THE LIMITING DISTRIBUTIONS OF CERTAIN STATISTICS[1]

By J. L. Doob

There have been many advances in the theory of probability in recent years, especially relating to its mathematical basis. Unfortunately, there appears to be no source readily available to the ordinary American statistician which sketches these results and shows their application to statistics. It is the purpose of this paper to define the basic concepts and state the basic theorems of probability, and then, as an application, to find the limiting distributions for large samples of a large class of statistics. One of these statistics is the tetrad difference, which has been of much concern to psychologists.

I

Let $F(x)$ be a monotone non-decreasing function, continuous on the left, defined at every point of the x-axis, and satisfying the conditions

$$\lim_{x\to-\infty} F(x) = 0, \qquad \lim_{x\to\infty} F(x) = 1. \tag{1}$$

Then the function $F(x)$ is said to be the distribution function of a chance variable $\mathbf{x}$, and $F(x)$ is said to be the probability that $\mathbf{x} < x$. The curve $y = F(x)$ is sometimes called the ogive in statistics. The chance variable $\mathbf{x}$ itself is merely the function x, taken in conjunction with the monotone function $F(x)$.

If $\int_{-\infty}^{\infty} x dF(x)$ exists as an absolutely convergent Stieltjes integral, the value of the integral is called the expectation of $\mathbf{x}$, and will be denoted by $E(\mathbf{x})$.

II

Let $F(x_1, \cdots, x_n)$ be a function defined over n-dimensional space, which is monotone, non-decreasing, continuous on the left in each coördinate if the others are held fast, and which satisfies the conditions

$$\lim_{x_j\to-\infty} F(x_1, \cdots, x_n) = 0, \quad j = 1, \cdots, n, \qquad \lim_{x_1,\cdots,x_n\to\infty} F(x_1, \cdots, x_n) = 1 \tag{2}$$

where in the last limit, $x_1, \cdots, x_n$ become infinite together. Then $F(x_1, \cdots, x_n)$ is said to be the distribution function of a set of chance variables $\mathbf{x}_1, \cdots, \mathbf{x}_n$, and $F(x_1, \cdots, x_n)$ is said to be the probability that all the inequalities $\mathbf{x}_j < x_j$, $(j = 1, \cdots, n)$, hold simultaneously. It can be shown that the function $F_j(x) = \lim_{\xi_1,\cdots,\xi_{n-1}\to\infty} (\xi_1, \cdots \xi_{j-1}, x, \xi_j, \cdots \xi_{n-1})$ is of the type discussed in §I. The

[1] Research under a grant-in-aid from the Carnegie Corporation.

function $F_j(x)$ is called the distribution function of $\mathbf{x}_j$. The chance variables $\mathbf{x}_1, \cdots, \mathbf{x}_n$ are called independent if $F(x_1, \cdots, x_n) = \prod_{j=1}^{n} F_j(x_j)$. The chance variables $\mathbf{x}_1, \cdots, \mathbf{x}_n$ are merely the functions $x_1, \cdots, x_n$ defined over n-dimensional space, taken in conjunction with the function $F(x_1, \cdots, x_n)$.

If $a_1, \cdots, a_n$ are any real numbers, the number $F(a_1, \cdots, a_n)$, the probability that $\mathbf{x}_j < a_j$, $j = 1, \cdots, n$, is also called the probability that a sample $(x_1, \cdots, x_n)$ shall be in the region of n-dimensional space determined by $x_j < a_j$, $j = 1, \cdots, n$. Thus regions of this special type have probabilities attached to them. Using the usual additivity rules, probabilities can be attached to more general regions, and in fact probability can be defined on a collection C of regions including all open sets, closed sets and all sets which can be obtained from them by repeatedly taking sums, products, and complements. (Such point sets are called Borel measurable). The resulting function of point sets is non-negative and completely additive.[2]

