A FREQUENCY-FUNCTION FORM OF THE CENTRAL LIMIT THEOREM

By WALTER L. SMITH

Received 15 December 1952

1. *Introduction.* The central limit theorem in the calculus of probability has been extensively studied in recent years. In its simplest form the theorem states that if $X_1, X_2, \ldots$ is a sequence of independent, identically distributed random variables of mean zero, then under general conditions the distribution function of $Z_n = (X_1 + \ldots + X_n)/\sqrt{n}$ converges as $n \to \infty$ to the normal or Gaussian distribution function. This form of the theorem in terms of *distribution functions* is the one required in statistical work, since it enables statements to be made about the limiting behaviour of prob $\{a < Z_n < b\}$.

When the random variables X_i have a frequency function $f(x)$, the frequency function of Z_n is determined, after a change of scale, by the n-fold convolution of $f(x)$ with itself. In certain analytical problems † and in statistical mechanics (4) it is required to approximate to, not prob $\{a < Z_n < b\}$, but the n-fold convolution of $f(x)$ with itself when n is large. For this we need a version of the central limit theorem in terms of *frequency functions*.

It is clear that a distribution-function form of the central limit theorem may be true while a frequency-function version is false. For example, if the distribution of X_i is discrete then at no stage does a frequency function of the Z_n exist, although the distribution function of the Z_n may tend to the normal distribution function.

Frequency-function versions of the central limit theorem have been proved previously by Khintchine (4) and by Cramér (1). The conditions of validity of their proofs are fairly strong; in particular, the X_i are assumed to have moments higher than the second. In the present paper the theorem is proved under less restrictive conditions making it more suitable to analytical applications. The general problem of deducing the convergence of frequency functions from the convergence of the corresponding distribution functions has recently been considered by Hammersley (3). His conditions refer, however, to the distribution functions of the Z_n and not to the individual distribution functions of the X_i.

2. *A class of distributions.* Let X be a random variable with distribution function $F(x)$. The characteristic function of X is given by the integral

$$\phi(t) = \int_{-\infty}^{+\infty} e^{itx}\, dF(x). \tag{1}$$

We shall say that a distribution (or its distribution function or frequency function) belongs to the class J if the associated characteristic function satisfies

$$|\phi(t)| \leqslant \frac{A}{|t|^{\alpha}} \quad \text{for all real } t, \quad |t| \geqslant R, \tag{2}$$

† One such application has been described in a paper by Dr D. R. Cox and myself (at present unpublished), 'A direct proof of a fundamental theorem of renewal theory'.

https://doi.org/10.1017/S0305004100028632 Published online by Cambridge University Press

for some positive A, R, α. Clearly the distribution function must be continuous, since a saltus would lead to a term in $\phi(t)$ of constant modulus for all t.

It follows from the Riemann-Lebesgue theorem for functions of bounded variation that if a distribution function is absolutely continuous, with a frequency function of bounded variation, then (2) will hold with $\alpha = 1$. This shows that a very large number of the common frequency functions belong to J, but there exist notable exceptions, such as the χ^2 distribution with one degree of freedom, which do not have frequency functions of bounded variation.

A theorem of Titchmarsh (6) after a little alteration informs us that if a frequency function is of the form

$$\mathfrak{f}(x) = \frac{\mathfrak{g}(x)}{|x|^{\beta}}, \tag{3}$$

where $0 < \beta < 1$ and $\mathfrak{g}(x)$ is of bounded variation in $(-\infty, +\infty)$, then (2) is true for $\alpha = 1-\beta$. There is obviously no need for the singularity of $\mathfrak{f}(x)$ to be at the origin, and whenever a frequency function can be expressed in the form

$$\mathfrak{f}(x) = \mathfrak{g}_0(x) + \sum_1^N \frac{\mathfrak{g}_r(x)}{|x-b_r|^{\beta_r}}, \tag{4}$$

where $\mathfrak{g}_r(x)$ for $r = 0, 1, \ldots, N$ are of bounded variation in $(-\infty, +\infty)$ and $0 < \beta_r < 1$, then it will be of the class J. Thus the class J is seen to include every common frequency function and many uncommon ones.

We next show that if a frequency function satisfies a certain Lipschitz type of condition (which, incidentally, is satisfied by the frequency functions (4)), then it will be of the class J.

A sufficient condition for a frequency function $\mathfrak{f}(x)$ *to belong to the class* J *is that as* $|h| \to 0$,

$$\int_{-\infty}^{+\infty} |\mathfrak{f}(x+h) - \mathfrak{f}(x)|\, dx = O(|h|^{\alpha}).$$

Proof. If $\phi(t)$ is the Fourier transform of $\mathfrak{f}(x)$ then $e^{-i\pi}\phi(t)$ is the Fourier transform of $\mathfrak{f}(x+\pi/t)$. Thus

$$(e^{-i\pi} - 1)\,\phi(t) = \int_{-\infty}^{+\infty} e^{ixt}\{\mathfrak{f}(x+\pi/t) - \mathfrak{f}(x)\}\, dx.$$

Hence

$$|\phi(t)| \leqslant \frac{1}{2}\int_{-\infty}^{+\infty} |\mathfrak{f}(x+\pi/t) - \mathfrak{f}(x)|\, dx = O(|t|^{-\alpha}).$$

The subclass of J consisting of distributions possessing a variance we call J'', and for this subclass $\phi(t)$ and its first two derivatives are bounded and continuous. We shall write B for the first absolute moment, which is an upper bound of $|\phi'(t)|$. An upper bound of $|\phi''(t)|$ is easily seen to be the variance.

