A CLASS OF "CENTRAL LIMIT THEOREMS" FOR CONVOLUTION PRODUCTS OF GENERALIZED FUNCTIONS

BY
REUBEN HERSH

Introduction and summary. From the naive viewpoint of a nonprobabilist, the remarkable fact about the central limit theorem is that (under suitable conditions) the average of a large number of random variables converges (in a suitable sense) to a limit *which is independent of the things being averaged, but is completely determined by the averaging process itself.*

In the present paper, this situation is abstracted from its connection with probability. As we shall see, the restriction to positive functionals which characterizes probability is in no way necessary for validity of the limit theorems. Moreover, we are able to dispense completely with measure theory, in both definitions and proofs. In a purely algebraic and function-space context, we find *a whole sequence of averaging processes, each with its own limit function*; the first two of them generalize the weak law of large numbers and the central limit theorem.

We prove convergence in three different senses, under three different sets of assumptions: first convergence of moments, then uniform convergence of characteristic functions, and finally weak convergence of distribution functions. Our first theorem is purely combinatorial; our next, essentially a calculus theorem; in our third we use function-space techniques borrowed from the modern theory of partial differential equations. By doing so, we obtain rather refined and precise results on the strongest norm in which we can assert convergence. In this respect our results may be of interest even in the classical probabilistic cases. In particular (see Corollary 2 below) we show that if a very mild extra regularity condition is imposed on the summands in the classical central limit theorem, then not only do the partial sums converge in law to the Gaussian, but *every derivative* of the c.d.f. of the partial sum exists and converges to the corresponding derivative of the Gaussian. Theorem 3 possesses an interpretation in terms of electrical engineering, which is given in the Remark following Corollary 2. It also has a surprising application to numerical analysis, which is described in the Postscript.

We begin with an abstract algebra A over the real numbers, equipped with a linear functional $\mathscr{L}$. (If $a \in A$ were a random variable, $\mathscr{L}(a^k)$ would be its kth

Presented to the Society, November 25, 1966 under the title *A family of nonprobabilistic central limit theorems*; received by the editors November 1, 1966 and, in revised form, November 20, 1967.

moment.) For $\alpha=1, 2, \ldots$ we define an α-order average, $S_{N,\alpha}(a_n)=N^{-1/\alpha}\sum_{n=1}^{N} a_n$. Under natural hypotheses on the a_n, we find in Theorem 1 that as $N\to\infty$ the kth moment of the average, $\mathscr{L}(S_{N,\alpha})^k$, converges to a limit $c_{\alpha,k}$ which is independent of the $\langle a_n\rangle$.

With each $a\in A$ we next associate a characteristic "function," the formal power series $\chi(a;t)=\sum \mathscr{L}(a^k)(it)^k/k!$. If the $\mathscr{L}(a_n^k)$ do not grow too wildly, χ exists in some interval as a genuine function. In that case, we find that $\chi(S_{N,\alpha};t)$ converges uniformly to the limit

$$\chi_\alpha(t) = \sum_{k=0}^{\infty} c_{\alpha,k}\frac{(it)^k}{k!} = \exp\left[\frac{(it)^\alpha}{\alpha!}\right].$$

For all α not a multiple of 4, χ_α is the Fourier transform of a function $\mathscr{N}_\alpha(x)$. $\mathscr{N}_2(x)$ is of course the Gaussian normal density function, as required by the central limit theorem; $\mathscr{N}_1(x)$ is a Dirac delta function, as required by the weak law of large numbers. $\mathscr{N}_3(x)$ turns out to be the Airy function of classical geometrical optics. For $\alpha=4$, 8, etc., χ_α does not have a Fourier transform even in the sense of Schwartz distributions.