If $f(x_1, \cdots, x_n)$ is any function of $x_1, \cdots, x_n$ let E_x be the set of points $(x_1, \cdots, x_n)$ where $f < x$. Suppose that E_x is in the collection C for all values of x, and let $F(x)$ be the probability attached to the set E_x. Then it is readily seen that $F(x)$ has the properties discussed in §I and is therefore the distribution function of a new chance variable $\mathbf{x}$, which will be denoted by $f(\mathbf{x}_1, \cdots, \mathbf{x}_n)$. The chance variable $f(\mathbf{x}_1, \cdots, \mathbf{x}_n)$ is merely the function $f(x_1, \cdots, x_n)$ taken in conjunction with the distribution function $F(x_1, \cdots, x_n)$. (An example is $f(x_1, \cdots, x_n) = x_1 + \cdots + x_n$, determining the chance variable $\mathbf{x}_1 + \cdots + \mathbf{x}_n$.) Suppose that $E(\mathbf{x})$ exists,

$$E(\mathbf{x}) = \int_{-\infty}^{\infty} x\,dF(x). \tag{3}$$

Then it can be shown that the n-dimensional (Lebesgue)-Stieltjes integral

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \cdots, x_n)\,dF(x_1, \cdots, x_n) \tag{4}$$

exists and has the value $E(\mathbf{x})$. Conversely the existence of the integral (4) implies that of (3).

If there is a Lebesgue-integrable function $\varphi(x_1, \cdots, x_n)$ such that

$$F(x_1, \cdots, x_n) = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} \varphi(x_1, \cdots, x_n)\,dx_1 \cdots dx_n, \tag{5}$$

[2] That is, if $p(E)$ is the value of the set function on the set E, and if $E_1, E_2, \cdots$ are point sets with no common points, and which are in C, $p\left(\sum_{m=1}^{\infty} E_m\right) = \sum_{m=1}^{\infty} p(E_m)$.

the function φ is said to be the density function of the distribution. In this case (4) becomes

$$(4') \qquad \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \cdots, x_n)\, \varphi(x_1, \cdots, x_n)\, dx_1 \cdots dx_n.$$

The probability attached to a point set E in the collection C is the integral (4) (or (4′) if there is a density function), where $f = 1$ over E and $f = 0$ elsewhere.

III

Let $\mathbf{x}, \mathbf{x}_1, \mathbf{x}_2, \cdots$ be a sequence of chance variables. We suppose that for every integer n, $\mathbf{x}, \mathbf{x}_n$ determine a bivariate distribution. Then it is readily seen from §II that there is a chance variable $|\mathbf{x}_n - \mathbf{x}|$ and therefore that $P\{|\mathbf{x}_n - \mathbf{x}| \leqq \lambda\}$[3] is defined for every number λ. If

$$(6) \qquad \lim_{n \to \infty} P\{|\mathbf{x}_n - \mathbf{x}| \leqq \lambda\} = 1$$

for every positive number λ, the sequence $\mathbf{x}_n$ is said to converge stochastically, or to converge in probability, to $\mathbf{x}$. If α is a constant, $P\{|\mathbf{x}_n - \alpha| \leqq \lambda\}$ is also defined for every number λ, and there is a corresponding definition of stochastic convergence to α. The usual theorems about limits hold: if $\mathbf{x}_n$, $\mathbf{y}_n$ converge stochastically to $\mathbf{x}$, $\mathbf{y}$, $\mathbf{x}_n + \mathbf{y}_n$ converges stochastically to $\mathbf{x} + \mathbf{y}$, etc.

An example of stochastic convergence is given by the law of large numbers. Let $\mathbf{x}$ be a chance variable with distribution function $F(x)$ and suppose that $E(\mathbf{x})$, $E(\mathbf{x}^2)$ exist, i.e. that

$$\int_{-\infty}^{\infty} x\,dF(x), \qquad \int_{-\infty}^{\infty} x^2 dF(x)$$

are absolutely convergent integrals. Let $\mathbf{x}_1, \cdots, \mathbf{x}_n$ be chance variables whose n-variate distribution function is $\prod_{j=1}^{n} F(x_j)$: we are thus supposing that the variables all have the same distribution and form an independent set. Then $\frac{1}{n}\sum_{j=1}^{n} \mathbf{x}_j$ is a new chance variable, and Tchebycheff's inequality furnishes an immediate proof that $\frac{1}{n}\sum_{j=1}^{n} \mathbf{x}_j$ converges stochastically to $E(\mathbf{x})$.[4]

---

[3] Throughout this paper, if γ represents a set of conditions on chance variables, $P\{\gamma\}$ will denote the probability that those conditions are satisfied.