3. *The central limit theorem in* J''. Throughout this section we shall be studying a sequence of independent, identically distributed, random variables $X_1, X_2, \ldots$ whose distribution function $F(x)$ is assumed (except in Lemma 2) to be of the class J''. We assume throughout this section, without loss of generality, that the mean of the X_i is zero and the variance unity. Define a new sequence of random variables by the equations

$$Z_n = \sum_{i=1}^{n} X_i/\sqrt{n}. \tag{5}$$

https://doi.org/10.1017/S0305004100028632 Published online by Cambridge University Press

We denote the frequency function of Z_n, when it exists, by $\mathfrak{h}_n(x)$ and the corresponding characteristic function by $\theta_n(t)$.

It follows in the usual way that

$$\theta_n(t) = \{\phi(t/\sqrt{n})\}^n. \tag{6}$$

Since the first two derivatives of $\phi(t)$ exist we also have

$$\theta_n''(t) = (n-1)\{\phi'(t/\sqrt{n})\}^2\{\phi(t/\sqrt{n})\}^{n-2} + \phi''(t/\sqrt{n})\{\phi(t/\sqrt{n})\}^{n-1}. \tag{7}$$

LEMMA 1. *For $n > 2 + 2/\alpha$, $\mathfrak{h}_n(x)$ exists and $\mathfrak{h}_n(x)$, $x^2\mathfrak{h}_n(x)$, are given by the absolutely convergent integrals*

$$\mathfrak{h}_n(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-itx}\theta_n(t)\,dt, \quad x^2\mathfrak{h}_n(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-itx}\theta_n''(t)\,dt, \tag{8}$$

and are consequently continuous and bounded.

For $|t| > R$ we have the following inequalities:

$$|\theta_n(t)| \leqslant \frac{A^n n^{\frac{1}{2}\alpha n}}{|t|^{n\alpha}} \leqslant \frac{A^n n^{\frac{1}{2}\alpha n}}{|t|^2}, \tag{9}$$

$$|\theta_n''(t)| \leqslant \frac{(n-1)B^2 n^{\frac{1}{2}(n-2)\alpha}A^{n-2}}{|t|^{(n-2)\alpha}} + \frac{A^{n-1}n^{\frac{1}{2}(n-1)\alpha}}{|t|^{(n-1)\alpha}} = O(|t|^{-2}). \tag{10}$$

Since $\theta_n(t)$ and $\theta_n''(t)$ are bounded for all t it follows from (9) and (10) that they are $\mathscr{L}(-\infty, +\infty)$. Thus the integrals (8) are absolutely convergent and the inversions (8) hold. This incidentally proves the existence of $\mathfrak{h}_n(x)$ for $n > 2 + 2/\alpha$. The continuity and boundedness of $\mathfrak{h}_n(x)$ and $x^2\mathfrak{h}_n(x)$ follow in the usual way and, in fact, by the Riemann-Lebesgue theorem, these two functions tend to zero as $|x| \to \infty$. We shall see later that the bounds on $\mathfrak{h}_n(x)$ and $x^2\mathfrak{h}_n(x)$ are independent of n.

LEMMA 2. *For all $F(x)$ having a variance,*

$$\lim_{n\to\infty} \theta_n(t) = e^{-\frac{1}{2}t^2}$$

and

$$\lim_{n\to\infty} \theta_n''(t) = (t^2 - 1)e^{-\frac{1}{2}t^2}$$

for all t.

The proof that $\theta_n(t) \to e^{-\frac{1}{2}t^2}$ follows from the usual argument adopted to prove the distribution-function form of the central limit theorem for identically distributed X_i (see, for example, (2)). It involves recognizing that

$$\phi(t) = 1 - \tfrac{1}{2}t^2 + o(t^2) \tag{11}$$

by Taylor's theorem.

Since $\phi(0) = 1$ and $\phi(t)$ is continuous we have, from (7),

$$\lim_{n\to\infty} \theta_n''(t) = e^{-\frac{1}{2}t^2} \lim_{n\to\infty} [(n-1)\{\phi'(t/\sqrt{n})\}^2 + \phi''(t/\sqrt{n})]. \tag{12}$$

The variance of $F(x)$ exists and equals unity, so that $\phi''(t)$ is continuous at $t = 0$ and $\phi''(0) = -1$. Thus

$$\lim_{n\to\infty} \phi''(t/\sqrt{n}) = -1. \tag{13}$$

Further, $\phi'(t)$ is differentiable, and since the mean is zero, $\phi'(0) = 0$. Thus

$$\phi'\left(\frac{t}{\sqrt{n}}\right) = \frac{t}{\sqrt{n}}\phi''\left(\vartheta\frac{t}{\sqrt{n}}\right) \quad (0 < \vartheta < 1), \tag{14}$$

https://doi.org/10.1017/S0305004100028632 Published online by Cambridge University Press

so that
$$\lim_{n\to\infty} \sqrt{n}\,\phi'\left(\frac{t}{\sqrt{n}}\right) = -t. \tag{15}$$

The lemma follows from (12), (13) and (15).