In order to go from convergence of $\chi(S_{N,\alpha};t)$ to convergence of their Fourier transforms $\mathscr{N}(S_{N,\alpha};x)$, we introduce the Sobolev-Schwartz-Lax spaces G_r and then obtain weak convergence in G_r of convolution products of generalized functions (Schwartz distributions) for all α not a multiple of 4. These spaces, though now standard tools in partial differential equations, do not seem to have been applied to the central limit problem—perhaps because they lack the property of closure under convolution that makes L_1 such a convenient space for probabilists. On the other hand, with their use we can state and prove a single theorem that encompasses both stronger and weaker kinds of convergence than are taken into account by the standard methods in probability theory. Then, under appropriate conditions, Sobolev's lemma leads easily to conclusions about pointwise convergence, uniform convergence or convergence of measures.

If the $\mathscr{N}(S_{N,\alpha};x)$ are distributional derivatives of measures then the rth order indefinite integral of $\mathscr{N}(S_{N,\alpha};x)$ converges pointwise to a corresponding integral of $\mathscr{N}_\alpha(x)$. In particular, if $r=1$, the $\mathscr{N}(a_n;x)$ are signed measures of total mass 1. If they should happen to be positive measures (which is possible only for $\alpha=1$ or 2) we are back to the usual forms of the classical limit theorems of probability theory.

To make the analogy complete, we conclude by defining $\mathscr{N}(S_{N,\alpha};x)$ directly from the $\mathscr{N}(a_n;x)$, so that our final convergence theorem is proved without assuming existence of moments higher than the αth. Thus our last step is to discard the algebra A with which we started.

The striking exceptional character of the case when α is a multiple of 4 seems to demand some intuitive clarification. To my mind this is the most intriguing of the many open questions suggested by these results.

Convergence of moments. Let there be given a real algebra A with identity, on which is defined a linear functional $\mathscr{L}$. A is not necessarily commutative. Elements of A will be written a_n. We imbed the real numbers as a subset of A by regarding 1 as the multiplicative identity in A. We set $a^0=1$ for all a, and assume $\mathscr{L}(1)=1$.

Motivated by the notion of *independent*, *standardized* random variables, we introduce, for sequences $\langle a_n \rangle$ from A and for every positive integer α, the following hypotheses:

HYPOTHESIS I. $\mathscr{L}[\prod_{n=1}^{N} a_n^{k_n}]=\prod \mathscr{L}(a_n^{k_n})$ *if all* a_n *are distinct*, $n=1,\ldots, N$;

$$\mathscr{L}(a_n^k) = 0 \quad \textit{for } 1 \leqq k < \alpha; \qquad \mathscr{L}(a_n^\alpha) = 1.$$

(The k_n are positive integers. If A is not commutative, $\prod a_n$ must of course be understood as an ordered product.)

HYPOTHESIS II. *For all positive integers k, there exists a real number $M(k)$ such that* $|\mathscr{L}(a_n^k)| \leqq M(k)$.

REMARK. If the a_n are random variables all of whose moments are finite, the first statement in Hypothesis 1 amounts to requiring that all powers of a_n and a_m are uncorrelated; this is a slightly weaker hypothesis than independence. The second statement, for $\alpha=2$, is the usual normalization that sets the mean at 0 and the variance at 1. Hypothesis II is the simplest possible "uniformity" condition; it could be weakened, but in view of Theorems 2 and 3 below, we have not considered it worthwhile to strain for maximum generality here.

It is evident that a sequence satisfying Hypothesis I cannot include any real multiple of the identity element 1.

Before proceeding further it may be desirable to give a concrete example of an algebra A containing a sequence $\langle a_n \rangle$ satisfying Hypotheses I and II.

For this purpose, let $\chi(t)$ be a C^∞ function with compact support, such that $\chi(0)=1$, $\chi^{(k)}(0)=0$ for $1 \leqq k \leqq \alpha-1$, $\chi^{(\alpha)}(0)=i^\alpha$. Let $\mathscr{F}$ stand for the Fourier transform and let

$$\mathscr{N}(x) = \mathscr{F}^{-1}[\chi] = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-ixt)\chi(t)\,dt.$$

Since $\mathscr{N}$ dies out at ∞ faster than any power, we can take A to be the algebra of continuous functions $f(x_1,\ldots,x_r)$ of an arbitrary (finite) number of real variables x_n, having at most polynomial growth at ∞, with

$$\mathscr{L}(f) = \int f(x_1,\ldots,x_r)\mathscr{N}(x_1)\cdots\mathscr{N}(x_r)\,dx_1\cdots dx_r.$$

Then the sequence of coordinate functions $\langle f_n \equiv x_n \rangle$ satisfies Hypotheses I and II, since, for all k, $\chi^{(k)}(0)=(i)^k \int_{-\infty}^{\infty} x^k \mathscr{N}(x)\,dx$.

