have different distributions. We shall only prove an extension to this case of Theorem 7.1 (or 7.2), confining ourselves, for the sake of simplicity, to the distribution of a single U-statistic.

The extension of Theorems 7.3 and 7.5 with $g = 1$ to this case is immediate. One has only to replace the reference to Theorem 7.1 by that to the following Theorem 8.1, and θ and ζ_1 by $E\{U\}$ and $\zeta_{1,n}$.

THEOREM 8.1. *Let $X_1, \cdots, X_n$ be n independent random vectors of r components, X_α having the d.f. $F_\alpha(x) = F_\alpha(x^{(1)}, \cdots, x^{(r)})$. Let $\Phi(x_1, \cdots, x_m)$ be a function symmetric in its m vector arguments $x_\beta = (x_\beta^{(1)}, \cdots, x_\beta^{(r)})$ which does not involve n, and let*

$$\bar{\Psi}_{1(\nu)}(x) = \binom{n-1}{m-1}^{-1} \sum_{(\neq \nu)}' \Psi_{1(\nu)\alpha_1,\cdots,\alpha_{m-1}}(x), \qquad (\nu = 1, \cdots, n), \tag{8.1}$$

where Ψ is defined by (5.15), *and the summation is extended over all subscripts α such that*

$$1 \leq \alpha_1 < \alpha_2 < \cdots < \alpha_{m-1} \leq n, \qquad \alpha_i \neq \nu, \qquad (i = 1, \cdots, m).$$

Suppose that there is a number A such that for every $n = 1, 2, \cdots$

$$\int \cdots \int \Phi^2(x_1, \cdots, x_m)\, dF_{\alpha_1}(x_1) \cdots dF_{\alpha_m}(x_m) < A, \tag{8.2}$$

$$(1 \leq \alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_m \leq n),$$

that

$$E\,|\bar{\Psi}^3_{1(\nu)}(X_\nu)\,| < \infty, \qquad (\nu = 1, 2, \cdots, n), \tag{8.3}$$

and

$$\lim_{n\to\infty} \sum_{\nu=1}^{n} E\,|\bar{\Psi}^3_{1(\nu)}(X_\nu)\,| \Big/ \left\{\sum_{\nu=1}^{n} E\{\bar{\Psi}^2_{1(\nu)}(X_\nu)\}\right\}^{3/2} = 0. \tag{8.4}$$

Then, as $n \to \infty$, the d.f. of $(U - E\{U\})/\sigma(U)$ tends to the normal d.f. with mean 0 *and variance* 1.

The proof is similar to that of Theorem 7.1.

Let

$$W = \frac{m}{n} \sum_{\nu=1}^{n} \bar{\Psi}_{1(\nu)}(X_\nu).$$

It will be shown that
(a) the d.f. of

$$V = \frac{W - E\{W\}}{\sigma(W)}$$

tends to the normal d.f. with mean 0 and variance 1, and that
(b) the d.f. of

$$V' = \frac{U - E\{U\}}{\sigma(U)}$$

tends to the same limit as the d.f. of V.

Part (a) follows immediately from (8.3) and (8.4) by Liapounoff's form of the Central Limit Theorem.

According to Lemma 7.1, (b) will be proved when it is shown that

$$\lim_{n\to\infty} E\{V' - V\}^2 = \lim\left\{2 - 2\,\frac{\sigma(U, W)}{\sigma(U)\sigma(W)}\right\} = 0$$

or

$$\lim_{n\to\infty} \frac{\sigma(U, W)}{\sigma(U)\sigma(W)} = 1. \tag{8.5}$$

Let c be an integer, $1 \leq c \leq m$, and write

$$\boldsymbol{x} = (x_1, \cdots, x_c), \qquad \boldsymbol{y} = (y_1, \cdots, y_{m-c}), \qquad \boldsymbol{z} = (z_1, \cdots, z_{m-c})$$

$$F_{(\alpha)}(\boldsymbol{x}) = F_{\alpha_1}(x_1) \cdots F_{\alpha_c}(x_c), \qquad F_{(\beta)}(\boldsymbol{y}) = F_{\beta_1}(y_1) \cdots F_{\beta_{m-c}}(y_{m-c}),$$

$$F_{(\gamma)}(\boldsymbol{z}) = F_{\gamma_1}(z_1) \cdots F_{\gamma_{m-c}}(z_{m-c}).$$

Then, by Schwarz's inequality,

$$\int \cdots \int \Phi(\boldsymbol{x}, \boldsymbol{y})\Phi(\boldsymbol{x}, \boldsymbol{z})\, dF_{(\alpha)}(\boldsymbol{x})\, dF_{(\beta)}(\boldsymbol{y})\, dF_{(\gamma)}(\boldsymbol{z})$$

$$\leq \left\{\int \cdots \int \Phi^2(\boldsymbol{x}, \boldsymbol{y})\, dF_{(\alpha)}(\boldsymbol{x})\, dF_{(\beta)}(\boldsymbol{y}) \cdot \int \cdots \int \Phi^2(\boldsymbol{x}, \boldsymbol{z})\, dF_{(\alpha)}(\boldsymbol{x})\, dF_{(\gamma)}(\boldsymbol{z})\right\}^{\frac{1}{2}},$$

which, by (8.2), is $< A$ for any set of subscripts.

By the inequality for moments, $\theta_{\alpha_1,\cdots,\alpha_m}$, as defined by (5.14), is also uniformly bounded, and applying these inequalities to (5.16), it follows that there exists a number B such that

$$|\zeta_{c(\alpha_1,\cdots,\alpha_c)\beta_1,\cdots,\beta_{m-c};\,\gamma_1,\cdots,\gamma_{m-c}}| < B, \qquad (c = 1, \cdots, m), \tag{8.6}$$

for every set of subscripts satisfying the inequalities

$$\alpha_g \neq \alpha_h, \quad \beta_g \neq \beta_h, \quad \gamma_g \neq \gamma_h \quad \text{if } g \neq h, \quad \alpha_i \neq \beta_j, \quad \alpha_i \neq \gamma_j,$$
$$(i = 1, \cdots, c; j = 1, \cdots, m - c).$$