THEOREM 1. *Subject to the notations and assumptions given above,*

$$\lim_{n\to\infty} |x|^m \mathfrak{h}_n(x) = \frac{|x|^m}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2}$$

for $0 \leqslant m \leqslant 2$ *uniformly in* x.

It is sufficient to prove the theorem for the cases $m = 0$ and $m = 2$. The proof consists in showing that there exists a function which is $\mathscr{L}(-\infty, +\infty)$ and which dominates $\theta_n(t)$ and $\theta_n''(t)$ for all $n \geqslant n_0 = 2 + 4/\alpha$. Without loss of generality we may assume $R > A^{1/\alpha}\sqrt{e}$. We take a fixed t such that $|t| > R\sqrt{n_0}$ and write $\rho = |t|/\sqrt{n}$. The dominating function is constructed in two stages.

(i) $n_0 \leqslant n \leqslant t^2/R^2$. By the property of the characteristic functions of distributions of the class J we have

$$(n-1)\left|\phi\left(\frac{t}{\sqrt{n}}\right)\right|^{n-2} \leqslant \left\{\frac{|t|^2}{\rho^2} - 1\right\}\left\{\frac{A^{1/\alpha}}{\rho}\right\}^{\alpha(|t|^2/\rho^2-2)}.$$

The function x^{-a/x^2} is strictly increasing for $x > \sqrt{e}$ if $a > 0$. Thus

$$(n-1)\left|\phi\left(\frac{t}{\sqrt{n}}\right)\right|^{n-2} \leqslant \left\{\frac{|t|^2}{R^2} - 1\right\}\left\{\frac{A^{1/\alpha}\sqrt{n_0}}{|t|}\right\}^{4+2\alpha}\left\{\frac{|t|}{A^{1/\alpha}\sqrt{n_0}}\right\}^{2\alpha} = O\left\{\frac{1}{|t|^2}\right\}.$$

This proves the existence of a function which is integrable over the set $|t| > R\sqrt{n_0}$ and which dominates $(n-1)\{\phi(t/\sqrt{n})\}^{n-2}$ over this set if $n_0 \leqslant n \leqslant |t|^2/R^2$. From this result and equations (6) and (7) we see that, since $\phi(t)$, $\phi'(t)$ and $\phi''(t)$ are bounded for all t, there exists a function which is integrable over $|t| > R\sqrt{n_0}$ and which dominates $\theta_n(t)$ and $\theta_n''(t)$ in this set for $n_0 \leqslant n \leqslant |t|^2/R^2$.

(ii) $t^2/R^2 \leqslant n < \infty$. We use at this point a lemma due to Cramér ((2), p. 26).

LEMMA 3. *If* $\psi(t)$ *is a characteristic function such that* $|\psi(t)| \leqslant k < 1$ *for all* $|t| \geqslant R$, *then for* $|t| < R$,

$$|\psi(t)| \leqslant 1 - \frac{(1-k^2)}{8R^2}t^2.$$

We can certainly find a k such that the conditions of Cramér's lemma are satisfied by $\phi(t)$. It follows that we can find a positive γ such that

$$|\phi(t)| < e^{-\gamma t^2} \quad \text{if} \quad |t| \leqslant R.$$

Thus, if $\rho \leqslant R$,

$$|\theta_n(t)| < e^{-\gamma t^2} \tag{16}$$

and
$$|\phi(t/\sqrt{n})|^{n-2} < e^{-\gamma(1-2/n_0)t^2}. \tag{17}$$

Further, $|\phi''(t)|$ is bounded by unity for all t, and hence, by (14), $|\sqrt{n}\,\phi'(t/\sqrt{n})/t|$ is also bounded by unity. It is thus apparent from (17) and (7) that constants E and F may be found such that

$$|\theta_n''(t)| < (Et^2 + F)\,e^{-\gamma(1-2/n_0)t^2} \quad (0 < \rho \leqslant R). \tag{18}$$

From (16) and (18) we see that $\theta_n(t)$ and $\theta_n''(t)$ are both dominated by functions of the class $\mathscr{L}(-\infty, +\infty)$ for the range $n > |t|^2/R^2$.

https://doi.org/10.1017/S0305004100028632 Published online by Cambridge University Press

