LIMITING DISTRIBUTIONS OF QUADRATIC AND BILINEAR FORMS[1, 2]

By William G. Madow

1. Introduction. In a previous paper [15], several generalizations of the theorem of Fisher, [6, p. 97] and Cochran, [2, p. 178] on the joint distribution of quadratic forms in normally and independently distributed random variables were derived. The chief purpose of this paper is a demonstration that the Fisher-Cochran theorem and its generalizations are valid in the limit under conditions completely analogous to those under which the Laplace-Liapounoff theorem holds. Applications to the analysis of variance, periodogram analysis and multivariate analysis are discussed.

Our general procedure will be to find algebraic conditions on the matrices of quadratic and bilinear forms which enable us to assert that the limiting distributions of these forms are those which they would have had if the variables, the squares or products of which appear in their canonical forms, had been normally and independently distributed.[3] One thing which makes this possible is the fact that many frequently used quadratic and bilinear forms have the same rank no matter what may be the number of variables of which they are functions. For example, the rank of the square of the arithmetic mean, $\bar{x}_n$, where

$$\bar{x}_n = \frac{1}{n}(x_1 + \cdots + x_n),$$

is one for all values of n. In this case the quadratic form,

$$\frac{1}{n^2}\sum_{\mu,\nu=1}^{n} x_\mu x_\nu,$$

is a function of the n variables $x_1, x_2, \cdots, x_n$.

In paragraph 2 we state the vector form of the Laplace-Liapounoff theorem and several corollaries. The joint limiting distributions of quadratic and bilinear forms are derived in paragraph 3. The final paragraph is devoted to a statement of a few applications of the theorems.

---

[1] Much of this research was done under a grant-in-aid from the Carnegie Corporation of New York.

[2] The material contained in this paper was presented in part to the American Statistical Association, December 28, 1937, and in part to the Institute of Mathematical Statistics, December 27, 1938.

[3] We shall be chiefly concerned with conditions under which the limiting distributions are not themselves normal. If the limiting distributions are normal, then generally under the conditions we state, the Laplace-Liapounoff theorem will have been directly applicable.

2. The Laplace-Liapounoff theorem.[4] We shall first state some definitions and terminology which will be used throughout the paper.

If used as subscripts or superscripts, or as indices of summation or multiplication, the letters i, j will take on all integral values from 1 through p, the letters μ, ν will take on all integral values from 1 through n, the letters γ, δ will take on all integral values from 1 through m, the letter α will take on all integral values from 1 through k, and the letter β will take on all integral values from 1 through k–1, unless explicit statement to the contrary is made.

The totality of all sets of ν real numbers will be denoted by R^ν. Thus R^ν is the combinatory product of the spaces $R^1, R^1, \cdots, R^1$, (ν times).

If $x_1, \cdots, x_n$ are random variables, and if Δ is a proposition concerning $x_1, \cdots, x_n$, then by $P\{\Delta\}$ we shall mean "the probability that Δ." The distribution function of the random variables $x_1, \cdots, x_n$ will be denoted by $F(x_1, \cdots, x_n)$, i.e.

$$F(x_1^0, \cdots, x_n^0) = P\{x_1 < x_1^0, \cdots, x_n < x_n^0\}$$

for all sets of n real numbers. Thus F will have an operational meaning in this paper.

If $\Delta(x_1, \cdots, x_n)$ is a function of $x_1, \cdots, x_n$ defined on R^n and measurable[5] with respect to $F(x_1, \cdots, x_n)$, then $E\{\Delta(x_1, \cdots, x_n)\}$ will be defined by the equation,

$$E\{\Delta(x_1, \cdots, x_n)\} = \int_{R^n} \Delta(x_1, \cdots, x_n)\, dF(x_1, \cdots, x_n),$$

where the integral is a Lebesgue-Stieltjes or Radon integral. Hence $|\Delta(x_1, \cdots, x_n)|$ is assumed to be integrable with respect to $F(x_1, \cdots, x_n)$.

If $\Omega(y_1, \cdots, y_p)$ is a single valued measurable function of $y_1, \cdots, y_p$ on R^p, and if y_i is a real single valued Borel measurable[6] function of $x_1, \cdots, x_n$ on R^n, then upon substituting for $y_1, \cdots, y_p$ it is seen that $\Omega(y_1, \cdots, y_p)$

[4] Although the theorems will be stated in terms of probability distributions, Borel measurability, and Lebesgue-Stieltjes integrability, it may simplify the reading if the words "probability distributions" are replaced by probability densities or statistical distributions, "Borel measurability" are replaced by continuity, and "Lebesgue-Stieltjes integrability" are replaced by Riemann integrability.

[5] A function $\Delta(x_1, \ldots, x_n)$ defined on R^n is said to be measurable with respect to a distribution function $F(x_1, \ldots, x_n)$ if the set $E(t)$ of all $x_1, \ldots, x_n$ such that $\Delta(x_1, \ldots, x_n) < t$ is such that $\int_{E(t)} dF(x_1, \ldots, x_n)$ is defined for all t.

[6] All subsets of R^n which may be formed from the totality of intervals of R^n by repeated summations or multiplications of not more than a denumerable number of intervals of R^n, and R^n itself, constitute the totality of Borel sets of R^n. The function $y(x_1, \ldots, x_n)$, defined on R^n, is a Borel measurable function of $x_1, \ldots, x_n$ on R^n if the set of values of $x_1, \ldots, x_n$ such that $y(x_1, \ldots, x_n) < t$ is a Borel set for all t. The class of continuous functions is contained in the class of Borel measurable functions. For further details, see [3, chs. 1, 2], [11, ch. 3] and [17, chs. 1, 2, 3].

is a single-valued measurable function, $\Delta(x_1, \cdots, x_n)$ of $x_1, \cdots, x_n$ on R^n. If $x_1, \cdots, x_n$ are random variables, then $y_1, \cdots, y_p$ are random variables, and[7]

$$(2.1) \qquad E\{\Omega(y_1, \cdots, y_p)\} = E\{\Delta(x_1, \cdots, x_n)\}.$$

We shall call $E(x_i)$ the mean value of x_i, σ_{ij} the covariance of x_i and x_j, and σ_{ii} or σ_i^2 the variance of x_i, where $\sigma_{ij} = E\{(x_i - Ex_i)(x_j - Ex_j)\}$.

The Laplace-Liapounoff, or Central Limit theorem states conditions under which linear functions of random variables have a normal limiting distribution. The general characteristic of the proofs of the theorem is that conditions are placed on the random variables so that they may virtually be assumed to be bounded. The Lindeberg[8] condition, which we shall use, is perhaps the least restrictive of all the conditions which require finite means and variances.

The Lindeberg condition[9], $\mathfrak{L}_p$: A set of random variables $x_{i\nu n}$ will be said to satisfy the Lindeberg condition $\mathfrak{L}_p$ if there exists, for any preassigned positive real numbers δ and ϵ, a positive integer n_0 such that if $n > n_0$, then

$$\sum_\nu \int_{|z_{\nu n}|>\epsilon} z_{\nu n}^2 \, dF(x_{1\nu n}, \cdots, x_{p\nu n}) < \delta,$$

where

$$z_{\nu n}^2 = x_{1\nu n}^2 + x_{2\nu n}^2 + \cdots + x_{p\nu n}^2$$

and

$$\sigma_{i1n}^2 + \sigma_{i2n}^2 + \cdots + \sigma_{inn}^2 = 1.$$

If

$$x_{i\nu n} = \frac{x_{i\nu}}{s_{in}} \quad \text{where} \quad s_{in}^2 = \sigma_{i1}^2 + \cdots + \sigma_{in}^2,$$

and the $x_{i\nu n}$ satisfy $\mathfrak{L}_p$ then we shall say that the $x_{i\nu}$ satisfy $\mathfrak{L}_p$.

Suppose that the random variables $y_{11}, \cdots, y_{pm_p}$ have a normal multivariate distribution with zero means and with covariance parameters $\sigma_{i\gamma j\delta}$ where

$$\sigma_{i\gamma j\delta} = E(y_{i\gamma}y_{j\delta}), \gamma = 1, \cdots, m_i \, ; \delta = 1, \cdots, m_j,$$

and denote the distribution function of $y_{11}, \cdots, y_{pm_p}$ by $N(y)$. Then we may state the Laplace-Liapounoff theorem as:

---

[7] It is noted that $\Omega(y_1, \ldots, y_p)$ is integrated with respect to $F(y_1, \ldots, y_p)$ and $\Delta(x_1, \ldots, x_n)$ is integrated with respect to $F(x_1, \ldots, x_n)$.

[8] See Cramer [3, pp. 57, 60, 114], and the references there given.