Now, we have

$$E\{W\} = 0$$

and

$$\sigma^2(W) = \frac{m^2}{n^2} \sum_{\nu=1}^{n} E\{\bar{\Psi}^2_{1(\nu)}(X_\nu)\} \tag{8.7}$$

or, inserting (8.1) and recalling (5.16),

$$\sigma^2(W) = \frac{m^2}{n^2}\binom{n-1}{m-1}^{-2} \sum_{\nu=1}^{n} \sum_{(\neq\nu)}' \sum_{(\neq\nu)}' \zeta_{1(\nu)\alpha_1,\cdots,\alpha_{m-1};\beta_1,\cdots,\beta_{m-1}}, \tag{8.8}$$

the two sums Σ' being over $\alpha_1 < \cdots < \alpha_{m-1}$, $(\alpha_i \neq \nu)$, and $\beta_1 < \cdots < \beta_{m-1}$, $(\beta_i \neq \nu)$, respectively. By (5.17), the sum of the terms whose subscripts $\nu, \alpha_1, \cdots, \alpha_{m-1}, \beta_1, \cdots, \beta_{m-1}$ are all different is equal to

$$\frac{n(n-1)\cdots(n-2m+2)}{(m-1)!(m-1)!}\zeta_{1,n} = n\binom{n-1}{m-1}\binom{n-m}{m-1}\zeta_{1,n}.$$

The number of the remaining terms is of order n^{2m-2}. Since, by (8.6), they are uniformly bounded, we have

$$\sigma^2(W) = \frac{m^2}{n}\zeta_{1,n} + O(n^{-2}). \tag{8.9}$$

Similarly, we have from (5.18)

$$\sigma^2(U) = \frac{m^2}{n}\zeta_{1,n} + O(n^{-2}),$$

and hence

$$\sigma(U) = \sigma(W) + O(n^{-1}). \tag{8.10}$$

The covariance of U and W is

$$\sigma(U, W) = \binom{n}{m}^{-1}\frac{m}{n}\sum_{\nu=1}^{n}\sum' E\{\bar{\Psi}_{1(\nu)}(X_\nu)\Psi_{m(\alpha_1,\cdots,\alpha_m)}(X_{\alpha_1}, \cdots, X_{\alpha_m})\}. \tag{8.11}$$

All terms except those in which one of the α's $= \nu$, vanish, and for the remaining ones we have, for fixed $\alpha_1, \cdots, \alpha_m$,

$$\begin{aligned} &E\{\bar{\Psi}_{1(\nu)}(X_\nu)\Psi_{m(\alpha_1,\cdots,\alpha_m)}(X_{\alpha_1}, \cdots, X_{\alpha_m})\} \\ &\quad = \binom{n-1}{m-1}^{-1}\sum_{(\neq\nu)}' E\{\Psi_{1(\nu)\beta_1,\cdots,\beta_{m-1}}(X_\nu)\Psi_{1(\nu)\gamma_1,\cdots,\gamma_{m-1}}(X_\nu)\} \\ &\quad = \binom{n-1}{m-1}^{-1}\sum_{(\neq\nu)}' \zeta_{1(\nu)\beta_1,\cdots,\beta_{m-1};\gamma_1,\cdots,\gamma_{m-1}} \end{aligned}$$

where the summation sign refers to the β's, and $\gamma_1, \cdots, \gamma_{m-1}$ are the α's that are $\neq \nu$. Inserting this in (8.11) and comparing the result with (8.8), we see that

$$\sigma(U, W) = \sigma^2(W). \tag{8.12}$$

From (8.12) and (8.10) we have

$$\frac{\sigma(U, W)}{\sigma(U)\sigma(W)} = \frac{\sigma(W)}{\sigma(U)} = \frac{n\sigma(W)}{n\sigma(W) + O(1)}.$$

Comparing condition (8.4) with (8.7), we see that we must have $n\sigma(W) \to \infty$ as $n \to \infty$. This shows the truth of (8.5). The proof of Theorem 8.1 is complete.

For some purposes the following corollary of Theorem 8.1 will be useful, where the conditions (8.2), (8.3), and (8.4) are replaced by other conditions which are more restrictive, but easier to apply.

THEOREM 8.2. *Theorem* 8.1 *holds if the conditions* (8.2), (8.3), *and* (8.4) *are replaced by the following:*
There exist two positive numbers C, D *such that*

$$\int \cdots \int |\Phi^3(x_1, \cdots, x_m)| \, dF_{\alpha_1}(x_1) \cdots dF_{\alpha_m}(x_m) < C \tag{8.13}$$

for $\alpha_i = 1, 2, \cdots, (i = 1, \cdots, m)$, *and*

$$\zeta_{1(\nu)\alpha_1,\cdots,\alpha_{m-1};\beta_1,\cdots,\beta_{m-1}} > D \tag{8.14}$$

for any subscripts satisfying

$$1 \leq \alpha_1 < \alpha_2 < \cdots < \alpha_{m-1}, \quad 1 \leq \beta_1 < \beta_2 < \cdots < \beta_{m-1}, \quad 1 \leq \nu \neq \alpha_i, \beta_i.$$

We have to show that (8.2), (8.3), and (8.4) follow from (8.13) and (8.14).

(8.13) implies (8.2) by the inequality for moments. By a reasoning analogous to that used in the previous proof, applying Hölder's inequality instead of Schwarz's inequality, it follows from (8.13) that

$$E|\bar{\Psi}^3_{1(\nu)}(X_\nu)| < C'. \tag{8.15}$$

On the other hand, by (8.7), (8.8), and (8.14),

$$\sum_{\nu=1}^{n} E\{\bar{\Psi}^2_{1(\nu)}(X_\nu)\} > nD. \tag{8.16}$$

(8.15) and (8.16) are sufficient for the fulfillment of (8.4).

9. Applications to particular statistics.

(a) *Moments and functions of moments.* It has been seen in section 4 that the k-statistics and the unbiased estimates of moments are U-statistics, while the sample moments are regular functionals of the sample d.f. By Theorems 7.1, 8.1, and 7.4 these statistics are asymptotically normally distributed, and by Theorem 7.5 the same is true for a function of moments, if the respective conditions are satisfied. These results are not new (cf., for example, Cramér [2]).

(b) *Mean difference and coefficient of concentration.* If $Y_1, \cdots, Y_n$ are n independent real-valued random variables, Gini's mean difference (without repetition) is defined by

$$d = \frac{1}{n(n-1)} \sum_{\alpha \neq \beta} |Y_\alpha - Y_\beta|.$$

If the Y_α's have the same distribution F, the mean of d is

$$\delta = \int\int |y_1 - y_2| \, dF(y_1) \, dF(y_2),$$

and the variance, by (5.13) is

$$\sigma^2(d) = \frac{2}{n(n-1)} \{2\zeta_1(\delta)(n-2) + \zeta_2(\delta)\},$$

where

$$\zeta_1(\delta) = \int \left\{ \int |y_1 - y_2| \, dF(y_2) \right\}^2 dF(y_1) - \delta^2, \tag{9.1}$$

$$\zeta_2(\delta) = \int\int (y_1 - y_2)^2 \, dF(y_1) \, dF(y_2) - \delta^2 = 2\sigma^2(Y) - \delta^2. \tag{9.2}$$

The notation $\zeta_1(\delta)$, $\zeta_2(\delta)$ serves to indicate the relation of these functionals of F to the functional $\delta(F)$; δ is here merely the symbol of the functional, not a particular value of it. In a similar way we shall write $\Phi(y_1, y_2 \mid \delta) = |y_1 - y_2|$, etc. When there is danger of confusing $\zeta_1(\delta)$ with $\zeta_1(F)$, we may write $\zeta_1(F \mid \delta)$.