Combining the results of (i) and (ii) we have that $\theta_n(t)$ and $\theta_n''(t)$ are dominated in the set $|t| > R\sqrt{n_0}$ by an integrable function (obtained by summing the functions obtained at each stage). In the set $|t| \leqslant R\sqrt{n_0}$ the functions $|\theta_n(t)|$ and $|\theta_n''(t)|$ are bounded by unity. Thus the convergence proved in Lemma 2 is dominated by a function of $\mathscr{L}(-\infty, +\infty)$. Therefore by Lebesgue's theorem on dominated convergence the inversion integrals (8) converge to the limits given in the statement of Theorem 1. It is easy to prove that the convergence is uniform. For example, for all x,

$$\left| x^2 h_n(x) - \frac{x^2 e^{-\frac{1}{2}x^2}}{\sqrt{(2\pi)}} \right| \leqslant \int_{-\infty}^{+\infty} |\theta_n''(t) - (t^2-1)e^{-\frac{1}{2}t^2}|\, dt,$$

and since the integrand $|\theta_n''(t) - (t^2-1)e^{-\frac{1}{2}t^2}|$ converges to zero and is dominated in $\mathscr{L}_1$-norm, the integral may be made arbitrarily small, uniformly in x.

If, for integral r (>0), we write $J^{(r)}$ for the subclass of J in which the distributions possess absolute moments of the first r orders, then by methods which are essentially similar to those used in the proof of Theorem 1 we can prove the following generalization.

THEOREM 2. *If $F(x)$ belongs to $J^{(r)}$ ($r \geqslant 2$) (and has zero mean and unit variance) then*

$$\lim_{n\to\infty} |x|^m \mathfrak{h}_n(x) = \frac{|x|^m}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2},$$

for $0 \leqslant m \leqslant r$, uniformly in x.

It may be remarked here that the existence of a moment of even order is sufficient to ensure the existence of absolute moments of all lower orders.

We draw attention to the fact that the methods used in this section may easily be extended to prove that if k is any arbitrary integer, however large, then for all sufficiently large n, $\mathfrak{h}_n(x)$ possesses derivatives up to and including the kth, and all these derivatives possess the appropriate limiting properties (i.e. tend, uniformly, etc., to the corresponding derivatives of the normal density function).

4. *A frequency function form of Lindeberg's theorem.* In § 3 we considered identically distributed random variables, and we now consider a sequence of independent random variables $X_1, X_2, \ldots$ with distribution functions $F_1(x), F_2(x), \ldots$ which are not necessarily identical. The means of the X_i are assumed to be zero (without loss of generality) and the variances are written $\sigma_1^2, \sigma_2^2, \ldots$. We write

$$s_n^2 = \sum_{i=1}^{n} \sigma_i^2$$

and form the new sequence of random variables

$$Z_n = \sum_{i=1}^{n} X_i / s_n.$$

Let $H_n(x)$ be the distribution function of the Z_n. Then the sufficiency part of Lindeberg's theorem ((2), p. 56) states that: *If as $n\to\infty$, $s_n\to\infty$ and $\sigma_n/s_n\to 0$, and if for any $\epsilon > 0$,*

$$\lim_{n\to\infty} \sum_{\nu=1}^{n} \frac{1}{s_n^2} \int_{|x| > \epsilon s_n} x^2 dF_\nu(x) = 0,$$

then

$$\lim_{n\to\infty} H_n(x) = \Phi(x)$$

for all x, where $\Phi(x)$ is the normal distribution function.

https://doi.org/10.1017/S0305004100028632 Published online by Cambridge University Press

In this section we shall obtain a fairly simple and intuitively reasonable condition under which a frequency-function version of Lindeberg's theorem will be true; the derivation of any wider sufficient conditions by the present methods seems difficult.

Let the characteristic functions of $X_1, X_2, \ldots$ be $\phi_1(t), \phi_2(t), \ldots$. We shall say that the sequence of random variables contains a *smoothing subsequence* if there exists a subsequence $\phi_{n_1}(t), \phi_{n_2}(t), \ldots$ such that for some positive A, R, α,

$$|\phi_{n_i}(t)| \leqslant A/|t|^\alpha \quad \text{for all real } t \quad (|t| \geqslant R),$$

where we again may take $R > A^{1/\alpha}\sqrt{e}$. The idea of a smoothing subsequence arises naturally from the work in § 3.

LEMMA 4. *There exists a positive function* $\Omega(t)$ *such that*

$$\begin{aligned} \Omega(t) &= A/|t|^\alpha \quad \textit{for} \quad |t| \geqslant R \\ &= 1-\gamma t^2 \quad \textit{for} \quad |t| < R, \end{aligned}$$

where γ *is some positive number, and if* $\phi_{n_i}(t)$ *belongs to a smoothing subsequence, then for all* t

$$|\phi_{n_i}(t)| \leqslant \Omega(t).$$

We make use of Lemma 3. For any $\phi_{n_i}(t)$ of the smoothing subsequence we have $|\phi_{n_i}(t)| < A/R^\alpha < 1$, if $|t| \geqslant R$. Lemma 3 will thus supply a positive function bounding all $|\phi_{n_i}(t)|$ in the interval $|t| < R$, and in this range we have taken $\Omega(t)$ to be this bounding function. For $|t| \geqslant R$ it is clear that we may take $\Omega(t) = A/|t|^\alpha$.