[9] It is not difficult to show that the Lindeberg condition will be satisfied if moments of order greater than two exist, [3, p. 60], or if the conditions stated by Levy [13, p. 207] and [14, p. 106] are satisfied.

THEOREM I. *Suppose that, for each value of n, the random variables $x_{i\gamma\nu n}$, which are independent for different values of ν, have zero means and covariance parameters $\sigma_{i\gamma j\delta\nu n}$, where*

$$\sigma_{i\gamma j\delta\nu n} = E(x_{i\gamma\nu n}x_{j\delta\nu n}).$$

Denote by d'_n the maximum of the variances $\sigma_{i\gamma i\gamma\nu n}$. If the functions $y_{i\gamma n}$ are defined by the equations

$$y_{i\gamma n} = \sum_\nu x_{i\gamma\nu n},$$

it follows that

$$\sigma_{i\gamma j\delta n} = E(y_{i\gamma n}y_{j\delta n}) = \sum_\nu \sigma_{i\gamma j\delta\nu n}.$$

If $\lim_{n\to\infty} \sigma_{i\gamma j\delta n} = \sigma_{i\gamma j\delta}$ and if $\lim_{n\to\infty} d'_n = 0$, then a necessary and sufficient condition that as $n \to \infty$, the limiting distribution[10] *of $y_{11n}, \cdots, y_{pm_pn}$ be $N(y)$ is that the condition $\mathfrak{L}_{pm_p}$ be satisfied.*

The proof of this theorem is omitted. It may readily be developed from the proofs of Cramer, [3, pp. 57, 113].

Before stating certain corollaries which are of interest, some additional definitions are necessary.

Let $C_n, C_{n+1}, \cdots$ be a sequence of m rowed real matrices

$$C_n = \| c_{\gamma\nu n} \|, \qquad n = m, m+1, \cdots,$$

and let the greatest of the absolute values of the elements of C_n be denoted by d_n. The inner product of any two rows of C_n will be denoted by $\rho_{\gamma\delta n}$, i.e.

$$\rho_{\gamma\delta n} = \sum_\nu c_{\gamma\nu n}c_{\delta\nu n}.$$

Let $X_1, X_2, \cdots$ be a sequence of random vectors of p components defined on R^p, and let the components of X_μ be denoted by $x_{1\mu}, \cdots, x_{p\mu}$. Let the components of the chance matrix $Y_n = \| y_{i\gamma n} \|$ which has p rows and m columns, be defined by the equations

$$y_{i\gamma n} = \sum_\nu c_{\gamma\nu n}x_{i\nu} \tag{2.2}$$

for each value of n, $(n = m, \cdots; m \geq p)$.

---

[10] The distribution functions $F(X_n)$ will be said to converge to the distribution function $F(X)$ if and only if

$$\lim_{n\to\infty} \int_{-\infty}^{x} dF(X_n) = F(X)$$

for every X at which $F(X)$ is continuous. If $F(X)$ is continuous throughout R^n, then the convergence is uniform.

Suppose that

$$E(x_{i\nu}) = 0 \tag{2.3}$$

and

$$E(x_{i\nu}x_{j\mu}) = \sigma_{ij}\delta_{\mu\nu}, \tag{2.4}$$

where $\delta_{\mu\nu} = 1$ if $\mu = \nu$ and $\delta_{\mu\nu} = 0$ if $\mu \neq \nu$. (There should be no confusion of this use of the letter δ with its use as an index.) It is easy to see that if the $c_{\gamma\nu n}$ are real numbers, then

$$E(y_{i\gamma n}) = 0$$

and

$$E(y_{i\gamma n}y_{j\delta n}) = \sigma_{ij}\rho_{\gamma\delta n}.$$

Let the determinant of the positive definite symmetric matrix, $(\sigma) = \| \sigma_{ij} \|$ be denoted by σ. Let the inverse matrix of (σ) be denoted by $(\sigma)^{-1} = \| \sigma^{ij} \|$ where σ^{ij} is the cofactor of σ_{ij} in (σ) divided by σ. The determinant of $(\sigma)^{-1}$ is σ^{-1}

By $N_d(x_1, \cdots, x_p; (\sigma))$ we shall mean the normal probability density with zero means and covariance parameters σ_{ij}, i.e.,

$$N_d(x_1, \cdots, x_p; (\sigma)) = (2\pi\sigma)^{-\frac{1}{2}} \exp\left[-\tfrac{1}{2}\sum_{i,j}\sigma^{ij}x_ix_j\right], \qquad (-\infty < x_i < \infty),$$

where (σ) is a positive definite matrix. If the random variables $x_1, \cdots, x_p$ have probability density $N_d(X; (\sigma)) \equiv N_d(x_1, \cdots, x_p; (\sigma))$, where X is a vector, then we shall say that X has a distribution function $N(X; (\sigma))$, i.e.

$$\frac{\partial^p}{\partial x_1 \cdots \partial x_p} N(X; (\sigma)) = N_d(X; (\sigma))$$

or

$$\int_{\infty}^{x_p} \cdots \int_{\infty}^{x_1} N_d(t_1, \cdots, t_p; (\sigma))\, dt_1 \cdots dt_p = N(X; (\sigma)).$$

Inasmuch as certain hypotheses will be used on several occasions in this paper, they are stated here.

If $x_1, x_2, \cdots$ are independently distributed, if (2.3) and (2.4) hold and if the x's satisfy the condition $\mathcal{L}_p$ then we shall say that $\mathcal{H}_p$ is true.

If C_n is such that, for all n, the equations $\rho_{\gamma\delta n} = \delta_{\gamma\delta}$ are true, we shall say that $\mathcal{C}$ is true.

The following corollary is useful in deriving limiting distributions in the analysis of variance.

CORROLLARY I. *Let $\mathcal{H}_p$ and $\mathcal{C}$ be true. Then a sufficient condition that*

$$\lim_{\infty} F(Y_n) = \prod_{\gamma} N(y_{1\gamma}, \cdots, y_{p\gamma}; (\sigma))$$

is $\lim_{n\to\infty} d_n = 0$.

The proof is based on the fact that the $x_{i\gamma\nu n}$ of Theorem I are given by $c_{\gamma\nu n}x_{i\nu}$. The details are omitted.

The pm rowed square matrix, $(\tau) = ||'\tau_{ns}||$ is defined as follows: If $r \leq m$, $s \leq m$; then $\tau_{rs} = \sigma_{11}\rho_{rs}$; and if $km < r \leq (k+1)m$, $lm < s \leq (l+1)m$, $l, k = 0, \cdots, p-1$, then $\tau_{rs} = \sigma_{k+1\,l+1}\rho_{r-km\,s-ln}$. The inverse matrix of (τ), and the determinants of (τ) and $(\tau)^{-1}$ are defined as are $(\sigma)^{-1}$, σ and σ^{-1}.

Corollary II. *Let $\mathcal{H}_p$ be true, and let*

$$\lim_{n\to\infty} \rho_{\gamma\delta n} = \rho_{\gamma\delta}, \qquad \rho_{\gamma\gamma} = 1.$$

Then, if $\lim_{n\to\infty} d_n = 0$, *it follows that*

$$\lim_{n\to\infty} F(Y_n) = F(Y),$$

where $F(Y)$ is the distribution function determined by the probability density

$$(2\pi)^{-\frac{pm}{2}}\tau^{-\frac{1}{2}}\exp\left[-\frac{1}{2}\sum_{r,s=1}^{pm}\tau^{rs}y_{k+1\;r-km}\,y_{l+1\;s-lm}\right]$$

where, if $r \leq m$, $s \leq m$, then $k = 0, l = 0$; if $r \leq m$, $m < s \leq 2m$, then $k = 0$, $l = 1$; and so on.

The proof is omitted.