U. S. Nair [19] has evaluated $\sigma^2(d)$ for several particular distributions.

By Theorem 7.1, $\sqrt{n}(d - \delta)$ is asymptotically normal if $\zeta_2(\delta)$ exists.

If $Y_1, \cdots, Y_n$ do not assume negative values, the coefficient of concentration (cf. Gini [8]) is defined by

$$G = \frac{d}{2\bar{Y}},$$

where $\bar{Y} = \Sigma Y_\alpha / n$. G is a function of two U-statistics. If the Y_α's are identically distributed, if $E\{Y^2\}$ exists, and if $\mu = E\{Y\} > 0$, then, by Theorem 7.5, $\sqrt{n}(G - \delta/2\mu)$ tends to be normally distributed with mean 0 and variance

$$\frac{\delta^2}{4\mu^4} \zeta_1(\mu) - \frac{\delta}{\mu^3} \zeta_1(\mu, \delta) + \frac{1}{\mu^2} \zeta_1(\delta),$$

where

$$\zeta_1(\mu) = \int y^2 \, dF(y) - \mu^2 = \sigma^2(Y),$$

$$\zeta_1(\mu, \delta) = \int\int y_1 |y_1 - y_2| \, dF(y_1) \, dF(y_2) - \mu\delta,$$

and $\zeta_1(\delta)$ is given by (9.1).

(c) *Functions of ranks and of the signs of variate differences.* Let $s(u)$ be the signum function,

$$s(u) = \begin{cases} -1 \text{ if } u < 0; \\ 0 \text{ if } u = 0; \\ 1 \text{ if } u > 0, \end{cases} \tag{9.3}$$

and let

$$c(u) = \tfrac{1}{2}\{1 + s(u)\} = \begin{cases} 0 \text{ if } u < 0; \\ \tfrac{1}{2} \text{ if } u = 0; \\ 1 \text{ if } u > 0. \end{cases} \tag{9.4}$$

If

$$x_\alpha = (x_\alpha^{(1)}, \cdots, x_\alpha^{(r)}), \qquad (\alpha = 1, \cdots, n)$$

is a sample of n vectors of r components, we may define the *rank* $R_\alpha^{(i)}$ of $x_\alpha^{(i)}$ by

$$\begin{aligned} R_\alpha^{(i)} &= \tfrac{1}{2} + \sum_{\beta=1}^{n} c(x_\alpha^{(i)} - x_\beta^{(i)}) \\ &= \frac{n+1}{2} + \tfrac{1}{2}\sum_{\beta=1}^{n} s(x_\alpha^{(i)} - x_\beta^{(i)}), \quad (i = 1, \cdots, r). \end{aligned} \tag{9.5}$$

If the numbers $x_1^{(i)}, x_2^{(i)}, \cdots, x_n^{(i)}$ are all different, the smallest of them has rank 1, the next smallest rank 2, etc. If some of them are equal, the rank as defined by (9.5) is known as the mid-rank.

Any function of the ranks is a function of expressions $c(x_\alpha^{(i)} - x_\beta^{(i)})$ or $s(x_\alpha^{(i)} - x_\beta^{(i)})$.

Conversely, since

$$s(x_\alpha^{(i)} - x_\beta^{(i)}) = s(R_\alpha^{(i)} - R_\beta^{(i)}),$$

any function of expressions $s(x_\alpha^{(i)} - x_\beta^{(i)})$ or $c(x_\alpha^{(i)} - x_\beta^{(i)})$ is a function of the ranks.

Consider a regular functional $\theta(F)$ whose kernel $\Phi(x_1, \cdots, x_m)$ depends only on the signs of the variate differences,

$$s(x_\alpha^{(i)} - x_\beta^{(i)}), \qquad (\alpha, \beta = 1, \cdots, m; i = 1, \cdots, r). \tag{9.6}$$

The corresponding U-statistic is a function of the ranks of the sample variates.

The function Φ can take only a finite number of values, $c_1, \cdots, c_N$, say. If $\pi_i = P\{\Phi = c_i\}$, $(i = 1, \cdots, N)$, we have

$$\theta = c_1\pi_1 + \cdots + c_N\pi_N, \qquad \sum_{i=1}^{N} \pi_i = 1.$$

π_i is a regular functional whose kernel $\Phi_i(x_1, \cdots, x_m)$ is equal to 1 or 0 according to whether $\Phi = c_i$ or $\neq c_i$. We have

$$\Phi = c_1\Phi_1 + \cdots + c_N\Phi_N .$$

In order that $\theta(F)$ exist, the c_i must be finite, and hence Φ is bounded. Therefore, $E\{\Phi^2\}$ exists, and if $X_1, X_2, \cdots$ are identically distributed, the d.f. of $\sqrt{n}(U - \theta)$ tends, by Theorem 7.1, to a normal d.f. which is non-singular if $\zeta_1 > 0$.

In the following we shall consider several examples of such functionals.

(d) *Difference sign correlation.* Consider the bivariate sample

$$(x_1^{(1)}, x_1^{(2)}), (x_2^{(1)}, x_2^{(2)}), \cdots, (x_n^{(1)}, x_n^{(2)}). \tag{9.7}$$

To each two members of this sample corresponds a pair of signs of the differences of the respective variables,

$$s(x_\alpha^{(1)} - x_\beta^{(1)}),\ s(x_\alpha^{(2)} - x_\beta^{(2)}), \qquad (\alpha \neq \beta;\ \alpha, \beta = 1, \cdots, n). \tag{9.8}$$

(9.8) is a population of $n(n-1)$ pairs of difference signs. Since

$$\sum_{\alpha \neq \beta} s(x_\alpha^{(i)} - x_\beta^{(i)}) = 0, \qquad (i = 1, 2),$$

the covariance t of the difference signs (9.8) is

$$t = \frac{1}{n(n-1)} \sum_{\alpha \neq \beta} s(x_\alpha^{(1)} - x_\beta^{(1)}) s(x_\alpha^{(2)} - x_\beta^{(2)}). \tag{9.9}$$

t will be briefly referred to as the *difference sign covariance* of the sample (9.7).