[4] If $\bar{\mathbf{x}}_n = \frac{1}{n}\sum_{j=1}^{n} \mathbf{x}_j$, $E(\bar{\mathbf{x}}_n) = E(\mathbf{x})$, $E(\bar{\mathbf{x}}_n^2) = \frac{1}{n}E(\mathbf{x}^2)$. Then if λ is any positive number $P\{|\bar{\mathbf{x}}_n - E(\mathbf{x})| > \lambda\} \leqq \frac{E\{[\mathbf{x} - E(\mathbf{x})]^2\}}{n\lambda^2}$ which implies (6).

There is also another kind of convergence, called convergence with probability 1. The sequence $\{\mathbf{x}_n\}$ converges with probability 1 to $\mathbf{x}$ if

$$\lim_{n\to\infty} P\{|\mathbf{x}_n - \mathbf{x}| \leqq \lambda, |\mathbf{x}_{n+1} - \mathbf{x}| \leqq \lambda, \cdots, |\mathbf{x}_{n+p} - \mathbf{x}| \leqq \lambda\} = 1 \tag{7}$$

for every value of $p \geqq 0$, uniformly in $p \geqq 0$ for every positive number λ. If $p = 0$ in (7), (7) becomes (6), so that convergence with probability 1 implies stochastic convergence. Although the converse is not true, if $\{\mathbf{x}_n\}$ is a sequence of chance variables converging stochastically to $\mathbf{x}$, there is a subsequence of $\{\mathbf{x}_n\}$ which converges with probability 1 to $\mathbf{x}$.[5] The usual limit theorems hold here also: if $\mathbf{x}_n$, $\mathbf{y}_n$ converge with probability 1 to $\mathbf{x}$, $\mathbf{y}$, $\mathbf{x}_n + \mathbf{y}_n$ converges with probability 1 to $\mathbf{x} + \mathbf{y}$, etc.

An example of convergence with probability 1 is the following. If in the previous example the hypothesis that $E(\mathbf{x}^2)$ exists is removed, so that only the weaker hypothesis of the existence of $E(\mathbf{x})$ is supposed, the Tchebycheff inequality can no longer be applied, but a different method shows that $\frac{1}{n}\sum_{j=1}^{n} \mathbf{x}_j$ converges with probability 1 (and therefore stochastically) to $E(x)$.[6] This result is known as the strong law of large numbers.

IV

Let $\mathbf{x}, \mathbf{x}_1, \mathbf{x}_2, \cdots$ be a sequence of chance variables with distribution functions $F(x), F_1(x), F_2(x), \cdots$ respectively. Then if $\lim_{n\to\infty} F_n(x) = F(x)$ for every value of x, the distribution of $\mathbf{x}_n$ is said to converge to a limiting distribution with distribution function $F(x)$.

As an example, consider the Laplace-Liapounoff theorem. Let $\mathbf{x}_1, \mathbf{x}_2, \cdots$ be a sequence of independent chance variables (i.e. any finite number of them form an independent set) with the same distribution functions, and let $E(\mathbf{x}_n)$, $E(\mathbf{x}_n^2)$ exist. We suppose that $\sigma^2 = E\{[\mathbf{x}_n - E(\mathbf{x}_n)]^2\} > 0$ so that the distribution of $\mathbf{x}_n$ is not merely confined to one point. Then the distribution of

$$n^{-\frac{1}{2}} \sum_{j=1}^{n} [\mathbf{x}_j - E(\mathbf{x}_j)] \tag{8}$$

---

[5] The theories of probability and of measure are fundamentally identical. Chance variables correspond to measurable functions. Stochastic convergence corresponds to convergence in measure, and convergence with probability 1 corresponds to convergence almost everywhere. The relation between these two types of convergence is discussed (in the terminology of the measure theory) in E. W. Hobson, The Theory of Functions of a Real Variable, second edition Vol. 2, pp. 239–244.

[6] Cf. for instance J. L. Doob, Transactions of the American Mathematical Society, Vol. 36 (1934), pp. 764–765.

converges to a limiting distribution with distribution function[7]

$$(9) \qquad \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{x^2}{2\sigma^2}}\,dx\,.$$