If $Z_1, \cdots, Z_t$ are random variables, then $F(X_1, \cdots, X_k \mid Z_1, \cdots, Z_t)$ is the distribution function of the random vectors $X_1, \cdots, X_k$ for fixed values of $Z_1, \cdots, Z_t$, i.e. for any fixed values of $Z_1, \cdots, Z_t$,

$$P\{X_1 < X_1, \cdots, X_k < X_k\} = F(X_1, \cdots, X_k \mid Z_1, \cdots, Z_t).$$

We shall now assume that the elements $c_{\gamma\nu n}$ of the matrix C_n are Borel measurable functions of a set of random variables[11] $Z_1, \cdots, Z_{t_n}$. Then the matrix C_n may be called a random matrix defined on a space W_n which is the combinatory product of the spaces on which $Z_1, \cdots, Z_{t_n}$ are defined. If, for each value of n, and for all X^n and Z^n, the equation

$$F(X^n, Z^n) = F(Z^n)\cdot\prod_\nu F(X_\nu \mid Z^n) \tag{2.5}$$

is satisfied, then we shall say that $\mathcal{I}$ is true. It is obvious that sufficient conditions for the truth of $\mathcal{I}$ are

$$F(X^n, Z^n) = F(Z^n)\cdot\prod_\nu F(X_\nu)$$

or, if $t_n \geq n$

$$F(X^n, Z^n) = F(Z_{n+1}, \cdots Z_{t_n})\cdot\prod F(X_\nu, Z_\nu)$$

[11] The symbol X^n will stand for the set of variables $X_1, \ldots, X_n$, and the symbol Z^n will stand for the set of variables $Z_1, \ldots, Z_{t_n}$.

or, if $t_n \leq n$

$$F(X^n, Z^n) = \prod_{\nu=1}^{t_n} F(X_\nu, Z_\nu) \cdot \prod_{\nu=t_n+1}^{n} F(X_\nu).$$

Inasmuch as we shall often use Fubini's theorem, it is now stated here.[12]

THEOREM II. *Let the distribution function of X^n, Z^n be $F(X^n, Z^n)$, let the distribution function of X^n for fixed values of Z^n be $F(X^n \mid Z^n)$, and let the distribution function of Z^n be $F(Z^n)$. Then if $\Delta(X^n, Z^n)$ is measurable with respect to $F(X^n, Z^n)$ and if*

$$\int_{R^{pn} \times W_n} |\Delta(X^n, Z^n)| \, dF(X^n, Z^n) < \infty,$$

it follows that

$$\int_{R^{pn}} |\Delta(X^n, Z^n)| \, dF(X^n \mid Z^n) < \infty$$

for almost all[13] sets of values of Z^n and

$$\int_{R^{pn} \times W_n} \Delta(X^n, Z^n) \, dF(X^n, Z^n) = \int_{W_n} \left[\int_{R^{pn}} \Delta(X^n, Z^n) \, dF(X^n \mid Z^n) \right] dF(Z^n).$$

In Corollary I an important condition was that the maximum of the absolute values of the elements of C_n should approach zero as n increased. In order to obtain a similar condition when the elements of C_n are random variables, we shall define the function $d(C_n)$ as follows: For each value of Z^n let $d(C_n)$ be the maximum of the absolute values of the elements of C_n. We shall denote $d(C_n)$ by d_n. If the elements of C_n are Borel measurable functions then d_n is a Borel measurable function of Z^n. Hence d_n is a random variable defined on W_n.

A sequence of random variables $d_1, d_2, \cdots$ is said to converge in probability to zero if, given $\epsilon > 0$, then

$$\lim_{n \to \infty} P\{|d_n| > \epsilon\} = 0.$$

If the sequence of functions $d_p, d_{p+1}, \cdots$ converges in probability to zero we shall say that $\mathfrak{Z}$ is true.

If $\mathcal{I}$ is true, and if, for almost all values of Z^n we have

$$\int_{R^p} x_{i\nu} \, dF(X_\nu, Z^n) = 0, \tag{2.6}$$

$$\int_{R^p} x_{i\nu} x_{j\nu} \, dF(X_\nu, Z^n) = \sigma_{ij}, \tag{2.7}$$

---

[12] Proofs of Fubini's theorem with the required amount of generality will be found in [5, p. 101] and [14, p. 73].

[13] A proposition concerning random variables is said to be true for almost all values of the variables, if it is true for all values of the variables, except perhaps for a set of probability zero with respect to the distribution function of the random variables.

and the condition $\mathfrak{L}_p$ is satisfied with respect to the X and the distribution functions $F(X_\nu, Z^n)$ then we shall say that $\mathfrak{K}_p^0$ is true.

If

$$\sum_\nu \int_{R^p \times W_n} c_{\gamma\nu n} c_{\delta\nu n} x_{i\nu} x_{j\nu} \, dF(X_\nu, Z^n) = \sigma_{ij} \delta_{\gamma\delta}, \tag{2.8}$$

then we shall say that $\mathcal{C}^0$ is true. It is noted that if $\mathcal{I}$ and (2.7) are true, then $\mathcal{C}^0$ is true if $\mathcal{C}$ is true for almost all sets of fixed values of Z^n.

Corollary III. *Let $\mathcal{C}^0$, $\mathcal{I}$ and $\mathfrak{K}_p^0$ be true. Then, if $\mathcal{Z}$ is true, it follows that*

$$\lim_{n \to \infty} F(Y_n) = \prod_\gamma N(y_{1\gamma}, \cdots, y_{p\gamma}; (\sigma)).$$

Proof. It is necessary to show that the condition $\mathfrak{L}_{pm}$ is satisfied by the variables $c_{\gamma\nu n} x_{i\nu}$ if the condition $\mathfrak{L}_p$ is satisfied by the variables $x_{i\nu}$ and that the condition $\mathcal{Z}$ implies that $\lim_{n \to \infty} d_n = 0$ when the $x_{i\gamma\nu n}$ of Theorem I are set equal to the $c_{\gamma\nu n} x_{i\nu}$ of Corollary III.

If we let $\Delta_{\nu n}^2 = \sum_{\gamma, i} (c_{\gamma\nu n} x_{i\nu})^2$, $\Delta_n^2 = \sum_\nu \Delta_{\nu n}^2$ and let $s_n^2 = E\{\Delta_n^2\}$, then, by (2.8), it is true that

$$s_n^2 = \sum_{\gamma, i} \sigma_{ii} = m \sum_i \sigma_{ii}.$$

From $\mathfrak{K}_p^0$ and the fact that for sufficiently large n, $| d_n^2(Z^n) | < 1$ for almost all Z^n we have for any preassigned ϵ and δ,

$$\frac{1}{s_n^2} \int_{\Delta_n > \epsilon s_n} \Delta_n^2 \, dF(X^n, Z^n) \le \frac{1}{s_n^2} \sum_\nu \int_{\Delta_n > \epsilon s_n} m \, d_n^2(Z^n) \sum_i x_{i\nu} \, dF(X_\nu, Z^n) < \delta$$

for sufficiently large n, since the set of x's and Z^n for which $\sum_{i,\nu} x_{i\nu}^2 > \epsilon s_n$ contains almost all the x's and Z^n for which $\Delta_n > \epsilon s_n$. Hence, the condition $\mathfrak{L}_{pm}$ is satisfied by the random variables $c_{\gamma\nu n} x_{i\nu}$ with respect to the distribution functions $F(X_\nu, Z^n)$.

We now show that

$$\lim_{n \to \infty} [\max E\{(c_{\gamma\nu n} x_{i\nu})^2\}] = 0.$$

It is clearly true that

$$E\{(c_{\gamma\nu n} x_{i\nu})^2\} \le \int_{R^p \times W_n} d_n^2 x_{i\nu}^2 \, dF(X_\nu, Z^n).$$

Since d_n converges in probability to zero, and since $d_n^2 \le 1$ for almost all Z, we can, for any $\epsilon > 0$, take n_0 so large that if $n > n_0$, then $P\{d_n^2 > \frac{1}{2}\epsilon\} < \frac{1}{2}\epsilon$. If E is the set on which $d_n^2 > \frac{1}{2}\epsilon$, we then have for all $n > n_0$, using (2.7),

$$E\{(c_{\gamma\nu n} x_{i\nu})^2\} \le \int_E \left[\int_{R^p} x_{i\nu}^2 \, dF(X_\nu \mid Z^n) \right] dF(Z^n) + \frac{\epsilon}{2} \int_{W_n} \left[\int_{R^p} x_{i\nu}^2 \, dF(X_\nu \mid Z^n) \right] dF(Z^n) \le \epsilon \sigma_{ii}$$

and this inequality is also satisfied for all $n > n_0$.

The following discussion is useful in obtaining the limiting distributions of statistics which occur in multivariate statistical analysis.

The letter f will assume all integral values from 1 through s, the letters μ, ν will assume all integral values from 1 through n_f, and the letters γ, δ will assume all integral values from 1 through m_f, for any f.

Let $X_1^f, \cdots$ be, for any fixed f, a sequence of random vectors of p_f components defined on R^{p_f}, and let the set of random variables $X_1^f, \cdots$ be independently distributed for any fixed f.

If, for each set of values of $n_1, \cdots, n_s$, (t_n is a function of $n_1, \cdots, n_s$),

$$F(X_1^1, \cdots, X_{n_s}^s, Z_1, \cdots, Z_{t_n}) = \prod_f \prod_\nu F(X_\nu^f \mid Z_1, \cdots, Z_{t_n}) \cdot F(Z_1, \cdots, Z_{t_n}),$$

we shall say that $\mathcal{I}_{sn}$ is true.