If all $x^{(1)}$'s and all $x^{(2)}$'s are different, we have

$$\sum_{\alpha \neq \beta} s^2(x_\alpha^{(i)} - x_\beta^{(i)}) = n(n-1), \qquad (i = 1, 2),$$

and then t is the product moment correlation of the difference signs.

It is easily seen that t is a linear function of the number of inversions in the permutation of the ranks of $x^{(1)}$ and $x^{(2)}$.

The statistic t has been considered by Esscher [6], Lindeberg [15], [16], Kendall [12], and others.

t is a U-statistic. As a function of a random sample from a bivariate population, t is an unbiased estimate of the regular functional of degree 2,

$$\tau = \int\int\int\int s(x_1^{(1)} - x_2^{(1)}) s(x_1^{(2)} - x_2^{(2)})\, dF(x_1)\, dF(x_2). \tag{9.10}$$

τ is the covariance of the signs of differences of the corresponding components of $X_1 = (X_1^{(1)}, X_1^{(2)})$ and $X_2 = (X_2^{(1)}, X_2^{(2)})$ in the population of pairs of independent vectors X_1, X_2 with identical d.f. $F(x) = F(x^{(1)}, x^{(2)})$. If $F(x^{(1)}, x^{(2)})$ is continuous, τ is the product moment correlation of the difference signs.

Two points (or vectors), $(x_1^{(1)}, x_1^{(2)})$ and $(x_2^{(1)}, x_2^{(2)})$ are called concordant or discordant according to whether

$$(x_1^{(1)} - x_2^{(1)})(x_1^{(2)} - x_2^{(2)})$$

is positive or negative. If $\pi^{(c)}$ and $\pi^{(d)}$ are the probabilities that a pair of vectors drawn at random from the population is concordant or discordant, respectively, we have from (9.10)

$$\tau = \pi^{(c)} - \pi^{(d)}.$$

If $F(x^{(1)}, x^{(2)})$ is continuous, we have $\pi^{(c)} + \pi^{(d)} = 1$, and hence

$$\tau = 2\pi^{(c)} - 1 = 1 - 2\pi^{(d)}. \tag{9.11}$$

If we put

$$\bar{F}(x^{(1)}, x^{(2)}) = \tfrac{1}{4}\{F(x^{(1)} - 0, x^{(2)} - 0) + F(x^{(1)} - 0, x^{(2)} + 0) + F(x^{(1)} + 0, x^{(2)} - 0) + F(x^{(1)} + 0, x^{(2)} + 0)\}, \tag{9.12}$$

we have

$$\Phi_1(x \mid \tau) = 1 - 2\bar{F}(x^{(1)}, \infty) - 2\bar{F}(\infty, x^{(2)}) + 4\bar{F}(x^{(1)}, x^{(2)}), \tag{9.13}$$

and we may write

$$\tau = E\{\Phi_1(X_1 \mid \tau)\}. \tag{9.14}$$

The variance of t is, by (5.13),

$$\sigma^2(t) = \frac{2}{n(n-1)}\{2\zeta_1(\tau)(n-2) + \zeta_2(\tau)\}, \tag{9.15}$$

where

$$\zeta_1(\tau) = E\{\Phi_1^2(X_1 \mid \tau)\} - \tau^2, \tag{9.16}$$

$$\zeta_2(\tau) = E\{s^2(X_1^{(1)} - X_2^{(1)})s^2(X_1^{(2)} - X_2^{(2)})\} - \tau^2. \tag{9.17}$$

If $F(x^{(1)}, x^{(2)})$ is continuous, we have $\zeta_2(\tau) = 1 - \tau^2$, and $\bar{F}(x^{(1)}, x^{(2)})$ in (9.13) may be replaced by $F(x^{(1)}, x^{(2)})$.

The variance of a linear function of t has been given for the continuous case by Lindeberg [15], [16].

If $X^{(1)}$ and $X^{(2)}$ are independent and have a continuous d.f., we find $\zeta_1(\tau) = \frac{1}{9}$, $\zeta_2(\tau) = 1$, and hence

$$\sigma^2(t) = \frac{2(2n+5)}{9n(n-1)}. \tag{9.18}$$

In this case the distribution of t is independent of the univariate distributions of $X^{(1)}$ and $X^{(2)}$. This is, however, no longer true if the independent variables are discontinuous. Then it appears that $\sigma^2(t)$ depends on $P\{X_1^{(i)} = X_2^{(i)}\}$ and $P\{X_1^{(i)} = X_2^{(i)} = X_3^{(i)}\}$, $(i = 1, 2)$.

By Theorem 7.1, the d.f. of $\sqrt{n}(t - \tau)$ tends to the normal form. This result has first been obtained for the particular case that all permutations of the ranks of $X^{(1)}$ and $X^{(2)}$ are equally probable, which corresponds to the independence of the continuous random variables $X^{(1)}$, $X^{(2)}$ (Kendall [12]). In this case t can be represented as a sum of independent random variables (cf. Dantzig [5] and Feller [7]). In the general case the asymptotic normality of t has been shown by Daniels and Kendall [4] and the author [10].

The functional $\tau(F)$ is stationary (and hence the normal limiting distribution of $\sqrt{n}(t - \tau)$ singular) if $\zeta_1 = 0$, which, in the case of a continuous F, means that the equation $\Phi_1(X \mid \tau) = \tau$ or

$$4F(X^{(1)}, X^{(2)}) = 2F(X^{(1)}, \infty) + 2F(\infty, X^{(2)}) - 1 + \tau \tag{9.19}$$

is satisfied with probability 1. This is the case if $X^{(2)}$ is an increasing function of $X^{(1)}$. Then $t = \tau = 1$ with probability 1, and $\sigma^2(t) = 0$. A case where (9.19) is fulfilled and $\sigma^2(t) > 0$ is the following: $X^{(1)}$ is uniformly distributed in the interval (0, 1), and

$$X^{(2)} = X^{(1)} + \tfrac{1}{2} \text{ if } 0 \leq X^{(1)} < \tfrac{1}{2},\ X^{(2)} = X^{(1)} - \tfrac{1}{2} \text{ if } \tfrac{1}{2} \leq X^{(1)} \leq 1. \tag{9.20}$$

In this case $\tau = 0, \zeta_2 = 1, \sigma^2(t) = 2/n(n-1)$.

