An Elementary Proof of the Central Limit Theorem

By H. F. Trotter in Kingston (Canada)

Introduction. The proof referred to in the title is elementary in the sense that it avoids the use of characteristic functions. To a mathematician already familiar with Fourier analysis, the characteristic function is a natural tool, but to a student of probability or statistics, confronting a proof of the central limit theorem for the first time, it may appear as an ingenious but artificial device. Thus, although knowledge of characteristic functions remains indispensable for the study of general limit theorems, there may be some interest in an alternative way of attacking the basic normal approximation theorem.

Our method of proof is in principle the same as that used by Lindeberg [1]. We start with the elementary observation that a measure is determined by the values of the integrals of continuous functions with respect to it. In probabilistic language, the distribution of a random variable x is known if the expectations $E\{f(x)\}$ are known for sufficiently many continuous functions. Trying to evaluate $E\{f(x+y)\}$ when x and y are independent random variables leads one to consider the operators defined in section 1. The "product rule" and convergence criterion then follow readily. It should be noted that the proof of the convergence criterion follows a fairly obvious and intuitive line of reasoning. The corresponding result for characteristic functions is precisely the point of greatest technical difficulty in the usual development.

The reader will observe that the "$\varepsilon - \delta$" part of the proof, in sections 2 and 3, is essentially the same as in the conventional treatment. The basic idea is that of using a second-order Taylor series expansion, and the same techniques of estimation apply.

The proof has the interesting feature that the only facts about the normal distribution used in the argument are that it has a finite variance, and that the sum of independent normal variates is again normal. In particular, the explicit analytic form of the distribution plays no part.

The first three sections of the paper are intended to be self-contained, and require little more than acquaintance with the notions of random variable, distribution function, and mathematical expectation. The theory of Stieltjes integration of continuous functions (including the reduction of a double integral to repeated single integrals) is also assumed; if one is willing to restrict attention to random variables with continuous density functions, the arguments can be recast easily in terms of the Riemann integral.

We treat first the very simple case of identically distributed random variables, which illustrates the method of proof with a minimum of complicating detail. In section 3 we show how LINDEBERG's result (as formulated in [2]) can be obtained. Section 4 contains a brief sketch, without proofs, of how the theory can be extended to the multivariate case.

I wish to thank Professor F. J. ANSCOMBE for a number of helpful comments.

1. Preliminaries. Let C be the class of bounded uniformly continuous real-valued functions on $(-\infty, \infty)$. For each $f \in C$, define $\|f\|$ (the *norm* of f) to be $\sup_x |f(x)|$. It is obvious that if f and g are both in C then so is the function $f + g$, and

$$\|f + g\| \leqq \|f\| + \|g\| .$$

A function A, defined on the class C and taking values in C is called a *linear operator* on C if

$$A(af + bg) = aAf + bAg$$

for all $f, g \in C$ and all real numbers a, b. (It is customary to write Af instead of the ordinary functional notation $A(f)$.) If A and B are linear operators on C their sum, $A + B$, is defined by taking $(A + B)f = Af + Bf$ for all $f \in C$; similarly $A - B$ is defined by taking $(A - B)f = Af - Bf$. The product AB is defined by taking $ABf = A(Bf)$; two operators A and B are said to commute if $AB = BA$. It is easily checked that the sum, difference and product of two linear operators are again linear operators. An operator A with the property that $\|Af\| \leqq \|f\|$ for all $f \in C$ is called a *contraction* operator.

Let x be a random variable with distribution function V, so that for every real number y

$$Pr\{x \leqq y\} = V(y) .$$

If f is any function in C the mathematical expectation of $f(x)$ exists and

$$E\{f(x)\} = \int f(y)\, dV(y) .$$

We define a linear operator associated with x in the following way. For any $f \in C$, we define $T_x f$ by setting

$$(T_x f)(y) = E\{f(x + y)\} = \int f(x + y)\, dV(x)$$

for every real number y. Since V is a probability distribution function,

$$|(T_x f)(y)| \leqq \|f\|$$

for all y, and

$$|(T_x f)(y_1) - (T_x f)(y_2)| \leqq \sup_x |f(x + y_1) - f(x + y_2)|$$

for all y_1, y_2. The second inequality shows that the uniform continuity of f implies that of $T_x f$, while the first shows that $T_x f$ is bounded. Hence T_x is a function from C to C, and it is clear from the definition that it is linear. The first inequality actually shows that *T_x is a contraction operator*. Note that if two random variables have the same distribution, their associated operators are the same.