Let, for any fixed value of f, the matrix[14] $C_n^f = \| c_{\gamma\nu n}^f \|$ where the $c_{\gamma\nu n}^f$ are Borel measurable functions of X_μ^k, ($k < f$), and[15] Z^n, have the same properties as C_n, and let $d(C_n^f)$ be the same function of C_n^f that $d(C_n)$ is of C_n. We shall denote $d(C_n^f)$ by d_n^f.

Let

$$y_{i\gamma n}^f = \sum_\nu c_{\gamma\nu n}^f x_{i\nu}^f$$

and let $Y_n^f = \| y_{\gamma\delta n}^f \|$.

For fixed f, the p_f rowed square matrix (σ_f), its inverse, and so on are defined as were the same functions of the σ_{ij} earlier in this paragraph but with σ_{ijf} replacing σ_{ij}, where

$$E\{x_{i\nu}^f\} = 0$$

and

$$E\{x_{i\nu}^f x_{j\nu}^f\} = \sigma_{ijf}.$$

If $\mathcal{I}_{sn}$ is true, and if for almost all values of Z^n we have

$$\int_{R^{p_f}} x_{i\nu}^f \, dF(X_\nu^f, Z^n) = 0, \tag{2.9}$$

$$\int_{R^{p_f}} x_{i\nu}^f x_{j\nu}^f \, dF(X_\nu^f, Z^n) = \sigma_{ijf}, \tag{2.10}$$

and the condition $\mathcal{L}_{p_f}$ is satisfied with respect to the X_ν^f and the distribution functions $F(X_\nu^f, Z^n)$ then we shall say that $\mathcal{K}_{p_f}^f$ is true.

If

$$\sum_\nu \int c_{\gamma\nu n}^f c_{\delta\nu n}^f x_{i\nu}^f x_{j\nu}^f \, dF(X_\nu^f, Z^n) = \sigma_{ijf}\delta_{\gamma\delta}, \tag{2.11}$$

[14] The superscripts f and k will not indicate multiplication but will only be indices.

[15] See footnote 11.

then we shall say that $\mathcal{C}^f$ is true. It is noted that if $\mathcal{G}_f$ and (2.10) are true then $\mathcal{C}^f$ is true if $\mathcal{C}$ is true for almost all sets of fixed values of $X_1^1, \cdots, X_n^{f-1}, Z^n$.

If d_n^f converges in probability to zero as n increases we shall say that $\mathcal{Z}_f$ is true.

Corollary IV. *Let $\mathcal{C}^s$, $\mathcal{G}_s$ and $\mathcal{H}_{p_1}^1, \cdots, \mathcal{H}_{p_s}^s$ be true. Then, if $\mathcal{Z}_1, \cdots, \mathcal{Z}_s$ are true, it follows that*

$$\lim_{n_1, \cdots, n_s \to \infty} F(Y_{n_1}^1, \cdots, Y_{n_s}^s) = \prod_f F(Y^f),$$

where

$$F(Y^f) = \prod_\gamma N(y_{1\gamma}^f, \cdots, y_{p_f\gamma}^f; (\sigma_f)).$$

The proof is almost identical with the proof of Corollary III of which this corollary is an extension.

It is remarked that if the statistics, the limiting distributions of which are desired, are associated with the normal distribution, as are most statistics studied, then Corollary IV may not be the best tool to use. This is a consequence of the fact that such statistics are generally expressible as functions of uncorrelated random variables and hence are more simply discussed, using Corollary I.

3. Limiting distributions of quadratic and bilinear forms. We first assume the coefficients of the forms to be constants. For each set of values of i, j, and n, the matrix of the bilinear form with coefficients which are real numbers,

$$b_{ij}^n = \sum_{\mu,\nu} a_{\mu\nu n} x_{i\mu} x_{j\nu}, \tag{3.1}$$

will be denoted by A_n, and the rank of A_n will be denoted by m. The maximum of the absolute values of the elements of A_n will be denoted by b_n. We shall assume that there exists an orthogonal transformation,

$$y_{i\mu n} = \sum_\nu c_{\mu\nu n} x_{i\nu}, \tag{3.2}$$

of $x_{i1}, \cdots, x_{in}$ such that

$$b_{ij}^n = \sum_\delta \lambda_\delta y_{i\delta n} y_{j\delta n}, \tag{3.3}$$

where the coefficients λ_δ are non-negative.[16]

Lemma I. *If d_n is the maximum of the absolute values of the elements $c_{\mu\nu n}$ then a necessary and sufficient condition that* $\lim_{n\to\infty} b_n = 0$ *is* $\lim_{n\to\infty} d_n = 0$.

[16] Our theorems will not be applicable if some of the λ_δ are negative and some are positive. However if all the λ_δ are non-positive then the theorems will remain true.

PROOF: From (3.1) it follows that

$$a_{\mu\nu n} = \sum_{\delta} \lambda_\delta c_{\delta\mu n} c_{\delta\nu n}.$$

Hence, $b_n \geq a_{\mu\mu n} \geq \lambda_\mu c_{\delta\mu n}^2$ and $| a_{\mu\nu n} | \leq d_n^2 \left(\sum_\delta \lambda_\delta\right)$. The remainder of the proof is obvious.

The following theorem will be the basis for a large sample analogue of Wishart's distribution.

THEOREM III. *Let $\mathcal{H}_p$ be true. Then, a sufficient condition that*

$$\lim_{n\to\infty} F(Y_n) = \prod_\gamma N(y_{1\gamma}, \cdots, y_{p\gamma}; (\sigma)),$$

where $b_{ij}^n = \sum_\delta \lambda_\delta y_{i\delta n} y_{j\delta n}$ *is* $\lim_{n\to\infty} b_n = 0$.

PROOF. According to Lemma I, the fact that $\lim_{n\to\infty} b_n = 0$, implies that $\lim_{n\to\infty} d_n = 0$. The $y_{i\nu n}$ are such that $\mathcal{C}$ is true. Hence the hypotheses of Corollary I are satisfied and the theorem is proved.

Before stating the corollary to Theorem III, we shall prove an obvious lemma which is of constant service.

LEMMA II. *Let* $\lim_{n\to\infty} F(X_n) = F(X)$ *at all points of continuity of* $F(X)$, *and let*

$$g_{1n} = g_1(x_{1n}, \cdots, x_{pn}), \cdots, g_{kn} = g_k(x_{1n}, \cdots, x_{pn})$$

be Borel measurable functions of their indicated variables for each value of n, $(p \gtreqless k)$, *defined on* R^p.

Then

$$\lim_{n\to\infty} F(g_{1n}, \cdots, g_{kn}) = F(g_1, \cdots, g_k)$$

at all points of continuity of $F(g_1, \cdots, g_k)$, *where* $g_\alpha = g_\alpha(x_1, \cdots, x_p)$.

PROOF. By (2.1), we have

$$E[e^{i\sum_\alpha t_\alpha g_\alpha(x_{1n}, \cdots, x_{pn})}] = E[e^{i\sum_\alpha t_\alpha g_{\alpha n}}], \tag{3.4}$$

where since $g_\alpha(x_1, \cdots, x_p)$ is a Borel measurable function of $x_1, \cdots, x_p$ we know that $g_{1n}, \cdots, g_{kn}$ have a joint distribution function $F(g_{1n}, \cdots, g_{kn})$. Then, since $\lim_{n\to\infty} F(X_n) = F(X)$ at all points of continuity of $F(X)$ we have[17]

$$\lim_{n\to\infty} E[e^{i\sum_\alpha t_\alpha g_\alpha(x_{1n}, \cdots, x_{pn})}] = E[e^{i\sum_\alpha t_\alpha g_\alpha(x_1, \cdots, x_p)}]$$

uniformly in every $t_1, \cdots, t_p$ interval since

$$| E[e^{i\sum_\alpha t_\alpha g_\alpha(x_{1n}, \cdots, x_{pn})}] - E[e^{i\sum_\alpha t_\alpha g_\alpha(x_1, \cdots, x_p)}] |$$
$$\leqslant \int | dF_n(X_1, \cdots, X_p) - F(X_1, \cdots, X_p) |,$$

[17] See Cramer, [3, p. 30] and "Additional Note" at the end of the book.

where $F_n(X_1, \cdots, X_p)$ stands for $F(X_{1n}, \cdots, X_{pn})$, when X_i and X_{in} have the same numerical values. If follows from (3.4), that

$$\lim_{n\to\infty} E[e^{i\sum_\alpha t_\alpha g_{\alpha n}}] = E[e^{i\sum_\alpha t_\alpha g_\alpha}]$$

uniformly in every $t_1, \cdots, t_p$ interval, and consequently

$$\lim_{n\to\infty} F(g_{1n}, \cdots, g_{kn}) = F(g_1, \cdots, g_k)$$

at all points of continuity of $F(g_1, \cdots, g_k)$.