(e) *Rank correlation and grade correlation.* If in the sample $\{(x_\alpha^{(1)}, x_\alpha^{(2)})\}$, $(\alpha = 1, \cdots, n)$, all $x_\alpha^{(1)}$'s and all $x_\alpha^{(2)}$'s are different, the rank correlation coefficient, which we denote by k', is given by

$$k' = \frac{12}{n^3 - n} \sum_{\alpha=1}^{n} \left(R_\alpha^{(1)} - \frac{n+1}{2}\right)\left(R_\alpha^{(2)} - \frac{n+1}{2}\right).$$

Inserting (9.5) we have

$$k' = \frac{3}{n^3 - n} \sum_{\alpha=1}^{n} \sum_{\beta=1}^{n} \sum_{\gamma=1}^{n} s(x_\alpha^{(1)} - x_\beta^{(1)}) s(x_\alpha^{(2)} - x_\gamma^{(2)})$$

or

$$k' = \frac{(n-2)k + 3t}{n+1} \tag{9.21}$$

where t is the difference sign covariance (9.9), and

$$k = \frac{3}{n(n-1)(n-2)} \sum{}'' s(x_\alpha^{(1)} - x_\beta^{(1)}) s(x_\alpha^{(2)} - x_\gamma^{(2)}),$$

the summation being over all different subscripts α, β, γ.

k is a U-statistic, and as a function of a random sample from a population with d.f. F, k is an unbiased estimate of the regular functional of degree 3,

$$\begin{aligned} \kappa &= 3 \int \cdots \int s(x_1^{(1)} - x_2^{(1)}) s(x_1^{(2)} - x_3^{(2)})\, dF(x_1)\, dF(x_2)\, dF(x_3) \\ &= 3 \int\int \{2\bar{F}^{(1)}(x^{(1)}) - 1\}\{2\bar{F}^{(2)}(x^{(2)}) - 1\}\, dF(x), \end{aligned} \tag{9.22}$$

where $\bar{F}^{(1)}(x^{(1)}) = \bar{F}(x^{(1)}, \infty)$, $\bar{F}^{(2)}(x^{(2)}) = \bar{F}(\infty, x^{(2)})$.

If F is continuous, we have

$$\int \bar{F}^{(i)}(y)\, d\bar{F}^{(i)}(y) = \int_0^1 u\, du = \tfrac{1}{2},$$

$$\int \{\bar{F}^{(i)}(y) - \tfrac{1}{2}\}^2\, d\bar{F}^{(i)}(u) = \int_0^1 (u - \tfrac{1}{2})^2\, du = \tfrac{1}{12}, \qquad (i = 1, 2),$$

and in this case κ is the coefficient of correlation between the random variables

$$U^{(1)} = F^{(1)}(X^{(1)}), \qquad U^{(2)} = F^{(2)}(X^{(2)}).$$

$U^{(i)}$ has been termed the grade of the continuous variable $X^{(i)}$, and in the general case $\bar{F}^{(i)}(X^{(i)})$ may be called the grade of $X^{(i)}$ (cf., for instance, G. U. Yule and M. G. Kendall [22, p. 150]). In general, κ is 12 times the covariance of the grades.

From (9.21) we have for the expected value of k',

$$E\{k'\} = \frac{(n-2)\kappa + 3\tau}{n+1}.$$

In the continuous case the rank correlation coefficient k' is an estimate of the grade correlation κ, which is biased for finite n but unbiased in the limit.

The kernel $3s(x_1^{(1)} - x_2^{(1)})s(x_1^{(2)} - x_3^{(2)})$ of κ is not symmetric. Denoting by $\Phi(x_1, x_2, x_3 \mid \kappa)$ the symmetric kernel of κ, we have

$$\Phi(x_1, x_2, x_3 \mid \kappa) = \frac{1}{2} \sum_{\substack{\alpha\neq\beta\neq\gamma \\ \alpha\neq\gamma}}^{1,2,3} s(x_\alpha^{(1)} - x_\beta^{(1)})s(x_\alpha^{(2)} - x_\gamma^{(2)}) \tag{9.23}$$

For computing κ and the constants ζ_i an alternative expression for κ and Φ is sometimes more convenient. From three two-dimensional vectors x_1, x_2, x_3 we can form three pairs (x_1, x_2), (x_1, x_3), and (x_2, x_3). The number of concordant pairs among them can be 3, 2, 1, or 0. If γ is the probability that among the three pairs formed from three random elements of the population at least 2 are concordant, we have, if the d.f. F is continuous,

$$\kappa = 2\gamma - 1. \tag{9.24}$$

This is analogous to the expression (9.11) for τ.

The truth of (9.24) can be seen as follows: From the definition of γ we have

$$\gamma = E\{\Phi(x_1, x_2, x_3 \mid \gamma)\},$$

where $\Phi(x_1, x_2, x_3 \mid \gamma)$ is $= 1$ if at least two of the three expressions

$$(x_\alpha^{(1)} - x_\beta^{(1)})(x_\alpha^{(2)} - x_\beta^{(2)}), \qquad (\alpha < \beta;\ \alpha, \beta = 1, 2, 3) \tag{9.25}$$

are positive, and equal to zero, if no more than one of them is positive. Since, by the continuity of F, we may neglect the case of (9.25) being zero, we may write

$$\Phi(x_1, x_2, x_3 \mid \gamma) = c_{12,12}c_{23,23}c_{31,31} + c_{12,12}c_{23,23}c_{31,13} + c_{12,12}c_{23,32}c_{31,31} + c_{12,21}c_{23,23}c_{31,31},$$

where

$$c_{\alpha,\beta,\gamma,\delta} = c[(x_\alpha^{(1)} - x_\beta^{(1)})(x_\gamma^{(2)} - x_\delta^{(2)})]$$

and $c(u)$ is defined by (9.4).

$\Phi(x_1, x_2, x_3 \mid \gamma)$ is symmetric in x_1, x_2, x_3.

The identity

$$\Phi(x_1, x_2, x_3 \mid \kappa) = 2\Phi(x_1, x_2, x_3 \mid \gamma) - 1 \tag{9.26}$$

can be shown to hold either by algebraical calculation using (9.4) or by direct computation of each side for the different positions of the three points x_1, x_2, x_3.

From (9.26) it appears that in the continuous case the symmetric kernel $\Phi(x_1, x_2, x_3 \mid \kappa)$ can assume only two values, -1 and $+1$.