The convergence of a sequence of n-variate distributions is defined as the convergence of the distribution functions just as above for $n = 1$. Suppose that $(\mathbf{x}_{11}, \cdots, \mathbf{x}_{n1})$, $(\mathbf{x}_{12}, \cdots, \mathbf{x}_{n2})$, $\cdots$ are independent sets of chance variables (i.e. the distribution function of any finite number of sets is the product of the distribution functions of the sets) with the same distribution functions. We suppose that $E(\mathbf{x}_{j1})$, $E(\mathbf{x}_{j1}^2)$ exist, $j = 1, \cdots, n$ and that $\sigma_j^2 = E\{[\mathbf{x}_{j1} - E(\mathbf{x}_{j1})]^2\} > 0$. Then if $\bar{\mathbf{x}}_{jm} = m^{-\frac{1}{2}}\sum_{i=1}^{m}[\mathbf{x}_{ji} - E(\mathbf{x}_{ji})]$, the n-variate distribution of $\bar{\mathbf{x}}_{1m}, \cdots, \bar{\mathbf{x}}_{nm}$ converges to the normal distribution[8] about zero means with variances $\sigma_1^2, \cdots, \sigma_n^2$ and correlation coefficients $\{\rho_{ij}\}$ where $\sigma_i\sigma_j\rho_{ij} = E\{[\mathbf{x}_{i1} - E(\mathbf{x}_{i1})][\mathbf{x}_{j1} - E(\mathbf{x}_{j1})]\}$.

Three lemmas will be needed below in applying these concepts.

LEMMA 1. *If $\{\mathbf{x}_n\}$ is a sequence of chance variables whose distributions approach a limiting distribution and if $\{\mathbf{y}_n\}$ is a sequence of chance variables converging stochastically to 0, the sequence $\{\mathbf{x}_n\mathbf{y}_n\}$ converges stochastically to 0.*

For if $F(x)$ is the distribution function of the limiting distribution, and if λ, μ are any positive numbers,

$$(10) \qquad \begin{aligned} P\{|\mathbf{x}_n\mathbf{y}_n| < \lambda\} &\geqq P\{|\mathbf{x}_n\mathbf{y}_n| < \lambda, \ |\mathbf{y}_n| \leqq \mu\} \geqq P\{|\mathbf{x}_n| < \lambda/\mu, \ |\mathbf{y}_n| \leqq \mu\} \\ &\geqq P\{|\mathbf{y}_n| \leqq \mu\} - P\{|\mathbf{x}_n| \geqq \lambda/\mu\} = -P\{|\mathbf{y}_n| > \mu\} + P\{|\mathbf{x}_n| < \lambda/\mu\} \\ &\geqq -P\{|y_n| > \mu\} + P\{\mathbf{x}_n < \lambda/\mu\} - P\{\mathbf{x}_n < -\lambda/2\mu\}\,. \end{aligned}$$

Then, letting n become infinite,

$$(11) \qquad \liminf_{n\to\infty} P\{|\mathbf{x}_n\mathbf{y}_n| < \lambda\} \geqq F(\lambda/\mu) - F(-\lambda/2\mu)\,.$$ [9]

Letting μ approach 0, $F(\lambda/\mu)$ approaches 1, $F(-\lambda/2\mu)$ approaches 0, and the right hand side becomes 1, as was to be proved.

LEMMA 2. *Let $\{\mathbf{x}_n\}$, $\{\mathbf{y}_n\}$, $\{\mathbf{z}_n\}$ be sequences of chance variables such that the distribution of x_n approaches a limiting distribution with continuous distribution function $F(x)$ and such that the sequences $\{\mathbf{y}_n\}$, $\{\mathbf{z}_n\}$ converge stochastically to 0, 1 respectively. Then the distributions of $\{\mathbf{x}_n/\mathbf{z}_n\}$[10] and of $\mathbf{x}_n + \mathbf{y}_n$ approach limiting distributions with the same distribution function $F(x)$.*

---

[7] A. Khintchine, Ergebnisse der Mathematik, Vol. 2, No. 4: Asymptotische Gesetze der Wahrscheinlichkeitsrechnung, pp. 1–8.

[8] Ibid. pp. 11–16.

[9] If $\{a_n\}$ is a sequence of real numbers $\limsup_{n\to\infty} a_n$ is defined as $\lim_{n\to\infty}$ {least upper bound $a_n, a_{n+1}, \cdots$ }, and $\liminf_{n\to\infty} a_n$ is defined as $-\limsup_{n\to\infty}(-a_n)$. A necessary and sufficient condition that the sequence $\{a_n\}$ converge to a limit a is that $\liminf_{n\to\infty} a_n = \limsup_{n\to\infty} a_n = a$.