The real valued function $G_d(x; n, c)$ will be defined by the equations

$$G_2(0; 0, c) = 1, \qquad (-\infty < c < \infty),$$

$$G_d(x; n, c) = [\Gamma(\tfrac{1}{2}n)]^{-1} (2c)^{-\frac{1}{2}n} x^{\frac{1}{2}n-1} \exp\left[-\frac{x}{2c}\right], \qquad (0 < x < \infty; c > 0; n > 0),$$

and $G_d(x; n, c) = 0$ otherwise. The function $G(x; n, c)$ will be defined by the equation

$$G(x; n, c) = \int_0^x G_d(t; n, c)\, dt.$$

The real valued function $G_d(x_{11}, x_{12}, \cdots, x_{pp}; n, (\sigma))$ will be defined by the equations

$$G_d(0, \cdots, 0; p-1, (\sigma)) = 1$$

$$G_d(x_{11}, \cdots, x_{pp}; n; (\sigma)) = (2\pi)^{-\frac{1}{4}p(p-1)} \sigma^{-\frac{1}{2}n} \cdot [\prod_i \Gamma\tfrac{1}{2}(n-i+1)]^{-1} \cdot |x|^{\frac{1}{2}(n-p+1)-1}$$

$$\cdot \exp[-\tfrac{1}{2} \sum_{i,j} \sigma^{ij} x_{ij}], \qquad (0 < x_{ii} < \infty; x_{ij}^2 \leq x_{ii}x_{jj}); (\sigma) \text{ is positive definite},$$

where $|x|$ is the determinant $|x_{ij}|$ and $G_d(x_{11}, \cdots, x_{pp}; n, (\sigma)) = 0$ otherwise. The function $G(x_{11}, \cdots, x_{pp}; n, (\sigma))$ will be defined by the equation

$$G(x_{11}, \cdots, x_{pp}; n, (\sigma)) = \int_{-\infty}^{x_{pp}} \cdots \int_{-\infty}^{x_{11}} G_d(t_{11}, \cdots, t_{pp}; n, (\sigma))\, dt_{11} dt_{12} \cdots dt_{pp}.$$

We can now state the limiting distribution analogue of Wishart's distribution.

COROLLARY V. *If $\mathcal{H}_p$ is true, if $\lambda_\delta = 1$, and if $m \geq p$ then*

$$\lim_{n\to\infty} F(b_{11}^n, b_{12}^n, \cdots, b_{pp}^m) = G(b_{11}, \cdots, b_{pp}; m, (\sigma)).$$

PROOF. The conditions of Theorem III and Lemma II are satisfied.

Obviously for fixed i, the limiting distribution of b_{ii}^n is $G(b; m, \sigma_{ii})$, and if $i \neq j$, the limiting distribution of b_{ij}^n/m is the distribution of the covariance of x_i and x_j in a sample of m independent pairs of observations.[18]

[18] See Wishart and Bartlett, [1, p. 266].

We proceed to the analogue for limiting distributions of one of our generalizations of the Fisher-Cochran theorem. It is first desirable to give some additional definitions.

We consider the bilinear forms

$$b_{ij\alpha}^{n} = \sum_{\mu,\nu} a_{\mu\nu n}^{\alpha} x_{i\mu} x_{j\nu} \tag{3.5}$$

with real coefficients, and we denote the matrix of $b_{ij\alpha}^{n}$ by A_n^α. The rank of A_n^β is m_β, and the rank of A_n^k is m_{kn}. If the maximum of the absolute values of the elements of $A_n^1, \cdots, A_n^{k-1}$ is b_n, and if there exists an orthogonal transformation,

$$y_{i\mu n} = \sum_{\nu} c_{\mu\nu n} x_{i\nu}, \tag{3.6}$$

of $x_{i1}, \cdots, x_{in}$ such that

$$b_{ij\alpha}^{n} = \sum_{\delta} \lambda_\delta y_{i\delta n} y_{j\delta n},$$

where δ assumes all integral values from $m_1 + \cdots + m_{\alpha-1} + 1$ through $m_1 + \cdots + m_\alpha$ and λ_δ is non-negative, then it is easy to prove, as in Lemma I, that a necessary and sufficient condition that $\lim_{n\to\infty} b_n = 0$ is $\lim_{n\to\infty} d_n = 0$, where d_n is the maximum of the absolute values of the elements $c_{\mu\nu n}$.

LEMMA III. *Let* $m = m_1 + \cdots + m_{k-1}$ *and let*

$$\sum_{\alpha} b_{ij\alpha}^{n} = \sum_{\nu} x_{i\nu} x_{j\nu}. \tag{3.7}$$

Then, a necessary and sufficient condition that

$$b_{ij\alpha}^{n} = \sum_{\delta} y_{i\delta n} y_{j\delta n},$$

where the real linear functions, $y_{i\delta n}$, *of* $x_{i1}, \cdots, x_{in}$ *are given by* (3.6), *the linear functions* (3.6) *not now being assumed to be orthogonal, is*

$$m_{kn} = n - m.$$

Furthermore, the functions (3.6) *are orthogonal.*

The proof of this lemma for the case $p = 1$ is given in [16]. The procedure to follow in extending the lemma to the cases where $p > 1$, is given in [15, p. 473]. It is noted that this lemma is more general than the lemma in [15] inasmuch we we show that the orthogonality of the transformation is a consequence of our hypotheses and not one of the hypotheses.[19]

[19] It is noted, however, that the increase in generality affects only the necessity not the sufficiency of the theorem.

THEOREM IV. *Let* $\mathcal{H}_p$, (3.7) *and* (3.8) *be true for all values of* n, *and suppose that* $\lim_{n\to\infty} b_n = 0$. *Then*

$$\lim_{n\to\infty} F(y_n) = \prod_{\gamma} N(y_{1\gamma}, \cdots, y_{p\gamma}; (\sigma)),$$

where $b^n_{ij\alpha} = \sum_{\delta} y_{i\delta n} y_{j\delta n}$.

The proof is omitted.

COROLLARY VI. *If the hypotheses of Theorem IV are assumed, and if* $m_\beta \geq p$; $(\beta = 1, \cdots, h; h < k)$, *then*

$$\lim_{n\to\infty} F(b^n_{111}, \cdots, b^n_{pph}, y_{1\,h+1\,n}, \cdots, y_{pmn})$$

$$= \prod_{\gamma=1}^{h} G(b_{11\gamma}, \cdots, b_{pp\gamma}; m_\gamma, (\sigma)) \cdot \prod_{\gamma=h+1}^{m} N(y_{1\gamma}, \cdots, y_{p\gamma}; (\sigma)).$$

If $p = 1$ in Theorem IV and Corollary VI, we have the large sample analogue of the Fisher-Cochran theorem.

We now discuss limiting distributions of random variables which are bilinear and quadratic forms in one set of chance variables for fixed values of other random variables. We consider the coefficients $a_{\mu\nu n}$ and $a^\alpha_{\mu\nu n}$ of b^n_{ij} and $b^n_{ij\alpha}$ to be random variables. Hence the matrices A_n and A^α_n are random matrices.

To be more explicit, let $X^f_1, X^f_2, \cdots$ be a sequence of random vectors, the random vector X^f_n having p_f components $x^f_{1n}, \cdots, x^f_{p_f n}$, and being defined on R^{p_f}. The set of random vectors X^f_ν and $Z_1, \cdots, Z_{t_n}$ will be assumed to be independent.

For each value of f the coefficients of the bilinear forms

$$(3.9) \qquad b^{n_f}_{ij\alpha f} = \sum_{\mu,\nu=1}^{n_f} a^{n_f}_{\mu\nu\alpha f} x^f_{i\mu} x^f_{i\nu}, \qquad (i, j = 1, \cdots, p_f; \alpha = 1, \cdots, k_f)$$

will be assumed to be Borel measurable functions of the random vectors $X^1_\mu, \cdots, X^{f-1}_\mu$ and $Z_1, \cdots, Z_{t_n}$.

The matrix of $b^{n_f}_{ij\alpha f}$ is denoted by $A^{\alpha f}_{n_f}$. The rank of $A^{\beta f}_{n_f}$ is $m_{\beta f}$ and the rank of $A^{k_f f}_{n_f}$ is $m_{k_f n_f}$ for all sets of values of the $a^{n_f}_{\mu\nu\alpha f}$ except, perhaps, on a set E_{n_f} which is such that $\lim_{n_f\to\infty} P(E_{n_f}) = 0$.