Now suppose that x_1 and x_2 are independent random variables with distribution functions V_1, V_2. By definition, $(T_{x_1+x_2}f)(y) = E\{f(x_1 + x_2 + y)\}$. Because x_1 and x_2 are independent,

$$\begin{aligned} E\{f(x_1 + x_2 + y)\} &= \int\int f(x_1 + x_2 + y)\, dV_1(x_1)\, dV_2(x_2) = \\ &= \int\Big[\int f(x_1 + x_2 + y)\, dV_1(x_1)\Big]\, dV_2(x_2) = \\ &= \int (T_{x_1}f)(x_2 + y)\, dV_2(x_2) = (T_{x_2}T_{x_1}f)(y)\,. \end{aligned}$$

Hence $T_{x_1+x_2} = T_{x_2}T_{x_1}$. Note that by inverting the order of integration above, we would come out with $T_{x_1}T_{x_2}$, which shows that *T_{x_1} and T_{x_2} commute.* By induction on the number of random variables involved we obtain the rule:

If $x_1, x_2, \ldots, x_n$ are independent random variables, then the operator $T_{x_1+x_2+\cdots+x_n}$ associated with their sum is the product of the operators $T_{x_1}, T_{x_2}, \ldots, T_{x_n}$.

A sequence of random variables $x_1, x_2, \ldots$ with distribution functions $V_1, V_2, \ldots$ is said to *converge in distribution* to a random variable x with distribution function V if $\lim\limits_{n\to\infty} V_n(y) = V(y)$ for every point y at which $V(y)$ is continuous. (Since a distribution function is monotone and continuous from the right, it is determined once its value is known at every point at which it is continuous.) In order to reduce the problem of convergence of a sequence of random variables to one involving their associated operators, we shall prove the following

Lemma. *A sufficient condition for a sequence of random variables $x_1, x_2, \ldots$ to converge in distribution to a random variable x is that*

$$\lim_{n\to\infty} \| T_{x_n}f - T_x f\| = 0$$

for every $f \in C$ whose first and second derivatives exist and are also in C.

(For brevity, we shall say that functions satisfying the preceding condition are in the class C^2.)

We first remark that the hypothesis of the lemma trivially implies that

$$\lim_{n\to\infty} (T_{x_n}f)(0) = (T_x f)(0)$$

for all $f \in C^2$, which, from the definition of T_x, simply means that

$$\lim_{n\to\infty} E\{f(x_n)\} = E\{f(x)\} \text{ for all } f \in C^2\,.$$

This (apparently) weaker condition is all that we shall actually use in the proof. The key which enables us to pass from expectations of continuous functions to distribution functions is the obvious fact that if f has the value 1 on a set A and the value 0 on a set B, and if $0 \leqq f(x) \leqq 1$ for all values of x, then

$$Pr\{x \in A\} \leqq E\{f(x)\} \leqq 1 - Pr\{x \in B\}\,.$$

Let $V, V_1, V_2, \ldots$ be the distribution functions of $x, x_1, x_2, \ldots$. Suppose that y is a point at which V is continuous. Let ε be any positive number, and take δ small enough that $V(y + \delta) - V(y - \delta) < \varepsilon$. Construct a pair of functions f, g in C^2 such that $0 \leqq f(x) \leqq g(x) \leqq 1$ for all x and $f(x) = 1$ for $x \leqq y - \delta$, $g(x) = 1$ for

$x \leqq y$, $f(x) = 0$ for $y \leqq x$, and $g(x) = 0$ for $y + \delta \leqq x$. (This is easily done; the reader may find it helpful to draw a picture.) Then

$$V(y - \delta) \leqq E\{f(x)\} \leqq V(y) \leqq E\{g(x)\} \leqq V(y + \delta),$$

and for each n

$$E\{f(x_n)\} \leqq V_n(y) \leqq E\{g(x_n)\}.$$

By hypothesis

$$E\{f(x)\} = \lim_{n\to\infty} E\{f(x_n)\}$$

and

$$E\{g(x)\} = \lim_{n\to\infty} E\{g(x_n)\},$$

so that

$$E\{f(x)\} \leqq \liminf_{n\to\infty} V_n(y) \leqq \limsup_{n\to\infty} V_n(y) \leqq E\{g(x)\}.$$

Hence

$$V(y) - \varepsilon \leqq \liminf_{n\to\infty} V_n(y) \leqq \limsup_{n\to\infty} V_n(y) \leqq V(y) + \varepsilon,$$

and since this is true for every positive ε, $\lim\limits_{n\to\infty} V_n(y) = V(y)$.