The variance of k is, according to (5.13),

$$\sigma^2(k) = \frac{6}{n(n-1)(n-2)} \left\{ 3\binom{n-3}{2} \zeta_1(\kappa) + 3(n-3)\zeta_2(\kappa) + \zeta_3(\kappa) \right\},$$

where

$$\begin{aligned}
\zeta_1(\kappa) &= E\{\Phi_1^2(X_1 \mid \kappa)\} - \kappa^2, \\
\zeta_2(\kappa) &= E\{\Phi_2^2(X_1, X_2 \mid \kappa)\} - \kappa^2, \\
\zeta_3(\kappa) &= E\{\Phi^2(X_1, X_2, X_3 \mid \kappa)\} - \kappa^2, \\
\Phi_1(x_1 \mid \kappa) &= E\{\Phi(x_1, X_2, X_3 \mid \kappa)\}, \\
\Phi_2(x_1, x_2 \mid \kappa) &= E\{\Phi(x_1, x_2, X_3 \mid \kappa)\}.
\end{aligned}$$

We find for the continuous case

$$\zeta_3(\kappa) = 1 - \kappa^2,$$

$$\begin{aligned}
(9.27)\quad \Phi_1(x_1 \mid \kappa) &= [1 - 2F(x_1^{(1)}, \infty)][1 - 2F(\infty, x_1^{(2)})] - 2F(x_1^{(1)}, \infty) \\
&\quad - 2F(\infty, x_1^{(2)}) + 4\int F(x_1^{(1)}, y^{(2)}) dF(\infty, y^{(2)}) \\
&\quad + 4\int F(y^{(1)}, x_1^{(2)}) dF(y^{(1)}, \infty),
\end{aligned}$$

$$\begin{aligned}
\Phi_2(x_1, x_2 \mid \kappa) &= 1 + 2F(x_1^{(1)}, x_2^{(2)}) + 2F(x_2^{(1)}, x_1^{(2)}) - 2c(x_2^{(2)} - x_1^{(2)})F(x_1^{(1)}, \infty) \\
&\quad - 2c(x_1^{(2)} - x_2^{(2)})F(x_2^{(1)}, \infty) - 2c(x_2^{(1)} - x_1^{(1)})F(\infty, x_1^{(2)}) \\
&\quad - 2c(x_1^{(1)} - x_2^{(1)})F(\infty, x_2^{(2)}).
\end{aligned}$$

If $X^{(1)}, X^{(2)}$ are continuous and independent, we obtain $\kappa = 0$, $\zeta_1 = \frac{1}{9}$, $\zeta_2 = \frac{7}{18}$, $\zeta_3 = 1$, and hence

$$(9.28)\qquad \sigma^2(k) = \frac{n^2 - 3}{n(n-1)(n-2)}.$$

In the discontinuous case of independence the distribution of k, as that of t, depends on the distributions of $X^{(1)}$ and $X^{(2)}$, and $\sigma^2(k)$ can again be expressed in terms of $P\{X_1^{(i)} = X_2^{(i)}\}$ and $P\{X_1^{(i)} = X_2^{(i)} = X_3^{(i)}\}$, $(i = 1, 2)$.

The variance of the rank correlation coefficient k' is, by (9.21),

$$(9.29)\qquad \sigma^2(k') = \frac{(n-2)^2\sigma^2(k) + 6(n-2)\sigma(t, k) + 9\sigma^2(t)}{(n+1)^2}.$$

For $\sigma(t, k)$ we have, according to (6.5),

$$\sigma(t, k) = \frac{6}{n(n-1)} \{(n-3)\zeta_1(\tau, \kappa) + \zeta_2(\tau, \kappa)\},$$

where

$$\zeta_1(\tau, \kappa) = E\{\Phi_1(X_1 \mid \tau)\Phi_1(X_1 \mid \kappa)\} - \tau\kappa,$$
$$\zeta_2(\tau, \kappa) = E\{\Phi(X_1, X_2 \mid \tau)\Phi_2(X_1, X_2 \mid \kappa)\} - \tau\kappa.$$

In the case of independence we see from (9.13) and (9.27) that

$$\Phi_1(x \mid \tau) = \Phi_1(x \mid \kappa) = [1 - 2F(x^{(1)}, \infty)][1 - 2F(\infty, x^{(2)})],$$

and we obtain

$$\zeta_1(\tau, \kappa) = \zeta_1(\kappa) = \zeta_1(\tau) = \tfrac{1}{9}, \tag{9.30}$$
$$\zeta_2(\tau, \kappa) = \tfrac{5}{9},$$

$$\sigma(t, k) = \frac{2(n+2)}{3n(n-1)}. \tag{9.31}$$

On inserting (9.28), (9.31) and (9.18) in (9.29), we find

$$\sigma^2(k') = \frac{1}{n-1},$$

in accordance with the result obtained for this case by Student and published by K. Pearson [20].

According to Theorem 7.1, $\sqrt{n}(k - \kappa)$ tends to be normally distributed with mean 0 and variance $9\zeta_1(\kappa)$. The same is true for the distribution of the rank correlation coefficient, k', as follows from Theorem 7.3 in conjunction with (9.21). For the special case of independence the asymptotic normality of k' has been proved by Hotelling and Pabst [11].

From Theorem 7.3 it also follows that the joint distribution of $\sqrt{n}(t - \tau)$ and $\sqrt{n}(k - \kappa)$ (or $\sqrt{n}(k' - \kappa)$) tends to the normal form with the variances $4\zeta_1(\tau)$ and $9\zeta_1(\kappa)$ and the covariance $6\zeta_1(\kappa, \tau)$. In the case of independence we see from (9.30) that the correlation $\rho(t, k)$ between t and k tends to 1, and we have the asymptotic functional relation $3t = 2k$. This result has been conjectured by Kendall and others [14], and proved by Daniels [3]. In general, however, $\rho(t, k)$ does not approach unity. Thus, if $X^{(1)}$ is uniformly distributed in (0, 1), and

$$\begin{aligned} X^{(2)} &= \tfrac{1}{2} - X^{(1)} && \text{if } 0 \le X^{(1)} < \tfrac{1}{4}, \\ X^{(2)} &= \tfrac{1}{2} + X^{(1)} && \text{if } \tfrac{1}{4} \le X^{(1)} < \tfrac{1}{2}, \\ X^{(2)} &= X^{(1)} - \tfrac{1}{2} && \text{if } \tfrac{1}{2} \le X^{(1)} < \tfrac{3}{4}, \\ X^{(2)} &= \tfrac{3}{2} - X^{(1)} && \text{if } \tfrac{3}{4} \le X^{(1)} \le 1, \end{aligned} \tag{9.32}$$

we have $\tau = \kappa = 0$, $\zeta_1(\tau) = 0$, $\zeta_2(\tau) = 1$, $\zeta_1(\kappa) = \frac{1}{16}$, $\zeta_1(\kappa, \tau) = 0$, and hence $\rho(t, k) \to 0$.

(f) *Non-parametric tests of independence.* Suppose that the random variables $X^{(1)}$, $X^{(2)}$ have a continuous joint d.f. $F(x^{(1)}, x^{(2)})$, and we want to test the hypothesis H_0 that $X^{(1)}$ and $X^{(2)}$ are independent, that is, that

$$F(x^{(1)}, x^{(2)}) = F(x^{(1)}, \infty)\, F(\infty, x^{(2)}).$$

The distribution of any statistic involving only the ranks of the variables does not depend on the d.f. of the population when H_0 is true. For this reason several rank order statistics, among them the difference sign correlation t and the rank correlation k', have been suggested for testing independence.