Let the function $b(A^{\beta f}_{n_f})$ be defined as follows:
For each set of values of the X^f_μ and Z let $b(A^{\beta f}_{n_f})$ be the maximum of the absolute values of the elements of $A^{\beta f}_{n_f}$. We shall denote $b(A^{\beta f}_{n_f})$ by $b^{\beta f}_{n_f}$. Obviously, $b^{\beta f}_{n_f}$ is a Borel measurable function of X^f_μ and Z. Hence

$$b^{\beta f}_{n_f} = b(A^{\beta f}_{n_f})$$

is a random variable defined on $W \times R^{n_1 p_1 + \cdots + n_s p_s}$.

For each value of f, and for almost all sets of fixed values of the X_μ^h, $(h = 1, \cdots, f-1)$, we shall assume that there exists an orthogonal transformation,

$$y_{i\mu n_f}^f = \sum_\nu c_{\mu\nu n_f}^f x_{i\nu}^f, \tag{3.10}$$

of $x_{i1}^f, \cdots, x_{in_f}^f$ such that[20]

$$b_{ij\alpha f}^{n_f} = \sum_\lambda y_{i\lambda n_f}^f y_{j\lambda n_f}^f, \tag{3.11}$$

where λ assumes all integral values from $m_{if} + \cdots + m_{\alpha-1\,f} + 1$ through $m_{1f} + \cdots + m_{\alpha f}$. The coefficients $c_{\mu\nu n_f}^f$ of the linear forms (3.10) are real single valued Borel measurable functions of the coefficients $a_{\mu\nu\alpha f}^f$ of the bilinear forms (3.9) for fixed values of the X_μ^h and Z^n. Let $c_{\mu\nu n_f}^f$ be the same function of the functions $a_{\mu\nu\alpha f}^f$ that $c_{\mu\nu n_f}^f$ is of the coefficients of the bilinear forms having constant coefficients. Furthermore, let $d_{n_f}^f$ be the same function of the matrix $C_{n_f}^f = \| c_{\mu\nu n_f}^f \|$ where $m = m_{1f} + \cdots + m_{k_f-1\,f}$, that $b_{n_f}^{\alpha f}$ is of $A_{m_f}^{\alpha f}$.

LEMMA IV. *A necessary and sufficient condition that $b_{n_f}^f$ converge in probability to zero as n increases is that $d_{n_f}^f$ converge in probability to zero as n increases.*

PROOF. Since

$$\sum_{\beta=1}^{k_f-1} a_{\mu\nu\beta f}^{n_f} = \sum_\lambda c_{\lambda\mu n_f}^f c_{\lambda\nu n_f}^f,$$

we have

$$(k_f - 1) b_{n_f}^f \geq \sum_{\beta=1}^{k_f-1} a_{\mu\mu\beta f}^{n_f} \geq [c_{\lambda\mu n_f}^f]^2$$

and

$$|a_{\mu\nu\alpha f}^{n_f}| \leq \{\sum_\lambda [c_{\lambda\mu n_f}^f]^2 \cdot \sum_\lambda [c_{\lambda\nu n_f}^f]^2\}^{\frac{1}{2}} \leq m_{\alpha f}[d_{n_f}^f]^2,$$

where λ assumes all integral values from $m_{1f} + \cdots + m_{\alpha-1\,f} + 1$ through $m_{1f} + \cdots + m_{\alpha f}$. The remainder of the proof is obvious.

In proving Theorem V we shall use a generalization of Lemma III which is proved in [15, p. 473].

THEOREM V. *Let $\mathcal{H}_{p_1}^1 \cdots \mathcal{H}_{p_s}^s$ be true, and suppose that*

$$\sum_\alpha b_{ij\alpha f}^{n_f} = \sum_{\nu=1}^{n_f} x_{i\nu}^f x_{j\nu}^f.$$

Then, if $b_{n_f}^f$ converges in probability to zero as n increases and if $m_f = n_f - m_{k_f n_f}$ for all values of n_f, it follows that

$$\lim_{n_1, \cdots, n_s \to \infty} F(y_{11n_1}^1, \cdots, y_{p_s m_s n_s}^s) = \prod_f \prod_{\gamma=1}^{m_f} N(y_{1\gamma}^f, \cdots, y_{p_f\gamma}^f; (\sigma^f)).$$

The proof is omitted.

---

[20] It is not necessary that the λ_δ be set equal to one as in (3.11). It is only somewhat easier to state the results.

COROLLARY VII. *If $m_{\alpha f} \geq p_f$, then*

$$\lim_{n_1, \cdots, n_s \to \infty} F(b^{n_1}_{1111}, \cdots, b^{n_s}_{p_s p_s k_s - 1\, s}) = \prod_f \prod_{\beta=1}^{k_f-1} G(b_{11\beta f}, \cdots, b_{p_f p_f \beta f}; m_{\beta f}, (\sigma^f)).$$

The proof is omitted.

Finally, let us assume that the vectors X^f_ν, for fixed ν are uncorrelated and for fixed f are independent. By that, we shall mean that $E(x^f_{i\nu} x^g_{j\nu}) = \sigma^f_{ij} \delta_{fg}$ and that for all n the set of random vectors X^f_ν are independent for the same or different superscripts providing the subscripts are all different. Let us also assume that the coefficients of the forms (3.9) are real numbers. Thus we have weakened the hypotheses of Theorem V concerning the random vectors, and we have strengthened the hypotheses of Theorem V concerning the forms (3.9). Inasmuch as we are generally concerned with the limiting distributions of statistics which occur in the analysis of the normal distribution, and many such statistics have been shown to be invariant under transformations into uncorrelated random variables,[21] Theorem VI and Corollary VIII will often be applicable.

THEOREM VI. *The statement of Theorem V is repeated.*

COROLLARY VIII. *The statement of Corollary VII is repeated.*

Another extension of these theorems may be obtained by allowing all the n_f to be equal, i.e. $n_1 = \cdots = n_s = n$, and by putting conditions on the forms (3.9) which enable us to say that for fixed i, f, μ and n, the set of random variables $c^f_{\mu\nu n} x^f_{i\nu}$ are independently distributed. Theorem I could then be used to obtain a very general result. However, except for the case dealt with above, the condition of independence appears to be rather restrictive, and the theorem is omitted.

4. Applications.

We first state the strong law of large numbers and a lemma which is very useful in the discussion of limiting distributions.

A sequence of random variables $X_1, \cdots$ will be said to converge with probability one[22] to a random variable X if

$$\lim_{n \to \infty} P\{|X_n - X| < \epsilon, |X_{n+1} - X| < \epsilon, \cdots, |X_{n+p} - X| < \epsilon\} = 1$$

for every value of $p \geq 0$, uniformly in p for every positive number ϵ. Upon setting $p = 1$, it is seen that convergence with probability one implies convergence in probability.

The strong law of large numbers[23] asserts that if the independent random variables $X, X_1, \cdots$ all have the same distribution function, and if $E(X)$ is finite, then the sequence of arithmetic means $\frac{1}{n} \sum_\nu X_\nu$ converges with probability one to $E(X)$.

[21] The regression transformation which yields the uncorrelated variables will be found in [15, p. 476, (3.2)].

[22] See Doob [4, p. 163], and Frechet, [9, p. 228].

[23] See Doob [4, p. 163], and Frechet, [9, p. 259]. A complete proof is given by Frechet.

Hence, if $E(x_{i\nu}) = 0$ and if σ_{ij} is finite, then $\frac{1}{n}\sum_{\nu} x_{i\nu}x_{j\nu} = s'_{ijn}$ converges with probability one to σ_{ij}. Since $\sum_{\nu}(x_{i\nu} - \bar{x}_{in})(x_{j\nu} - \bar{x}_{jn}) = \sum_{\nu} x_{i\nu}x_{j\nu} - n\bar{x}_{in}\bar{x}_{jn}$ where $\bar{x}_{in}$ is the arithmetic mean of $x_{i1}, \cdots, x_{in}$, and since $\bar{x}_{in}$ converges with probability one to zero, it follows that $s_{ijn} = s'_{ijn} - \bar{x}_{in}\bar{x}_{jn}$ converges with probability one to σ_{ij}. It is, of course, assumed that the random variables $x_{i\nu}$, $x_{j\nu}$ have the same joint distribution function for all values of ν, and that the random vectors $X_1, \cdots$ are independently distributed. The process of the reduction of s_{ijn} to s'_{ijn} in the limit, is an example of the possible uses of:

LEMMA V. *If $\varphi(t_1, \cdots, t_p)$ is a continuous function of $t_1, \cdots, t_p$, and if the sequence of random variables x_{in} converges in probability, (with probability one) to x_i which may be a random variable or a constant, then the sequence of random variables $\varphi(x_{1n}, \cdots, x_{pn})$ converges in probability (with probability one) to $\varphi(x_1, \cdots, x_p)$, where some or all of the x's may be constants. If $x_1, \cdots, x_p$ are constants then $\varphi(t_1, \cdots, t_p)$ need only be continuous in the neighborhood of $x_1, \cdots, x_p$ and Borel measurable.*

For a proof of part of this lemma which may be extended to yield the entire proof, see, Frechet, [9, p. 178].