We shall use α to denote a standard normal variate; i.e., α is a random variable with distribution function

$$V(y) = (2\pi)^{-1/2} \int_{-\infty}^{y} e^{-u^2/2}\, du.$$

Then for any σ, $\sigma\alpha$ is a normally distributed random variable with mean zero and variance σ^2. Recall that if $\sigma_1\alpha_1$ and $\sigma_2\alpha_2$ are independent normal variates with mean zero, then their sum is normally distributed with mean zero and variance $\sigma_1{}^2 + \sigma_2{}^2$. This fundamental property of the normal distribution is all that will be needed in the following proofs. In terms of the related operators, it means that

$$T_{\sigma_1\alpha} T_{\sigma_2\alpha} = T_{\sigma\alpha}$$

where $\sigma = (\sigma_1{}^2 + \sigma_2{}^2)^{1/2}$.

2. The central limit theorem for identically distributed random variables. Suppose x is a random variable (with distribution function V) having mean 0 and variance 1, so that

$$\int x\, dV(x) = 0, \quad \int x^2\, dV(x) = 1.$$

Then one form of the central limit theorem may be stated as follows:

Theorem. *Let $x_1, x_2, \ldots$ be a sequence of independent random variables with the same distribution as x, and define $X_n = n^{-1/2}(x_1 + x_2 + \cdots + x_n)$. Then the random variables X_n converge in distribution to a standard normal variate.*

The proof uses the following lemma on linear operators. (Note that the hypotheses are satisfied for operators derived from random variables.)

Lemma. *Let A and B be two contraction operators which commute with each other. Then for any $f \in C$,*

$$\| A^n f - B^n f \| \leqq n \| Af - Bf \|.$$

To see this, observe that

$$A^n f - B^n f = \sum_{i=0}^{n-1} A^{n-i-1}(A - B) B^i f = \sum_{i=0}^{n-1} A^{n-i-1} B^i (A - B) f$$

since A and B commute. There are n terms on the right, and since A and B are contraction operators the norm of each term is less than or equal to $\|(A - B)f\|$. Since the norm of the sum is less than or equal to the sum of the norms, the desired conclusion follows.

According to the lemma of section 1, it will suffice to prove that

$$\lim_{n\to\infty} \| T_{X_n} f - T_\alpha f \| = 0$$

for every $f \in C^2$. Writing σ for $n^{-1/2}$, and recalling the rule derived in section 1, we have $T_{X_n} = T^n_{\sigma x}$. Also, because of the fundamental property of the normal distribution, $T_\alpha = T^n_{\sigma\alpha}$. Applying the preceding lemma on linear operators gives

$$\| T_{X_n} f - T_\alpha f \| \leqq n \, \| T_{\sigma x} f - T_{\sigma\alpha} f \| \, .$$

Thus we have reduced the problem to showing that $\lim\limits_{n\to\infty} n \, \| T_{\sigma x} f - T_{\sigma\alpha} f \| = 0$ for every $f \in C^2$.

For $f \in C^2$ we can obtain a Taylor series expansion in the form

$$\begin{aligned} f(x+y) &= f(y) + x f'(y) + \tfrac{1}{2} x^2 f''(\eta) = \\ &= f(y) + x f'(y) + \tfrac{1}{2} x^2 f''(y) + \tfrac{1}{2} x^2 (f''(\eta) - f''(y)) \end{aligned}$$

where η is some number between y and $y + x$. Now take $\varepsilon > 0$; since $f \in C^2$, f'' is uniformly continuous and we can find $\delta > 0$ so that $|f''(\eta) - f''(y)| < \varepsilon$ whenever $|\eta - y| < \delta$. We now calculate an approximation to $T_{\sigma x} f$.

$$\begin{aligned} (T_{\sigma x} f)(y) &= \int f(y + \sigma x)\, dV(x) = \\ &= f(y) \int dV(x) + f'(y) \sigma \int x \, dV(x) + \tfrac{1}{2} f''(y) \sigma^2 \int x^2 \, dV(x) + \\ &\qquad + \tfrac{1}{2} \sigma^2 \int (f''(\eta) - f''(y)) x^2 \, dV(x) \, . \end{aligned}$$

The first three terms reduce to $f(y) + \frac{1}{2} \sigma^2 f''(y)$ since V is the distribution function of a variate with mean 0 and variance 1. To estimate the last term, we split the integral into

$$\int_{|x| < \delta/\sigma} (f''(\eta) - f''(y)) x^2 \, dV(x) + \int_{|x| \geqq \delta/\sigma} (f''(\eta) - f''(y)) x^2 \, dV(x) \, .$$