[10] Since $\mathbf{z}_n$ converges stochastically to 1, the probability that $\mathbf{z}_n = 0$ approaches 0. The theorem is independent of the way $\mathbf{x}_n/\mathbf{z}_n$ is defined when $\mathbf{z}_n = 0$.

Since $\frac{\mathbf{x}_n}{\mathbf{z}_n} = \mathbf{x}_n + \mathbf{x}_n \frac{1 - \mathbf{z}_n}{\mathbf{z}_n}$ (neglecting the possibility that $\mathbf{z}_n$ may vanish), where the last term converges stochastically to 0 by Lemma 1, it is sufficient to prove the second part of the theorem. If $\epsilon > 0$, and if x is an arbitrary number,

$$(12)\quad P\{\mathbf{x}_n + \mathbf{y}_n < x\} = P\{\mathbf{x}_n + \mathbf{y}_n < x, |\mathbf{y}_n| \leqq \epsilon\} + P\{\mathbf{x}_n + \mathbf{y}_n < x, |\mathbf{y}_n| > \epsilon\}.$$

Since the sequence $\{\mathbf{y}_n\}$ converges stochastically to 0,

$$(13)\qquad \lim_{n\to\infty} P\{\mathbf{x}_n + \mathbf{y}_n < x, |y_n| > \epsilon\} \leqq \lim_{n\to\infty} P\{|\mathbf{y}_n| > \epsilon\} = 0$$

so that in the limit the second term in (12) can be neglected. Moreover

$$(14)\qquad P\{\mathbf{x}_n + \mathbf{y}_n < x, |\mathbf{y}_n| \leqq \epsilon\} \leqq P\{\mathbf{x}_n < x + \epsilon\}.$$

If we let n become infinite and then let ϵ approach 0, (14) becomes

$$(15)\qquad \limsup_{n\to\infty} P\{\mathbf{x}_n + \mathbf{y}_n < x\} \leqq F(x).$$

A similar argument shows that

$$(16)\qquad \liminf_{n\to\infty} P\{\mathbf{x}_n + \mathbf{y}_n < x\} \geqq F(x),$$

and (15), (16) taken together imply that

$$(17)\qquad \lim_{n\to\infty} P\{\mathbf{x}_n + \mathbf{y}_n < x\} = F(x),$$

as was to be proved.

LEMMA 3. *If $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4$ are chance variables whose distribution has density function*

$$\frac{1}{(2\pi)^2} e^{-\frac{1}{2}(x_1^2+x_2^2+x_3^2+x_4^2)}$$

the distribution of $\mathbf{z} = \mathbf{x}_1\mathbf{x}_2 - \mathbf{x}_3\mathbf{x}_4$ has density function $\frac{1}{2}e^{-|x|}$.

The distribution of $\mathbf{u} = \mathbf{x}_1\mathbf{x}_2$ and that of $\mathbf{v} = -\mathbf{x}_3\mathbf{x}_4$ have the same density function:

$$(18)\qquad \frac{1}{\pi}\int_0^\infty e^{-\frac{x^2}{2t^2}-\frac{t^2}{2}}\frac{dt}{t}.$$

Hence the distribution of $\mathbf{z}$ has density function

$$(19)\qquad \frac{1}{\pi^2}\int_{-\infty}^\infty\int_0^\infty\int_0^\infty e^{-\frac{(x-\lambda)^2}{2t^2}-\frac{t^2}{2}-\frac{\lambda^2}{2\tau^2}-\frac{\tau^2}{2}}\,d\lambda\,\frac{dt}{t}\,\frac{d\tau}{\tau}.$$

If we change to polar coördinates: $t = r\cos\theta$, $\tau = r\sin\theta$, and integrate out λ, we obtain

$$\frac{1}{\pi}\int_0^\infty\int_0^{\pi/2} e^{-\frac{x^2}{2r^2}-\frac{r^2}{2}}\,dr\,d\theta = \frac{1}{2}e^{-|x|}.$$