Using Lemma V it is easy to see that the coefficients r_n of least squares equations converge with probability one to their β values, where the β value is obtained by substituting σ_{ij} for s_{ijn} in the expression for r_n assuming, of course, independent random vectors which have the same distribution functions.

Since problems in the analysis of variance may be interpreted as problems in least squares the above comments and Lemma V will generally make it possible, when determining limiting distributions, to consider the statistics to be functions of deviations from "true" mean functions rather than "sample" mean functions.

We shall discuss, briefly, four applications of these results.

(a). *The limiting distribution of the regression coefficient.* Let r_n, the "sample" regression coefficient, be defined by the equation

$$r_n = \frac{\sum_{\nu} x_{i\nu}x_{j\nu}}{\sum_{\nu} x_{i\nu}^2},$$

where $x_{i\nu}$ and $x_{j\nu}$ are deviations from arithmetic means. If the random vectors $(x_{i\nu}, x_{j\nu})$ are independently distributed for fixed i, j, with the same distribution functions, and if $E(x_{i\nu}) = E(x_{j\nu}) = 0$, $E(x_{i\nu}x_{j\nu}) = \sigma_{ij}$, then it follows from the strong law of large numbers that $\sum_{\nu} x_{i\nu}x_{j\nu}/n$ converges to σ_{ij} with probability one, and from the Laplace-Liapounoff theorem that $\sum_{\nu} x_{i\nu}x_{j\nu}/\sqrt{n}$ has a normal limiting distribution with mean σ_{ij} and variance $E\{x_{i\nu}x_{j\nu} - \sigma_{ij})^2\}$. Hence, by Lemma V, $\sqrt{n}\left(r_n - \frac{\sigma_{ij}}{\sigma_{ii}}\right)$ has a normal limiting distribution with mean zero and variance $\lim_{n\to\infty} E\left\{n\left(r_n - \frac{\sigma_{ij}}{\sigma_{ii}}\right)^2\right\}$ unless that limit does not exist.

If the $x_{i\nu}$ are not random variables then, in order to apply Corollary I with $p = 1$, it is necessary that

$$\lim_{n \to \infty} \frac{x_{im}}{(\sum_\nu x_{i\nu}^2)^{\frac{1}{2}}} = 0. \tag{4.1}$$

In that case, the limiting distribution of $(\sum_\nu x_{i\nu}^2)^{\frac{1}{2}} \cdot r_n$ is normal with zero mean and variance σ_{jj}. If (4.1) is not satisfied then there is no assurance, unless the $x_{j\nu}$ are normally distributed, that the limiting distribution of $(\sum_\nu x_{i\nu}^2)^{\frac{1}{2}} r_n$ is normal.

(*b*). *The limiting distribution of the analysis of variance ratio.* The tests of significance which occur in the analysis of variance depend on the ratio of two quadratic forms, q_{1n} and q_{2n}, the denominator q_{2n} having rank (or degrees of freedom) m_{2n} increasing with n, and the numerator q_{1n} having rank m_1 not changing with n, i.e.,

$$v_n = \frac{q_{1n} m_{2n}}{q_{2n} m_1},$$

where $q_{1n} + q_{2n} + q_{3n} = \sum_\nu x_\nu^2$ and q_{3n} is a quadratic form of rank m_{3n} which will be identically zero if $n = m_1 + m_{2n}$. Since[24] q_{2n} is expressible as the variance of x about a least squares equation it follows from the previous discussion and Lemma IV that $\frac{q_{2n}}{m_{2n}}$ converges with probability one to σ^2 under the assumptions that the x_ν are independently distributed with zero means and variances σ^2. Hence the limiting distribution of v_n will depend only on the limiting distribution of q_{1n} and it will consequently be necessary to consider only the matrix of q_{1n}, in order to apply Corollary VI with $p = 1$. For example,[25] if there are pn independently distributed random variables $x_{i\nu}$ with zero means and variances σ^2 arranged in p blocks of n random variables each, then

$$\sum_{i,\nu} (x_{i\nu} - \bar{x})^2 = n \sum_i (\bar{x}_{in} - \bar{x}_n)^2 + \sum_{i,\nu} (x_{i\nu} - \bar{x}_{in})^2,$$

where $\bar{x}_{in}$ is the arithmetic mean of $x_{i1}, \cdots, x_{in}$ and $\bar{x}_n$ is the arithmetic mean of all the $x_{i\nu}$. Then

$$q_{1n} = n \sum_i (\bar{x}_{in} - \bar{x}_n)^2,$$

$$q_{2n} = \sum_{i,\nu} (x_{i\nu} - \bar{x}_{in})^2,$$

$$m_1 = p - 1,$$

$$m_{2n} = p(n - 1)$$

[24] This has been proved by Kolodziejczyk, [12, p. 161].

[25] Other schemes are given in Fisher, [8].

and the matrix of q_{1n} may be obtained by substituting for the $\bar{x}_{in}$ and $\bar{x}_n$. In this case it is sufficient to express q_{1n} as $\sum_{i,j} a_{ij}S_iS_j$ where $S_i = \sum_{\nu} x_{i\nu}$, $a_{ii} = (p - 1)/pn$, and, f, $i \neq j$, $a_{ij} = -1/pn$, to see that the condition that the maximum of the absolute values of the elements of the matrix of q_{1n} approaches zero as n increases. Hence, if the $x_{i\nu}$ satisfy the condition $\mathfrak{L}$, the limiting distribution of $m_1 v_n$ is $G(v; p - 1, 1)$.

Clearly, if only the rank of q_{3n} increases as n increases, the rank m_{2n} of q_{2n} being constant and if the maximum of the absolute values of the elements of the matrix of q_{2n} also approaches zero as n increases, then v_n will have a limiting distribution which is the analysis of variance distribution, and the limiting distribution of $\frac{q_{1n}}{q_{1n} + q_{2n}}$ will be the correlation ratio distribution.

(c). *Periodogram analysis.* We need only remark that the linear functions which are used in the analysis of the Schuster periodogram[26] meet all the requirements of Corollary I if the x_ν are independently distributed with zero means and constant variances and satisfy the condition $\mathfrak{L}$. Consequently the large sample theory of the Schuster periodogram is the same for non-normal as it is for normal distributions.

(d). *Multivariate analysis.* We shall assume that the random vectors $X_1, \cdots, (X_\nu$ has components $x_{1\nu}, \cdots, x_{p\nu})$, are independently distributed, that (2.3) and (2.4) are satisfied, and that the condition $\mathfrak{L}_p$ is satisfied. For any fixed n and α we shall call the determinant D_α^n of the forms (3.5) a generalized sum of squares, and the determinant V_α^n of the elements $b_{ij\alpha}^n/m_\alpha$ a generalized variance. We shall say that D_β^n and V_β^n have rank m_β and that D_k^n and V_k^n have rank n_{kn}. If m_β is constant, and if (3.7) and (3.8) are true then clearly the limiting distribution of D_β^n is the distribution of the generalized variance of m_β vector observations[27] from a normal distribution, with zero means and covariance parameters σ_{ij}. Under the same conditions, the limiting distribution of D_β^n/V_k^n is the distribution of the generalized variance of m_β vector observations from a normal distribution with zero means and covariance parameters δ_{ij}. Many other similar limiting distributions are immediately derivable.

Before completing our discussion of the limiting distributions of statistics occurring in multivariate analysis, we shall state a theorem on limiting distributions which is an obvious generalization of a theorem of Doob, [4, p. 166].

Suppose that the random variables $g(n)X_{1n}, \cdots, g(n)X_{pn}$ have a distribution function $F(g(n)X_{1n}, \cdots, g(n)X_{pn})$ which is such that

$$\lim_{n\to\infty} F(g(n)X_{1n}, \cdots, g(n)X_{pn}) = F(X_1, \cdots, X_p),$$

where $F(X_1, \cdots, X_p)$ is a continuous distribution function, and suppose that X_{in} converges in probability to the real number ξ_i. For example, if $\bar{x}_n =$

[26] The theory of the Schuster periodogram is given by Fisher [7].