For $|x| < \delta/\sigma$, $|\eta - y| \leqq |\sigma x| < \delta$, so the absolute value of the first integral is less than or equal to

$$\int_{|x| < \delta/\sigma} \varepsilon x^2 \, dV(x) \leqq \varepsilon \int x^2 \, dV(x) = \varepsilon \, .$$

For all values of η, $|f''(\eta) - f''(y)| \leqq 2 \| f'' \|$, and consequently the absolute value of the second integral is less than or equal to $2 \| f'' \| \int\limits_{|x| \geqq \delta/\sigma} x^2 \, dV(x)$.

Since $\int x^2 \, dV(x)$ is finite,

$$\lim_{k\to\infty} \int_{|x| \geqq k} x^2 \, dV(x) = 0 \, ;$$

hence the second integral will be less than ε if σ is sufficiently small (i.e., if n is sufficiently large). Thus we have shown that if n is sufficiently large,

$$\left|(T_{\sigma x}f)(y) - f(y) - \tfrac{1}{2}\sigma^2 f''(y)\right| \leqq \sigma^2\varepsilon = \varepsilon n^{-1}.$$

The preceding argument depended only on the fact that the random variable x had mean zero and variance 1. Since the standard normal variate also has mean zero and variance 1, a similar estimate for $(T_{\sigma\alpha}f)(y)$ is valid. Taking the difference between the two estimates gives

$$\left|(T_{\sigma x}f)(y) - (T_{\sigma\alpha}f)(y)\right| \leqq 2\varepsilon n^{-1},$$

or in other words,

$$n\,\|T_{\sigma x}f - T_{\sigma\alpha}f\| \leqq 2\varepsilon,$$

for sufficiently large values of n. This is true for any positive ε, and hence

$$\lim_{n\to\infty} n\,\|T_{\sigma x}f - T_{\sigma\alpha}f\| = 0.$$

3. Non-identically distributed random variables. Suppose that $x_1, x_2, \ldots$ is a sequence of independent random variables, with x_i having the distribution function V_i. We shall assume that each x_i has mean zero and a finite variance σ_i^2, so that

$$\int x\,dV_i(x) = 0 \quad \text{and} \quad \int x^2\,dV_i(x) = \sigma_i^2.$$

We set $s_n = \left(\sum_{i=1}^{n} \sigma_i^2\right)^{1/2}$. The sequence $x_1, x_2, \ldots$ is said to satisfy the *Lindeberg condition* if for every positive δ,

$$\lim_{n\to\infty} s_n^{-2} \sum_{i=1}^{n} \int_{|x| \geqq \delta s_n} x^2\,dV_i(x) = 0.$$

We shall prove the following version of the central limit theorem.

Let $x_1, x_2, \ldots$ be a sequence of independent random variables as described above, and define $X_n = s_n^{-1}(x_1 + x_2 + \cdots + x_n)$. Then if the Lindeberg condition is satisfied, the random variables X_n converge in distribution to a standard normal variate.

We first generalize the lemma used in the preceding section.

Let $A_1, A_2, \ldots, A_n, B_1, B_2, \ldots, B_n$ be contraction operators which commute with each other. Then for any $f \in C$,

$$\|A_1A_2\ldots A_n f - B_1B_2\ldots B_n f\| \leqq \sum_{i=1}^{n} \|A_i f - B_i f\|.$$

We have

$$A_1A_2\ldots A_n f - B_1B_2\ldots B_n f = \sum_{i=1}^{n} A_1A_2\ldots A_{i-1}(A_i - B_i)B_{i+1}B_{i+2}\ldots B_n f.$$

Just as in the special case considered earlier, since the operators commute and are contraction operators, the norm of the i-th term on the right-hand side is less than or equal to $\|A_i f - B_i f\|$. Summing over i yields the desired conclusion.

We need one additional lemma. Let z be a random variable with distribution function V, having mean 0 and variance 1. If we define $z_i = \sigma_i z$, then z_i has the same variance as x_i. *If the sequence $x_1, x_2, \ldots$ satisfies the Lindeberg condition, then so does* *the sequence* $z_1, z_2, \dots$. (We shall apply this lemma to the special case in which z is a standard normal variate.)