V

THEOREM 1. *Let* $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4$ *determine a 4-variate distribution with distribution function* $F(x_1, x_2, x_3, x_4)$. *Suppose that* $E(\mathbf{x}_i)$, $E(\mathbf{x}_i^2)$, $E(\mathbf{x}_i^2\mathbf{x}_j^2)$ *exist,* $i, j = 1, \cdots, 4$, *and suppose that* $E(\mathbf{x}_i) = 0$, $E(\mathbf{x}_i^2) = 1$,[11] $i, j = 1, 2, 3, 4$. *Let* $\mathbf{x}_{1j}, \mathbf{x}_{2j}, \mathbf{x}_{3j}, \mathbf{x}_{4j}$ *have the same 4-variate distribution as* $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4$, $j = 1, \cdots, n$, *and let the* $4n$*-variate distribution function of* $\{\mathbf{x}_{ij}\}$ *be* $\prod_{j=1}^{n} F(x_{1j}, x_{2j}, x_{3j}, x_{4j})$. *We shall use the following notation (which suppresses the dependence on* n*):*

$$(20)\qquad \xi_i = \frac{1}{n}\sum_{k=1}^{n}\mathbf{x}_{ik}, \qquad \mathbf{s}_{ij} = \frac{1}{n}\sum_{k=1}^{n}\mathbf{x}_{ik}\mathbf{x}_{jk}, \qquad \rho_{ij} = E(\mathbf{x}_i\mathbf{x}_j).$$

Let φ *be a function of* ξ_i, s_{ij}, *defined in a neighborhood* N *of* P: $\xi_i = 0$, $s_{ij} = \rho_{ij}$, *which, together with its second partial derivatives is continuous in* N. *Define* $\sigma \geqq 0$ *by*

$$(21)\qquad \sigma^2 = E\left\{\left[\sum_{i=1}^{4}\frac{\partial\varphi}{\partial\xi_i}\mathbf{x}_i - \sum_{i,j=1}^{4}\frac{\partial\varphi}{\partial s_{ij}}(\rho_{ij} - \mathbf{x}_i\mathbf{x}_j)\right]^2\right\},$$

where the partial derivatives are evaluated at P. *Then if* $\sigma > 0$, *the distribution of* $\sqrt{n}\,[\varphi - \varphi(P)]$ *(where* φ *has the arguments* ξ_i, $\mathbf{s}_{ij}$*) converges to a limiting distribution which is normal with mean* 0 *and variance* σ^2.

To prove this theorem we expand φ in the neighborhood of P, obtaining

$$(22)\qquad \sqrt{n}\,[\varphi - \varphi(P)] = \sum_{i=1}^{4}\frac{\partial\varphi}{\partial\xi_i}\sqrt{n}\,\xi_i - \sum_{i,j=1}^{4}\frac{\partial\varphi}{\partial s_{ij}}\sqrt{n}\,(\rho_{ij} - \mathbf{s}_{ij}) + \mathbf{R}_n$$

where the partial derivatives are evaluated at P, and where $\mathbf{R}_n$ consists of a linear combination of $\sqrt{n}\,\xi_i\xi_j$, $\sqrt{n}\,\xi_i(\rho_{jk} - \mathbf{s}_{jk})$, $\sqrt{n}\,(\rho_{ij} - \mathbf{s}_{ij})(\rho_{kl} - \mathbf{s}_{kl})$, with coefficients which are uniformly bounded as long as ξ_i, $\mathbf{s}_{ij}$ are in the neighborhood N. Now

$$(23)\qquad \lim_{n\to\infty}\xi_i = 0 \qquad\qquad \lim_{n\to\infty}\mathbf{s}_{ij} = \rho_{ij}$$

with probability 1, by the law of large numbers, and as n becomes infinite the distributions of $\sqrt{n}\,\xi_i$, $\sqrt{n}\,(\rho_{ij} - \mathbf{s}_{ij})$ converge to limiting distributions, by the

[11] The hypothesis that $E(\mathbf{x}_i) = 0$ involves no real restriction, since the general case can be reduced to this one by substituting $\mathbf{x}_i - E(\mathbf{x}_i)$ for $\mathbf{x}_i$. The hypothesis that $E(\mathbf{x}_i^2) = 1$ can be met by substituting $\mathbf{x}_i[E(\mathbf{x}_i^2)]^{-\frac{1}{2}}$ whenever $E(\mathbf{x}_i^2) > 0$, which will always be true unless $\mathbf{x}_i = 0$ with probability 1.

Laplace-Liapounoff theorem. Then by Lemma 1, the terms of $\mathbf{R}_n$ converge stochastically to 0. The other terms of $\sqrt{n}[\varphi - \varphi(P)]$ are sums to which the Laplace-Liapounoff theorem can be applied, giving the desired conclusion.