It follows from Lemma 4 that the variances of the members of a smoothing subsequence have a positive lower bound.

We next state a sufficient condition for a frequency-function form of Lindeberg's theorem to hold. Let n^* be the number of members of the smoothing subsequence in $X_1, X_2, \ldots, X_n$.

CONDITION A. $$\varliminf_{n\to\infty} n^*/s_n^2 > 0.$$

We may regard Condition A as expressing the intuitive idea that for a unit increase in the variance the average number of 'injections of smoothness' must not be too small.

THEOREM 3. *If the conditions of Lindeberg's theorem hold and if a smoothing subsequence exists which satisfies Condition* A, *then for all sufficiently large* n, $H_n(x)$ *is absolutely continuous with a frequency function* $\mathfrak{h}_n(x)$ *such that*

$$\lim_{n\to\infty} \mathfrak{h}_n(x) = \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2}$$

uniformly in x.

The characteristic function of $H_n(x)$ is

$$\theta_n(t) = \prod_{j=1}^{n} \phi_j\left(\frac{t}{s_n}\right). \tag{19}$$

Thus $$|\theta_n(t)| \leqslant \left\{\Omega\left(\frac{t}{s_n}\right)\right\}^{n^*}.$$

But by Condition A there exists a $\mu > 0$ such that for all sufficiently large n, $n^* > \mu s_n^2$. Since $|\Omega(t)| \leqslant 1$ for all t it follows that

$$|\theta_n(t)| < \{\Omega(t/s_n)\}^{\mu s_n^2}.$$

https://doi.org/10.1017/S0305004100028632 Published online by Cambridge University Press

The function $\Omega(t)$ possesses all the properties of the characteristic function $\phi(t)$ of § 3 which were used to derive bounds and dominating functions. It follows from the distribution-function form of Lindeberg's theorem that

$$\theta_n(t) \to e^{-\frac{1}{2}t^2} \quad \text{as} \quad n \to \infty,$$

and we have shown therefore that if Condition A is satisfied the convergence is dominated. Theorem 3 may thus be proved by an appeal to Lebesgue's theorem such as was made in the proof of Theorem 1.

To prove the result, analytically more useful, that $x^2\mathfrak{h}_n(x)$ possesses the appropriate limiting property, two additional difficulties must be overcome. We must first show that $\theta_n''(t)$ tends to the correct limit and then show that the convergence is dominated. The first part is easy to prove by the next lemma. The second part needs an additional condition to be satisfied.

LEMMA 5. *Let $\{\mathfrak{f}_n(x)\}$ be a sequence of frequency functions with zero mean and unit variance and suppose that the sequence tends uniformly to a frequency function $\mathfrak{f}(x)$ (which has also zero mean and unit variance). If $\{\kappa_n(t)\}$ is the sequence of characteristic functions corresponding to $\{\mathfrak{f}_n(x)\}$ and $\kappa(t)$ is the characteristic function of $\mathfrak{f}(x)$, then $\kappa_n''(t)$ tends to $\kappa''(t)$ as $n \to \infty$.*

Given $\epsilon > 0$ there exists $R = R(\epsilon)$ such that

$$\int_{-R}^{+R} x^2\mathfrak{f}(x)\,dx > 1 - \tfrac{1}{2}\epsilon.$$

Having fixed R there exists $n_0(\epsilon)$ such that for all $n > n_0(\epsilon)$

$$|\,\mathfrak{f}_n(x) - \mathfrak{f}(x)\,| < \frac{3}{4}\frac{\epsilon}{R^3}.$$

Thus
$$\int_{-R}^{+R} x^2\mathfrak{f}_n(x)\,dx > \int_{-R}^{+R} x^2\mathfrak{f}(x)\,dx - \frac{3}{4}\frac{\epsilon}{R^3}\int_{-R}^{+R} x^2\,dx$$
$$> 1 - \tfrac{1}{2}\epsilon - \tfrac{1}{2}\epsilon = 1 - \epsilon.$$

But
$$\int_{-\infty}^{+\infty} x^2\mathfrak{f}_n(x)\,dx = 1.$$

Thus
$$\int_{|x|>R} x^2\mathfrak{f}_n(x)\,dx < \epsilon$$

for all $n > n_0(\epsilon)$.

It is apparent that the second derivatives of $\kappa_n(t)$ and $\kappa(t)$ must exist since all the variances are finite. We have

$$\kappa_n''(t) - \kappa''(t) = -\int_{-\infty}^{+\infty} e^{itx}x^2\{\mathfrak{f}_n(x) - \mathfrak{f}(x)\}\,dx.$$

Thus for $n > n_0(\epsilon)$,

$$|\,\kappa_n''(t) - \kappa''(t)\,| \leqslant \int_{-R}^{+R} x^2\,|\,\mathfrak{f}_n(x) - \mathfrak{f}(x)\,|\,dx + \tfrac{1}{2}\epsilon + \epsilon$$
$$\leqslant \int_{-R}^{+R} x^2\left\{\frac{3}{4}\frac{\epsilon}{R^3}\right\}dx + \tfrac{3}{2}\epsilon$$
$$\leqslant \tfrac{1}{2}\epsilon + \tfrac{3}{2}\epsilon = 2\epsilon.$$

https://doi.org/10.1017/S0305004100028632 Published online by Cambridge University Press