The proof proceeds in two stages. Let $k_n = \max_{i \leq n} \{\sigma_i s_n^{-1}\}$. We shall show that the Lindeberg condition for $x_1, x_2, \dots$ implies that $\lim_{n\to\infty} k_n = 0$, and that this in turn implies the condition for $z_1, z_2, \dots$.

Let $j \leqq n$ be the index for which $\sigma_j = k_n s_n$. We have

$$s_n^{-2} \sum_{i=1}^{n} \int_{|x| \geqq \delta s_n} x^2 \, dV_i(x) \geqq s_n^{-2} \int_{|x| \geqq \delta s_n} dV_j(x) = s_n^{-2} \left(\sigma_j^2 - \int_{|x| < \delta s_n} x^2 \, dV_j(x)\right) \geqq s_n^{-2} (\sigma_j^2 - \delta^2 s_n^2) = k_n^2 - \delta^2 .$$

By assumption, the left-hand side tends to 0 as $n \to \infty$ for every $\delta > 0$. Hence $\limsup_{n\to\infty} k_n^2 \leqq \delta^2$ for every positive δ, which implies $\lim_{n\to\infty} k_n = 0$.

Now if V_i' is the distribution function of z_i we have

$$\int_{|z_i| \geqq \delta s_n} z_i^2 \, dV_i'(z_i) = \int_{|z| \geqq \delta s_n \sigma_i^{-1}} \sigma_i^2 z^2 \, dV(z) \leqq \sigma_i^2 \int_{|z| \geqq \delta k_n^{-1}} z^2 \, dV(z) .$$

Adding over i and dividing by $s_n^2 = \sum_{i=1}^{n} \sigma_i^2$, we obtain

$$s_n^{-2} \sum_{i=1}^{n} \int_{|z_i| \geqq \delta s_n} z_i^2 \, dV_i'(z_i) \leqq \int_{|z| \geqq \delta k_n^{-1}} z^2 \, dV(z) .$$

For any fixed positive δ, the right-hand side tends to zero as $n \to \infty$ because $k_n \to 0$. Consequently, the left-hand side also tends to zero, which means that the Lindeberg condition is satisfied.

As in the case of identically distributed random variables, it will be enough to show that

$$\lim_{n\to\infty} \| T_{X_n} f - T_\alpha f \| = 0$$

for every $f \in C^2$, where α is a standard normal variate. We have

$$T_{X_n} = T_{x_1 s_n^{-1}} T_{x_2 s_n^{-1}} \dots T_{x_n s_n^{-1}}$$

and

$$T_\alpha = T_{\sigma_1 \alpha s_n^{-1}} T_{\sigma_2 \alpha s_n^{-1}} \dots T_{\sigma_n \alpha s_n^{-1}} .$$

Applying the lemma on operators, the theorem will be proved if we can show that

$$\lim_{n\to\infty} \sum_{i=1}^{n} \| T_{x_i s_n^{-1}} f - T_{\sigma_i \alpha s_n^{-1}} f \| = 0$$

for every $f \in C^2$.

Exactly the same calculations as were used in section 2 (after replacing σ by s_n^{-1} and V by V_i, and recalling that $\int x^2 \, dV_i(x) = \sigma_i^2$ instead of 1) give the result that for $f \in C^2$ and any $\varepsilon > 0$,

$$| T_{x_i s_n^{-1}} f(y) - f(y) - \tfrac{1}{2} \sigma_i^2 s_n^{-2} f''(y) | \leqq \tfrac{1}{2} \varepsilon \sigma_i^2 s_n^{-2} + \| f'' \| \, s_n^{-2} \int_{|x| \geqq \delta s_n^{-1}} x^2 \, dV_i(x)$$

provided that δ is sufficiently small. Similarly,

$$| T_{\sigma_i \alpha s_n^{-1}} f(y) - f(y) - \tfrac{1}{2} \sigma_i^2 s_n^{-2} f''(y) | \leqq \tfrac{1}{2} \varepsilon \sigma_i^2 s_n^{-2} + \| f'' \| s_n^{-2} \int\limits_{|x| \geqq \delta s_n^{-1}} x^2 \, dV_i'(x) \,,$$

where V_i' is the distribution function of $\sigma_i \alpha$. Thus we have

$$\| T_{x_i s_n^{-1}} f - T_{\sigma_i \alpha s_n^{-1}} f \| \leqq \varepsilon \sigma_i^2 s_n^{-2} + \| f'' \| (\xi_{i,n} + \xi_{i,n}')$$

where $\xi_{i,n}$ and $\xi_{i,n}'$ stand for

$$s_n^{-2} \int\limits_{|x| \geqq \delta s_n^{-1}} x^2 \, dV_i(x) \quad \text{and} \quad s_n^{-2} \int\limits_{|x| \geqq \delta s_n^{-1}} x^2 \, dV_i'(x)$$

respectively. Adding, we obtain

$$\sum_{i=1}^{n} \| T_{x_i s_n^{-1}} f - T_{\sigma_i \alpha s_n^{-1}} f \| \leqq \varepsilon + \| f'' \| \sum_{i=1}^{n} (\xi_{i,n} + \xi_{i,n}') \,.$$