From the preceding results we can obtain the asymptotic power functions of the tests of independence based on t and k'. If H_0 is true, we have $E\{t\} = \tau = 0$, and the critical region of size ϵ of the t-test may be defined by $|\, t \,| > c_n$, where c_n is the smallest number satisfying the inequality

$$P\{\, |\, t \,| > c_n \mid H_0\} \leq \epsilon. \tag{9.33}$$

By Theorem 7.2 and (9.18) we may write $c_n = 2\lambda_n/3\sqrt{n}$, where λ_n tends to a positive constant λ depending on ϵ.

Since $\sigma^2(t) = O(n^{-1})$, the power function

$$P_n(H) = P\{\, |\, t \,| \geq 2\lambda_n/3\sqrt{n} \mid H\}$$

tends to one as $n \to \infty$ for any alternative hypothesis H with $\tau(F) \neq 0$. If, however, $\tau = 0$, we have $\lim P_n(H) < 1$. If $\tau = 0$ and $\zeta_1(\tau) < \frac{1}{9}$, we have even $\lim P_n(H) < \epsilon$, and with respect to these alternatives the test is biased in the limit. Thus, in the case of the distribution (9.20) we have even $P_n(H) \to 0$. In this case there is a functional relationship between the variables, and the distribution must be considered as considerably different from the case of independence.

For the rank correlation test we have a similar result. If c_n' is the smallest number satisfying $P\{\, |\, k' \,| > c_n' \mid \mathrm{H}_0\} \leq \epsilon$, we have $c_n' = \lambda_n'/\sqrt{n}$, where $\lim \lambda_n' = \lambda$, and the test is biased in the limit if $\kappa = 0$ and $\zeta_1(\kappa) < \frac{1}{9}$. This is fulfilled in the case of the distribution (9.32), where $\zeta_1(\kappa) = \frac{1}{16}$.

The question arises whether there exist non-parametric tests of independence which are unbiased or unbiased in the limit. This point will be discussed in a separate paper on tests of independence.

(g) *Mann's test against trend.* Let $Y_1, \cdots, Y_n$ be n independent real-valued random variables, Y_α having the continuous d.f. $F_\alpha(y)$, $(\alpha = 1, \cdots, n)$. The hypothesis of randomness,

$$H_1: \qquad F_1(y) = \cdots = F_n(y)$$

is to be tested against the alternative hypothesis of a "downward trend,"

$$H_2: \qquad F_1(y) < F_2(y) < \cdots < F_n(y).$$

H. B. Mann [17] has suggested a test of H_1 against H_2 based on the number T of inequalities $Y_\alpha < Y_\beta$, where $\alpha < \beta$. We may write

$$2T - \frac{n(n-1)}{2} = \sum_{\alpha<\beta} s(Y_\beta - Y_\alpha) = \sum_{\alpha<\beta} s(\alpha - \beta) s(Y_\alpha - Y_\beta).$$

The U-statistic

$$t = \{4T/n(n-1)\} - 1$$

is the same as (9.9) for the special case when one component is not a random variable.

Let

$$\tau_{\alpha\beta} = s(\alpha - \beta) \iint s(y_1 - y_2)\, dF_\alpha(y_1)\, dF_\beta(y_2)$$

$$= s(\alpha - \beta)\left\{ 2 \int F_\beta(y)\, dF_\alpha(y) - 1 \right\}.$$

We have $\tau_{\alpha\beta} = 0$ if H_1 is true and $\tau_{\alpha\beta} < 0$ if H_2 is true.

Since

$$E\{t\} = \tau_n = \frac{2}{n(n-1)} \sum_{\alpha<\beta} \tau_{\alpha\beta},$$

it follows that $E\{t\} = 0$ under H_1 and $E\{t\} < 0$ under H_2.

Mann's test against trend has the power function $P_n(H) = P\{t < a_n \mid H\}$, where a_n is the largest number satisfying $P\{t < a_n \mid H_1\} \leq \epsilon$.

Since $a_n \to 0$ and, by (5.18), $\sigma^2(t) = O(n^{-1})$, it follows from Tchebycheff's inequality that the test is consistent (that is, $P_n(H_2) \to 1$) and hence unbiased in the limit. This has been shown by Mann who also gave sufficient conditions under which the test is unbiased for finite n.

By Theorems 8.1 and 8.2 the distribution of $(t - \tau_n)/\sigma(t)$ is asymptotically normal if certain conditions are satisfied. Since (8.2), (8.3) and (8.13) are fulfilled, either of the conditions (8.4) and (8.14) is sufficient.

(h) *The coefficient of partial difference sign correlation.* Consider a three-variate sample $x_1, \cdots, x_n$; $x_\alpha = (x_\alpha^{(1)}, x_\alpha^{(2)}, x_\alpha^{(3)})$, $(\alpha = 1, \cdots, n)$. In a similar way as in section 9d we may form the set of the $n(n-1)$ triplets of difference signs,

$$s(x_\alpha^{(1)} - x_\beta^{(1)}), \quad s(x_\alpha^{(2)} - x_\beta^{(2)}), \quad s(x_\alpha^{(3)} - x_\beta^{(3)}), \tag{9.34}$$

$$(\alpha \neq \beta;\ \alpha, \beta = 1, \cdots, n).$$

We shall assume that all $x^{(1)}$'s, $x^{(2)}$'s, and $x^{(3)}$'s are different. Then the triplets (9.34) contain only two different numbers, $+1$ and -1. Hence the regression functions of the three-variate population (9.34) are linear.

If t_{12}, t_{13}, and t_{23} are the difference sign correlations of $\{s(x_\alpha^{(1)} - x_\beta^{(1)}), s(x_\alpha^{(2)} - x_\beta^{(2)})\}$, $\{s(x_\alpha^{(1)} - x_\beta^{(1)}), s(x_\alpha^{(3)} - x_\beta^{(3)})\}$ and $\{s(x_\alpha^{(2)} - x_\beta^{(2)}), s(x_\alpha^{(3)} - x_\beta^{(3)})\}$ respectively, we have for the coefficient $t_{12.3}$ of partial correlation between $s(x_\alpha^{(1)} - x_\beta^{(1)})$ and $s(x_\alpha^{(2)} - x_\beta^{(2)})$ with respect to $s(x_\alpha^{(3)} - x_\beta^{(3)})$,

$$t_{12.3} = \frac{t_{12} - t_{13}\, t_{23}}{\sqrt{(1 - t_{13}^2)(1 - t_{23}^2)}}. \tag{9.35}$$

This measure of partial correlation has been suggested by Kendall [13] who gave an alternative definition of $t_{12.3}$.