This proves the lemma. Before we prove that the convergence of $\theta_n''(t)$ is dominated we state

CONDITION B. $$\overline{\lim_{n\to\infty}} \sum_{\nu=1}^{n} \frac{1}{s_n^2} \int_{-\infty}^{+\infty} |x| \, dF_\nu(x) < \infty.$$

Condition B is a weak condition to prevent s_n^2 from increasing too slowly. Whenever $\overline{\lim}\, \sigma_i^2 > 0$ it is easy to show that Condition B is satisfied.

It is clear that if Condition B is satisfied,

$$\frac{1}{s_n^2} \sum_{\nu=1}^{n} \left| \phi_\nu'\left(\frac{t}{s_n}\right) \right| \leqslant \sum_{\nu=1}^{n} \frac{1}{s_n^2} \int_{-\infty}^{+\infty} |x| \, dF_\nu(x) \leqslant C$$

for all sufficiently large n, where C is some finite constant. Hence

$$\sum_{i\neq j}^{n} \sum^{n} \left| \phi_i'\left(\frac{t}{s_n}\right) \right| \left| \phi_j'\left(\frac{t}{s_n}\right) \right| \leqslant \left\{ \sum_{i=1}^{n} \left| \phi'\left(\frac{t}{s_n}\right) \right| \right\}^2 \leqslant C^2 s_n^4.$$

By differentiating (19) twice we obtain

$$s_n^2 \theta_n''(t) = \sum_{i=1}^{n} \phi_i''\left(\frac{t}{s_n}\right) \prod_{j\neq i} \phi_j\left(\frac{t}{s_n}\right) + \sum_{i\neq j}^{n} \phi_i'\left(\frac{t}{s_n}\right) \phi_j'\left(\frac{t}{s_n}\right) \prod_{k\neq i,j}^{n} \phi_k\left(\frac{t}{s_n}\right).$$

If we assume the conditions of Theorem 3 to hold, we have therefore for n sufficiently large,

$$\begin{aligned} s_n^2 \, |\, \theta_n''(t) \,| &\leqslant \left\{ \Omega\left(\frac{t}{s_n}\right) \right\}^{n^*-1} \sum_{i=1}^{n} \sigma_i^2 + \left\{ \Omega\left(\frac{t}{s_n}\right) \right\}^{n^*-2} \sum_{i\neq j}^{n} \left| \phi_i'\left(\frac{t}{s_n}\right) \right| \left| \phi_j'\left(\frac{t}{s_n}\right) \right| \\ &\leqslant s_n^2 \left\{ \Omega\left(\frac{t}{s_n}\right) \right\}^{n^*-1} + C^2 s_n^4 \left\{ \Omega\left(\frac{t}{s_n}\right) \right\}^{n^*-2}. \end{aligned}$$

Thus $$|\, \theta_n''(t) \,| \leqslant (1 + C^2 s_n^2) \left\{ \Omega\left(\frac{t}{s_n}\right) \right\}^{\mu s_n^2 - 2}. \tag{20}$$

As $s_n \to \infty$ the function on the right-hand side of (20) can be shown to be dominated in $\mathscr{L}_1$-norm by the method used for $(n-1) \, |\, \phi(t/\sqrt{n}) \,|^{n-2}$ in § 3.

THEOREM 4. *If, in addition to the conditions of Theorem* 3, *Condition* B *is satisfied, then*

$$\lim_{n\to\infty} x^2 \mathfrak{h}_n(x) = \frac{x^2}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2},$$

uniformly in x.

Proof. We have shown in Lemma 5 that $\theta_n''(t)$ tends to the correct limit, and this convergence has just been shown to be dominated in $\mathscr{L}_1$-norm. Theorem 4 follows on the lines of the previous theorems by an application of Lebesgue's convergence theorem.

5. *More general classes of distributions.* The distribution-function form of the central limit theorem holds under very general conditions. In proving the frequency-function form of the theorem we have relied on distributions of the class J to provide the 'system' with sufficient 'smoothness'. Intuitively it seems that even if most of the 'probability mass' in the basic distribution $F(x)$ were assigned to one or two points (or to a set of measure zero), eventually one could regard Z_n as being distributed in a 'smooth' frequency distribution if n were large enough and there were some 'smoothness' in the system to begin with.

https://doi.org/10.1017/S0305004100028632 Published online by Cambridge University Press

Our interest in this section is with the case of identically distributed X_i which has already been treated in § 3. The difficulty with distributions which are not absolutely continuous is that at no finite stage is the distribution function $H_n(x)$ of the Z_n (defined by (5)) necessarily absolutely continuous. While $H_n(x)$ must have a derivative almost everywhere, and we call the derivative $\mathfrak{h}_n(x)$ where it exists, nevertheless $\mathfrak{h}_n(x)$ will not necessarily be a true frequency function, since its integral over any set may not be the probability measure of that set. We seek a way of expressing the intuitive notions discussed above by means of the following definition.