[27] See Wilks, [18, p. 476] or Madow, [15, pp. 481, 484].

$\sum_{\nu} x_\nu/n$ where $E(x_\nu) = 0$, $E(x_\nu^2) = 1$, and $\mathfrak{L}$ is satisfied, then $\bar{x}_n$ converges to zero with probability one, and $\sqrt{n}\,\bar{x}_n$ has a limiting distribution which is normal with zero mean and unit variance, i.e.

$$\lim_{n\to\infty} | P\{\sqrt{n}\bar{x}_n < x\} - N(X; 1) | = 0.$$

THEOREM VII. *Let $\varphi_f(t_1, \cdots, t_p)$ be a function of $t_1, \cdots, t_p$ defined in a neighborhood N of $\xi_1, \cdots, \xi_p$ which, together with its $(k_f + 1)$-th partial derivatives is continuous in N. Suppose that k is the least value of r_f such that the random variables*[28]

$$[g(n)]^{r_f}\left[\sum_i (x_{in} - \xi_i)\cdot\frac{\partial\varphi_f(\xi_1, \cdots, \xi_p)}{\partial\xi_i}\right]^{(r_f)}$$

have a joint limiting distribution function $D(x_1, \cdots, x_s)$. Then the random variables $[g(n)]^{k_f}[\varphi_f(x_{1n}, \cdots, x_{pn}) - \varphi_f(\xi_1, \cdots, \xi_p)]$ have a joint limiting distribution which is given by $D(x_1, \cdots, x_s)$. The value k_f is greater than or equal to the minimum value for which not all the partial derivatives of order k_f vanish at $\xi_1, \cdots, \xi_p$.

The proof is almost word for word that of Doob, the only difference being the removal of the specializing words.

We now consider the limiting distribution of the ratio of generalized sums of squares L_n which is defined by

$$L_n = \frac{D_k^n}{D_{k+1}^n},$$

where D_{k+1}^n is the determinant of the forms $b_{ijk}^n + b_{ij1}^n = b_{ij\,k+1}^n$. It has been shown that[29]

$$L_n = \prod_i \frac{Y_{ik}^n}{Y_{ik+1}^n},$$

where Y_{ij}^n, $(j = k, k + 1)$, is a ratio of generalized sums of squares

$$Y_{ij}^n = \frac{| b_{rsj}^n |}{| b_{uvj}^n |}, \qquad (r, s = 1, \cdots, i; u, v = 1, \cdots, i - 1; b_{00j}^n = 1).$$

Since Y_{ij}^n/m_{jn} converges with the probability one to $| \sigma_{rs} |/| \sigma_{uv} |$, and since, by Corollarv VIII the joint limiting distribution of the $m_{k+1\,n}\left(1 - \frac{Y_{ik}^n}{Y_{ik+1}^n}\right)$ is

[28] See Goursat-Hedrick, [10, p. 107] for a statement of the Taylor expansion of functions of several variables, which we use here, by $\frac{\partial\varphi_f(\xi_1, \dots, \xi_p)}{\partial\xi_i}$ is meant the value of $\frac{\partial\varphi_f(x_1, \dots, x_p)}{\partial x_i}$ at the point $\xi_1, \dots, \xi_p$.

[29] See Madow, [15, p. 485].

$\prod_i G(x_i\,; m, 1)$ it follows, by Theorem VII, that the joint limiting distribution of the ratios of generalized sums of squares

$$\prod_{h=1}^{i} \frac{Y_{hk}^n}{Y_{hk+1}^n},$$

is

$$\prod_i G(x_i\,; im_1\,, 1)$$

and that the limiting distribution of $m_{k+1\;n}(1 - L_n)$ is[30]

$$G(x; pm_1\,, 1).$$

In a following paper, these results will be extended to quadratic forms in non-central random variables.

5. Summary. In Section 2, Theorem I, we stated a very general form of the Laplace-Liapounoff theorem based on the Lindeberg condition. In four corollaries, this theorem was shown to provide joint limiting distributions for systems of linear forms which are such that the maximum of the absolute values of their coefficients converge to zero with an increase in the size of the sample if the coefficients are constants, and converge in probability to zero with an increase in the size of the sample if the coefficients are themselves random variables. It was shown that under certain conditions functions of several random variables, which are such that each function is a linear function of certain random variables for fixed values of random variables of lower index, also have a normal multivariate limiting distribution.

These results were extended to include limiting distributions of quadratic and bilinear forms in Section 3. The method of extension was to show that necessary and sufficient conditions for the existence of systems of linear forms satisfying the conditions of Section 2 are provided by rather simple conditions, the most important of which is that the greatest of the absolute values of the elements of the matrices of the quadratic and bilinear forms approach zero if the size of the sample increases, the ranks of the forms remaining unaltered. This led to the theorem that quadratic and bilinear forms having such matrices have χ^2, or covariance, or Wishart's distribution as limiting distributions. It was then shown, in Theorem IV, that if the rank of the sum of the matrices of the quadratic and bilinear forms is equal to the sum of the ranks of the matrices, and if certain of these ranks do not change as the size of the sample increases, then the system of quadratic and bilinear forms have Wishart's distribution in the limit provided the other conditions are met. These results

[30] A generalization of Wilks' result, [19, p. 323] to the case where the variates are not assumed to have a normal multivariate distribution may readily be obtained.

were then extended in Theorem V to one of the cases occurring when the coefficients of the forms are themselves random variables.

Several simple illustrations of the uses of the methods were given in Section 4. It was shown that the analysis of the variance ratios, and statistics occurring in the theory of multivariate statistical analysis have the same limiting distributions which they would have had if their variables had been normally and independently distributed.

REFERENCES

[1] M. S. Bartlett and J. Wishart, "The generalized product moment distribution in a normal system," *Proc. Camb. Phil. Soc.*, Vol. 29 (1933), pp. 260–270.

[2] W. G. Cochran, "The distribution of quadratic forms in a normal system, with applications to the analysis of covariance," *Proc. Camb. Phil. Soc.*, Vol. 30 (1934), pp. 178–191.

[3] H. Cramer, Random Variables and Probability Distributions," *Camb. Tracts in Math.* and *Math. Physics*, No. 36, London, 1937.

[4] J. L. Doob, "The limiting distributions of certain statistics," *Annals of Math. Stat.*, Vol. 6 (1935), pp. 160-169.

[5] J. L. Doob, "Stochastic processes with an integral-valued parameter," *Trans. Am. Math. Soc.*, Vol. 44 (1938), pp. 87-150.

[6] R. A. Fisher, "Applications of 'Student's' distribution," *Metron*, Vol. 5 (1926), pp. 90–104.

[7] R. A. Fisher, "Tests of significance in harmonic analysis," *Proc. Roy. Soc.*, (A), Vol. 125 (1929), pp. 54-60.

[8] R. A. Fisher, *Statistical Methods for Research Workers*, 7th ed., Oliver and Boyd, London, 1938.

[9] M. Frechet, *Recherches Theoriques Modernes sur la Theorie des Probabilites*, Vol. 1, Gauthier-Villars, Paris, 1937.

[10] E. Goursat, *A Course in Mathematical Analysis*, Vol. 1, translated by E. R. Hedrick, Ginn and Co., New York, 1904.

[11] A. Kolmogoroff, "Grundbegriffe der Wahrscheinlichkeitsrechnung," *Ergebnisse der Mathematik*, Vol. 2, no. 3.

[12] S. Kolodziejczyk, "On an important class of statistical hypotheses," *Biometrika*, Vol. 27 (1935), pp. 161-190.

[13] P. Levy, *Calcul des Probabilites*, Gauthier-Villars, Paris, 1925.

[14] P. Levy, *Theorie de l'Addition des Variables Aleatoires*, Gauthier-Villars, Paris, 1937.

[15] W. G. Madow, "Contributions to the theory of multivariate statistical analysis," *Trans. Am. Math. Soc.*, Vol. 44 (1938), pp. 454-495.

[16] W. G. Madow, "The distribution of quadratic forms in non-central normal random variables," *Annals of Math. Stat.*, Vol. 11 (1940), pp. 100–103.

[17] S. Saks, *Theory of the Integral*, 2nd ed., G. E. Stechert and Co., New York, 1937.

[18] S. S. Wilks, "Certain generalizations in the analysis of variance," *Biometrika*, Vol. 24 (1932), pp. 472–494.

[19] S. S. Wilks, "On the Independence of k sets of normally distributed statistical variables," *Econometrika*, Vol. 3 (1935), pp. 309–326.

[20] J. Wishart and M. S. Bartlett, See [1].

Washington, D. C.