By assumption the sequence $x_1, x_2, \ldots$ satisfies the Lindeberg condition, and by the lemma proved above, so does the sequence $\sigma_1 \alpha, \sigma_2 \alpha, \ldots$. This condition asserts that $\lim\limits_{n \to \infty} \sum\limits_{i=1}^{n} \xi_{i,n}'$ and $\lim\limits_{n \to \infty} \sum\limits_{i=1}^{n} \xi_{i,n}$ are both equal to zero. Hence

$$\limsup_{n \to \infty} \sum_{i=1}^{n} \| T_{x_i s_n^{-1}} f - T_{\sigma_i \alpha s_n^{-1}} f \| \leqq \varepsilon$$

for every positive ε, which shows that the limit must actually be 0.

4. The multivariate case. It is quite easy to extend the methods of this paper to prove various forms of multivariate central limit theorems. We shall indicate briefly how such an extension can be carried out.

Much of the discussion can be carried over with no formal change; it is only necessary to reinterpret the symbols. Fixing some integer m, we consider x and y as random variables ranging over the space of vectors with m coordinates. Now if $x \leqq y$ is interpreted to mean that each coordinate of x is less than or equal to the corresponding coordinate of y, distribution functions can be defined by the same formula as before. C becomes the class of bounded uniformly continuous real-valued functions of m real variables. A function f is in the class C^2 if it and all its first and second partial derivatives (including the mixed second partials) are in C. Now the operators associated with random variables can be defined just as before. The product rule and convergence criterion stated in section 1 continue to hold, and they can be proved by arguments essentially the same as those of section 1.

The *mean* of a random variable x is a vector, whose i-th coordinate, m_i, is $E\{x_i\}$ where x_i is the i-th coordinate of x. The *variance* of x is an $m \times m$ matrix whose ij-th entry is $E\{(x_i - m_i)(x_j - m_j)\}$. (It is necessarily symmetric and positive semi-definite.) Suppose that x has mean m and variance matrix S. Let us consider x and m as column vectors, and write x' and m' for the corresponding row vectors. Let p be a row vector and P an $m \times m$ matrix. Then $E\{px\} = pm$ and $E\{x' P x\} = m' P m + \mathit{Trace}(PS)$, as may be verified by a straightforward calculation.

For every positive semi-definite symmetric matrix S, there is a normal distribution with mean 0 and variance matrix S. These normal distributions have the basic property that the sum of two independent normal variates with zero means and variance matrices S_1 and S_2 is normally distributed with mean zero and variance matrix $S_1 + S_2$. (If S is singular, the corresponding normal distribution will be degenerate.)

In order to obtain a Taylor expansion in a form similar to that which we have used already, we must give the proper interpretation to the first and second derivatives of a function. We shall define $f'(y)$ to be the row vector whose i-th component is the partial derivative of f with respect to the i-th variable, evaluated at y. We consider $f''(y)$ to be a matrix, whose ij-th entry is the second partial derivative of f with respect to the i-th and j-th variables, evaluated at y. Then for any $f \in C^2$ there is an expansion

$$f(x+y) = f(y) + f'(y)x + \tfrac{1}{2} x' f''(\eta) x$$

where η is a vector lying on the line segment joining x and y.

The arguments of section 2 can now be translated almost word for word to give a proof of the following theorem.

Let x be an m-dimensional random variable with mean zero and variance matrix S, and let $x_1, x_2, \ldots$ be a sequence of independent random variables with the same distribution as x. Define $X_n = n^{-1/2}(x_1 + x_2 + \cdots + x_n)$. Then the random variables X_n converge in distribution to a normal variate with mean zero and variance matrix S.

References

[1] J. W. Lindeberg, Über das Exponentialgesetz in der Wahrscheinlichkeitsrechnung. Math. Z. **15**, 211–225 (1922).

[2] M. Loève, Probability Theory. New York 1955.

Eingegangen am 7.2.1959