DEFINITION. *Suppose there exist two sequences of positive numbers* $\{p_n\}$, $\{q_n\}$ *such that* $p_n+q_n=1$ *for all* n. *Suppose further that there exist three sequences of distribution functions* $\{P_n(x)\}$, $\{Q_n(x)\}$, $\{H_n(x)\}$, *such that*

$$H_n(x) = p_n P_n(x) + q_n Q_n(x).$$

If the sequence $\{P_n(x)\}$ *possesses some limiting property* Ω, *and if* $\Sigma q_n < \infty$, *we shall say that the sequence* $\{H_n(x)\}$ *possesses the property* Ω *almost certainly.*

The expression 'almost certain' has previously been used solely in relation to random variables. There is a good reason for using it to describe the sequence of distributions which govern these random variables. A sample Z_n drawn from $H_n(x)$ can be considered as being drawn from the distribution function $P_n(x)$ with probability p_n or from $Q_n(x)$ with probability q_n. But in any realized sequence of the Z_n, since Σq_n converges, we know that with probability one all but a finite number of the Z_n will be derived from the sequence $\{P_n(x)\}$. This follows from the Borel-Cantelli lemma. In other words, with probability one there is in any realized sequence of the Z_n, a finite n_0 such that for all n greater than n_0 the Z_n arise from the $\{P_n(x)\}$. Successive values of the Z_n may, of course, be correlated.

The next theorem attempts to sum up the intuitive ideas we have mentioned. The notation is that of § 3.

THEOREM 5. *If* $F(x)$ *is a distribution function with zero mean and unit variance, and if in some interval, however small,* $F'(x)$ *equals almost everywhere a positive function of bounded variation, then almost certainly*

$$|x|^m \mathfrak{h}_n(x) \to \frac{|x|^m}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2}$$

for $0 \leqslant m \leqslant 2$ *uniformly in* x.†

Since $F(x)$ is bounded and monotonic it must have a derivative almost everywhere. Thus, almost everywhere in some interval (a, b), we have

$$F'(x) = \mathfrak{p}(x),$$

where $\mathfrak{p}(x)$ is some function of bounded variation in (a, b). It follows from a known result in real variable theory (cf. (5)) that we may write for $a \leqslant x \leqslant b$,

$$F(x) = \int_a^x \mathfrak{p}(x)\,dx + k(x), \tag{21}$$

† *Note added in proof.* The condition (*) 'in some interval, however small, $F'(x)$ equals almost everywhere a positive function of bounded variation' may be weakened at the expense of some simplicity. Minor, but fairly obvious, complications of the proof would allow us to replace (*) by 'for some $\delta > 0$ and some n, $H_n'(x) > \delta$ almost everywhere in some interval, however small'.

https://doi.org/10.1017/S0305004100028632 Published online by Cambridge University Press

where $k'(x) = 0$ almost everywhere and $k(x)$ is monotonic increasing. In (a, b), $\mathfrak{p}(x)$ may be taken as a frequency function (apart from a normalizing constant) which will be of the class J''. Let the corresponding distribution function be $P(x)$. Then if we put

$$p = \int_a^b \mathfrak{p}(x)\,dx \quad (>0),$$

we have

$$\begin{aligned} P(x) &= 0 \quad (x<a) \\ &= p^{-1}\int_a^x \mathfrak{p}(x)\,dx \quad (a \leqslant x \leqslant b) \\ &= 1 \quad (x>b). \end{aligned}$$

Let us write $q = 1-p$. Then by (21) the function $q^{-1}\{F(x)-pP(x)\}$ is a distribution function, and we call it $Q(x)$. Thus

$$F(x) = pP(x)+qQ(x); \tag{22}$$

in other words, we can express $F(x)$ as the linear sum of two distribution functions, one of which is absolutely continuous with a frequency function of bounded variation. It is clear that $Q(x)$ will have a finite mean and variance. Let us write

$$\psi(t) = \int_{-\infty}^{+\infty} e^{itx}\,dP(x), \quad \chi(t) = \int_{-\infty}^{+\infty} e^{itx}\,dQ(x).$$

Then

$$\phi(t) = p\psi(t)+q\chi(t),$$

so that

$$\theta_n(t) = \{\phi(t/\sqrt{n})\}^n = \{p\psi(t/\sqrt{n})+q\chi(t/\sqrt{n})\}^n. \tag{23}$$