As an example of the application of this theorem, we suppose that φ is a correlation coefficient:

$$\varphi = \frac{s_{12}}{(s_{11}\, s_{22})^{\frac{1}{2}}}, \qquad \varphi(P) = \rho_{12}. \tag{24}$$

Here σ^2 is $E\{[x_1x_2 - \frac{1}{2}\rho_{12}(x_1^2 + x_2^2)]^2\}$, (which reduces to the familiar result $1 - \rho_{12}^2$ when the bivariate distribution of x_1, x_2 is normal) and $\sigma = 0$ only when, with probability 1,

$$2\, x_1x_2 = \rho_{12}(x_1^2 + x_2^2). \tag{25}$$

As a second example we suppose that φ is a tetrad difference:

$$\varphi = \frac{s_{13}\, s_{24} - s_{14}\, s_{23}}{(s_{11}\, s_{22}\, s_{33}\, s_{44})^{\frac{1}{2}}}, \qquad \varphi(P) = \rho_{13}\rho_{24} - \rho_{14}\rho_{23}. \tag{26}$$

Here σ^2 becomes

$$\sigma^2 = E\left\{\left[\rho_{24}x_1x_3 + \rho_{13}x_2x_4 - \rho_{14}x_2x_3 - \rho_{23}x_1x_4 - \frac{\varphi(P)}{2}\sum_{j=1}^{4} x_j^2\right]^2\right\} \tag{27}$$

and $\sigma = 0$ only when the quantity in the brackets vanishes with probability 1.

If in either of the two above cases $s_{ij} - \xi_i\xi_j$ is substituted for s_{ij} (i.e. if the deviations from the sample mean, not those from the true mean, are used), the result is unaltered. This is true in general, since $\frac{\partial\varphi}{\partial\xi_i}, \frac{\partial\varphi}{\partial s_{ij}}$ are unaltered at P by this substitution.

There is a well-known δ-method used in statistics to find limiting variances of statistics of the type covered by Theorem 1,[12] and Theorem 1 shows an interpretation which can be given to the results obtained by this method.

We now investigate the necessary modification of Theorem 1 if $\sigma = 0$, i.e. if

$$\sum_{i=1}^{4} \frac{\partial\varphi}{\partial\xi_i} x_i - \sum_{i,j=1}^{4} \frac{\partial\varphi}{\partial s_{ij}}(\rho_{ij} - x_ix_j) = 0 \tag{28}$$

with probability 1. If we assume that φ has continuous third partial derivatives in the neighborhood N, we find that

---

[12] Examples of the use of this method can be found in T. L. Kelley, Crossroads in The Mind of Man, Stanford University (1928), pp. 49–50, and in an article by S. Wright, Annals of Mathematical Statistics, Vol. 5 (1934), p. 211.

$$n[\varphi - \varphi(P)] = \frac{n}{2} \sum \frac{\partial^2 \varphi}{\partial \xi_i \xi_j} \xi_i \xi_j + \frac{n}{2} \sum_{i,j,k} \frac{\partial^2 \varphi}{\partial \xi_i \partial s_{jk}} \xi_i (s_{jk} - \rho_{jk}) \tag{29}$$

$$+ \frac{n}{2} \sum_{i,j,k,l} \frac{\partial^2 \varphi}{\partial s_{ij} \partial s_{kl}} (s_{ij} - \rho_{ij})(s_{kl} - \rho_{kl}) + R'_n$$

where R'_n converges stochastically to 0. The second degree terms constitute a quadratic form in $\{\xi_i, s_{jk} - \rho_{jk}\}$. Now the multivariate distribution of $\{\sqrt{n}\xi_i, \sqrt{n}(s_{jk} - \rho_{jk})\}$, by the Laplace-Liapounoff theorem, converges to a normal distribution whose variances and correlation coefficients are those of $x_i, x_j x_k$. The distribution of $n[\varphi - \varphi(P)]$ thus converges to the distribution of the quadratic form

$$\frac{n}{2} \sum_{i,j} \frac{\partial^2 \varphi}{\partial \xi_i \partial \xi_j} \alpha_i \alpha_j + \frac{n}{2} \sum_{i,j,k} \frac{\partial^2 \varphi}{\partial \xi_i \partial s_{jk}} \alpha_i \beta_{jk} + \frac{n}{2} \sum_{i,j,k,l} \frac{\partial^2 \varphi}{\partial s_{ij} \partial s_{kl}} \beta_{ij} \beta_{kl}, \tag{30}$$

where $\{\alpha_i, \beta_{jk}\}$ have the multivariate distribution just described, unless the quadratic form vanishes identically. This reasoning can be continued, the general result being that there is some power ν of n, if φ is sufficiently regular, such that the distribution of $n^\nu[\varphi - \varphi(P)]$ converges to a limiting distribution.