If we have two independent three-dimensional random vectors $X_1 = (X_1^{(1)}, X_1^{(2)}, X_1^{(3)})$ and $X_2 = (X_2^{(1)}, X_2^{(2)}, X_2^{(3)})$ with the same continuous d.f. $F(x^{(1)}, x^{(2)}, x^{(3)})$, the distribution of the difference signs $s(X_1^{(i)} - X_2^{(i)})$, $(i = 1, 2, 3)$, has again linear regression functions, and we may define the partial difference sign correlation

$$\tau_{12.3} = \frac{\tau_{12} - \tau_{13}\tau_{23}}{\sqrt{(1 - \tau_{13}^2)(1 - \tau_{23}^2)}},$$

where τ_{ij} is the difference sign correlation of $X^{(i)}, X^{(j)}$.

If $t_{12.3}$ is a function of a random sample, and if $\tau_{13}^2 \neq 1$, $\tau_{23}^2 \neq 1$, the d.f. of $\sqrt{n}(t_{12.3} - \tau_{12.3})$ tends, by Theorem 7.5, so the normal d.f. with mean zero and variance

$$\sigma_{12.3}^2 = \frac{4}{(1 - \tau_{13}^2)(1 - \tau_{23}^2)} \Big\{ \zeta_1(\tau_{12}) + \frac{(\tau_{23} - \tau_{12}\tau_{13})^2}{(1 - \tau_{13}^2)^2} \zeta_1(\tau_{13}) + \frac{(\tau_{13} - \tau_{12}\tau_{23})^2}{(1 - \tau_{23}^2)^2} \zeta_1(\tau_{23}) - 2\frac{\tau_{23} - \tau_{12}\tau_{13}}{1 - \tau_{13}^2} \zeta_1(\tau_{12}, \tau_{13}) - 2\frac{\tau_{13} - \tau_{12}\tau_{23}}{1 - \tau_{23}^2} \zeta_1(\tau_{12}, \tau_{23}) + 2\frac{(\tau_{23} - \tau_{12}\tau_{13})(\tau_{13} - \tau_{12}\tau_{23})}{(1 - \tau_{13}^2)(1 - \tau_{23}^2)} \zeta_1(\tau_{13}, \tau_{23}) \Big\},$$

where

$$\zeta(\tau_{ij}) = E\{\Phi_1^2(X \mid \tau_{ij})\} - \tau_{ij}^2,$$

$$\zeta_1(\tau_{ij}, \tau_{gh}) = E\{\Phi_1(X \mid \tau_{ij})\Phi_1(X \mid \tau_{gh})\} - \tau_{ij}\tau_{gh},$$

and, for instance (cf. (9.13)),

$$\Phi_1(x \mid \tau_{12}) = 1 - 2F(x^{(1)}, \infty, \infty) - 2F(\infty, x^{(2)}, \infty) + 4F(x^{(1)}, x^{(2)}, \infty).$$

If $\tau_{13} = \tau_{23} = 0$, we have

$$\sigma_{12.3}^2 = 4\zeta_1(\tau_{12}),$$

and $\sqrt{n}(t_{12.3} - \tau_{12.3})$ has the same limiting distribution as $\sqrt{n}(t_{12} - \tau_{12})$. This is in particular the case when $X^{(1)}, X^{(2)}, X^{(3)}$ are independent.

REFERENCES

[1] H. Cramér, *Random Variables and Probability Distributions*, Cambridge Tracts in Math,, Cambridge, 1937.

[2] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946.

[3] H. E. Daniels, "The relation between measures of correlation in the universe of sample permutations," *Biometrika*, Vol. 33 (1944), pp. 129–135.

[4] H. E. Daniels and M. G. Kendall, "The significance of rank correlations where parental correlation exists," *Biometrika*, Vol. 34 (1947), pp. 197–208.

[5] G. B. Dantzig, "On a class of distributions that approach the normal distribution function," *Annals of Math. Stat.*, Vol. 10 (1939) pp. 247–253.

[6] F. ESSCHER, "On a method of determining correlation from the ranks of the variates," *Skandinavisk Aktuar. tids.*, Vol. 7 (1924), pp. 201–219.
[7] W. FELLER, "The fundamental limit theorems in probability," *Am. Math. Soc. Bull.*, Vol. 51 (1945), pp. 800–832.
[8] C. GINI, "Sulla misura della concentrazione e della variabilità dei caratteri," *Atti del R. Istituto Veneto di S.L.A.*, Vol. 73 (1913–14), Part 2.
[9] P. R. HALMOS, "The theory of unbiased estimation," *Annals of Math. Stat.*, Vol. 17 (1946), pp. 34–43.
[10] W. HÖFFDING, "On the distribution of the rank correlation coefficient τ, when the variates are not independent," *Biometrika*, Vol. 34 (1947), pp. 183–196.
[11] H. HOTELLING AND M. R. PABST, "Rank correlation and tests of significance involving no assumptions of normality," *Annals of Math. Stat.*, Vol. 7 (1936), pp. 29–43.
[12] M. G. KENDALL, "A new measure of rank correlation," *Biometrika*, Vol. 30 (1938), pp. 81–93.
[13] M. G. KENDALL, "Partial rank correlation," *Biometrika*, Vol. 32 (1942), pp. 277–283.
[14] M. G. KENDALL, S. F. H. KENDALL, AND B. BABINGTON SMITH, "The distribution of Spearman's coefficient of rank correlation in a universe in which all rankings occur an equal number of times," *Biometrika*, Vol. 30 (1939), pp. 251–273.
[15] J. W. LINDEBERG, "Über die Korrelation," VI *Skand. Matematikerkongres i København*, 1925, pp. 437–446.
[16] J. W. LINDEBERG, "Some remarks on the mean error of the percentage of correlation," *Nordic Statistical Journal*, Vol. 1 (1929), pp. 137–141.
[17] H. B. MANN, "Nonparametric tests against trend," *Econometrica*, Vol. 13 (1945), pp. 245–259.
[18] R. v. MISES, "On the asymptotic distribution of differentiable statistical functions," *Annals of Math. Stat.*, Vol. 18 (1947), pp. 309–348.
[19] U. S. NAIR, "The standard error of Gini's mean difference," *Biometrika*, Vol. 28 (1936), 428–436.
[20] K. PEARSON, "On further methods of determining correlation," *Drapers' Company Research Memoirs*, Biometric Series, IV, London, 1907.
[21] V. VOLTERRA, *Theory of Functionals*, Blackie, (authorized translation by Miss M. Long), London and Glasgow, 1931.
[22] G. U. YULE AND M. G. KENDALL, *An Introduction to the Theory of Statistics*, Griffin, 11th Edition, London, 1937.