In the expansion of this binomial expression we shall sum together all those terms which contain $\psi(t/\sqrt{n})$ to at least the power βn, where β is any positive number less than p. This sum may be considered, apart from a constant factor, to be the characteristic function $\Psi_n(t)$ of some probability distribution. A similar argument applies to the remaining terms giving a characteristic function $\Xi_n(t)$. Thus we may write

$$\theta_n(t) = p_n\Psi_n(t)+q_n\Xi_n(t). \tag{24}$$

The constants p_n and q_n may be obtained by putting $t = 0$ in the appropriate terms of (23). Obviously $p_n > 0$, $q_n \geqslant 0$ and $p_n+q_n = 1$. We can get a bound on q_n by the usual Bernoulli approximation to the binomial distribution

$$\begin{aligned} q_n &\sim \int_{-\infty}^{-\frac{(p-\beta)\sqrt{n}}{\sqrt{(pq)}}} \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2}\,dx, \\ &\sim \frac{C}{\sqrt{n}} e^{-cn}, \end{aligned}$$

for some constants C and c. Thus Σq_n converges. We must now prove that $\Psi_n(t)$ and $\Psi_n''(t)$ tend to the correct limits. It is not difficult to show that the variance $|\Xi''(0)|$ is bounded, so that $|\Xi_n(t)|$ and $|\Xi_n''(t)|$ are bounded for all t. Since $q_n \to 0$,

$$\lim_{n\to\infty} q_n |\Xi_n(t)| = 0$$

and

$$\lim_{n\to\infty} q_n |\Xi_n''(t)| = 0,$$

for all t. If we now use Lemma 2 and the fact that $p_n \to 1$ we can, from (24), prove that $\Psi_n(t)$ and $\Psi_n''(t)$ tend to the correct limits. But $\Psi_n(t)$ contains $\{\psi(t/\sqrt{n})\}^{\beta n}$ as a factor,

https://doi.org/10.1017/S0305004100028632 Published online by Cambridge University Press

and since $\psi(t)$ is J'' an argument of the type used in § 3 will prove the convergence of $\Psi_n(t)$ and $\Psi_n''(t)$ to be dominated. This proves the theorem.

Let the distribution functions corresponding to $\Psi_n(t)$ and $\Xi_n(t)$ be $P_n(x)$ and $Q_n(x)$. Then from (24) we have

$$H_n(x) = p_n P_n(x) + q_n Q_n(x).$$

For n large enough $P_n(x)$ will be absolutely continuous with a frequency function which we shall call $\mathfrak{p}_n(x)$. $H_n(x)$ and $Q_n(x)$ will have derivatives almost everywhere and we call these $\mathfrak{h}_n(x)$ and $\mathfrak{q}_n(x)$. Thus almost everywhere

$$\mathfrak{h}_n(x) = p^n \mathfrak{p}_n(x) + q^n \mathfrak{q}_n(x). \tag{25}$$

By an argument of the type employed in the proof of Lemma 5 it is possible to prove that

$$\lim_{n\to\infty} \int_{-\infty}^{+\infty} \left| x^m \mathfrak{p}_n(x) - \frac{x^m}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2} \right| dx = 0,$$

for $m = 0$ or 2. From (25) we have almost everywhere

$$\left| x^m \mathfrak{h}_n(x) - \frac{x^m}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2} \right| \leqslant \left| x^m \mathfrak{p}_n(x) - \frac{x^m}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2} \right| + x^m q_n \mathfrak{p}_n(x) + x^m q_n \mathfrak{q}_n(x).$$

Thus, using the theorem on the integral of a derivative and the boundedness of the variance of $P_n(x)$ and $Q_n(x)$,

$$\int_{-\infty}^{+\infty} \left| x^m \mathfrak{h}_n(x) - \frac{x^m}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2} \right| dx \leqslant C q_n + \int_{-\infty}^{+\infty} \left| x^m \mathfrak{p}_n(x) - \frac{x^m}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2} \right| dx,$$

for some constant C.

If we now merely assume that $q_n \to 0$ and not the full content, $\Sigma q_n < \infty$, of almost certain convergence, we obtain

COROLLARY. *If the conditions of Theorem* 5 *are satisfied,*

$$\lim_{n\to\infty} \int_{-\infty}^{+\infty} \left| x^m \mathfrak{h}_n(x) - \frac{x^m}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2} \right| dx = 0,$$

for $m = 0$ *or* 2.†

I am indebted to D. R. Cox for many discussions on the subject of this paper and to D. V. Lindley for critical comments on the draft.

REFERENCES

(1) CRAMÉR, H. *Skand. AktuarTidskr.* (1928), p. 13.
(2) CRAMÉR, H. *Random variables and probability distributions* (Camb. Tracts Math. no. 36, 1937).
(3) HAMMERSLEY, J. M. *Proc. Camb. phil. Soc.* 48 (1952), 592–9.
(4) KHINTCHINE, A. I. *Mathematical foundations of statistical mechanics* (New York, 1949), pp. 166–74.
(5) MCSHANE, E. J. *Integration* (Princeton, 1947), p. 202.
(6) TITCHMARSH, E. C. *Introduction to the theory of Fourier integrals* (Oxford, 1948), p. 172.

STATISTICAL LABORATORY
CAMBRIDGE

† *Note added in proof.* A number of results of this type have recently appeared in a paper (in Russian) by Prohorov, Yu. V. See *Doklady Akad. Nauk SSSR* (N.S.), 83 (1952), 797–800.

https://doi.org/10.1017/S0305004100028632 Published online by Cambridge University Press