When $\sigma = 0$ in the second example, unless the distribution of x_1, x_2, x_3, x_4 is confined with probability 1 to a 4-dimensional quadric, $\rho_{13} = \rho_{14} = \rho_{23} = \rho_{24} = 0$. Equation (29) becomes

$$n[\varphi - \varphi(P)] = s_{13} s_{24} - s_{14} s_{23} + R'_n. \tag{29'}$$

Now if x_1, x_2 are transformed by a linear homogeneous transformation with determinant Δ, it is readily seen that $s_{13}s_{24} - s_{14}s_{23}$ is multiplied by Δ. The same is true of x_3, x_4. If x_1, x_2 are transformed into x'_1, x'_2 so that $E(x'^2_j) = 1$, $E(x'_1 x'_2) = 0$, the determinant of the transformation is $\pm(1 - \rho_{12}^2)^{-\frac{1}{2}}$. Then transforming each pair $(x_1, x_2), (x_3, x_4)$ in this way into $(x'_1, x'_2), (x'_3, x'_4)$, the variables x'_1, x'_2, x'_3, x'_4 are uncorrelated. If $s'_{ij} = \frac{1}{n} \sum_{k=1}^{n} x'_{ik} x'_{jk}$,

$$s'_{13} s'_{24} - s'_{14} s'_{23} = \frac{s_{13} s_{24} - s_{14} s_{23}}{\pm(1 - \rho_{12}^2)^{\frac{1}{2}}(1 - \rho_{34}^2)^{\frac{1}{2}}}. \tag{31}$$

The limiting distribution of $s'_{13}s'_{24} - s'_{14}s'_{23}$ is the distribution of $\beta'_{13}\beta'_{24} - \beta'_{14}\beta'_{23}$ where these four chance variables are normally distributed, $E(\beta'_{13}) = E(\beta'_{24}) = E(\beta'_{14}) = E(\beta'_{23}) = 0$, $E(\beta'_{ij}) = E(x'_i x'_j)$, $E(\beta'_{ij}\beta'_{kl}) = E(x'_i x'_j x'_k x'_l)$. Now if x_1, x_2, x_3, x_4 are normally distributed—the most important case for statistical purposes—x'_1, x'_2, x'_3, x'_4 will also be distributed normally, and the vanishing of the correlation coefficients means that the chance variables are independent. If this is true

$$E(\beta'^2_{ij}) = 1, \qquad E(\beta_{ij}\beta_{kl}) = 0, \qquad (\beta_{ij} \neq \beta_{kl}). \tag{32}$$

Evidently, however, x_1', x_2', x_3', x_4' do not have to be independent to make these equations valid. It is more than sufficient if the pairs (x_1, x_2), (x_3, x_4) and therefore the pairs (x_1', x_2'), (x_3', x_4') are independent. If (32) is true, the β's are independent, each one being normally distributed with mean 0 and variance 1. Summarizing these results, and using Lemma 3: *if φ is the tetrad difference and if $\rho_{13} = \rho_{14} = \rho_{23} = \rho_{24} = 0$, the distribution of $n[\varphi - \varphi(P)]$ converges to a limiting distribution. If in addition the distribution of x_1, x_2, x_3, x_4 is normal, or if the pairs (x_1, x_2) (x_3, x_4) are independent, this limiting distribution has density function*

$$\frac{c}{2} e^{-c|x|}$$

where $c = (1 - \rho_{12}^2)^{-\frac{1}{2}} (1 - \rho_{34}^2)^{-\frac{1}{2}}$.

Wilks has investigated the case where x_1, x_2, x_3, x_4 are normally and independently distributed, and in this case found the exact variance of the tetrad difference as a function of n.[13]

COLUMBIA UNIVERSITY.

---

[13] Proceedings of the National Academy of Sciences, Vol. 18, (1932), pp. 562–565.