We now introduce two notations:
$S_{N,\alpha}=N^{-1/\alpha}\sum_{n=1}^{N} a_n$ where $\langle a_n\rangle$ satisfy Hypotheses I and II, and
$c_{\alpha,k}=\lim_{N\to\infty}\mathscr{L}[(S_{N,\alpha})^k]$.

THEOREM 1. *Under Hypotheses* I and II, $c_{\alpha,k}$ *exists, and is equal to zero, unless* k *is a multiple of* α; *if* $k=l\alpha$, $c_{\alpha,k}=k!/(\alpha!)^l l!$.

Proof. The result is trivial for $l=k=0$; suppose then that $k\geqq 1$. Now,

$$(S_{N,\alpha})^k = N^{-k/\alpha}\sum a_{n_1}\cdots a_{n_k},$$

where the summation is over all ordered sets of k subscripts $n_1\cdots n_k$ between 1 and N. By Hypothesis I, applying $\mathscr{L}$ allows us to permute factors in each product, and then to discard any product in which any factor is repeated fewer than α times. Of the remaining terms in the sum, we distinguish those in which every factor is taken to exactly the α power. In such a term there obviously must be exactly $l=k/\alpha$ distinct factors. We call these the "leading terms." The number of leading terms is just equal to (number of distinguishable permutations of a set of k objects of l kinds, with exactly α of each kind) times (number of ways of choosing l objects from a set of N objects). That is, it is zero unless α divides k; if $k=\alpha l$, it is

$$\frac{k!}{(\alpha!)^l}\cdot\frac{N(N-1)\cdots(N-l+1)}{l!}.$$

Now, $\mathscr{L}$ applied to each leading term yields 1; so, observing that the number of leading terms is a polynomial in N of degree $l=k/\alpha$, we see that

$$\lim_{N\to\infty}(N^{-k/\alpha}\cdot\mathscr{L}(\text{sum of leading terms})) = \frac{k!}{(\alpha!)^l l!}$$

if k is a multiple of α, and zero otherwise.

Let us call those terms in $S_{N,\alpha}^k$ still unaccounted for, "the remainder." These terms are products, where every factor is taken at least to the α power, and one or more of them is taken to a power greater than α; thus we see that any such term has fewer than k/α distinct factors. By Hypothesis II, $\mathscr{L}$ applied to any such term is not greater in absolute value than $\max_{r<k}|M(r)|^k$, an estimate independent of N. Moreover, the number of such terms is expressed as the sum of certain combinatorial coefficients *that do not depend on* N times the number of ways of choosing from N objects a subset of *fewer than* k/α. Thus the number of such terms is a polynomial in N of degree *less than* k/α, and so, dividing by $N^{k/\alpha}$, using our estimate for $\mathscr{L}$ applied to each term, and letting $N\to\infty$, we see that in the limit $\mathscr{L}$ applied to the remainder vanishes, and the proof is complete.

Before going on, we should confess that Theorem 1 is not really needed in the sequel. The reason is that here, as in the standard theory, the method of characteristic functions turns out to be simpler and neater than the method of moments. It might nevertheless be interesting to see if a proof of Theorem 3 below could indeed be based on Theorem 1 instead of on Theorem 2.

Convergence of characteristic functions. Next we define a "characteristic function" $\chi(a; t)$ for each $a \in A$, and a limiting characteristic function $\chi_\alpha(t)$ for each α:

$$\chi(a; t) = \sum_{k=0}^{\infty} \mathscr{L}(a^k) \frac{(it)^k}{k!}; \qquad \chi_\alpha(t) = \sum_0^\infty c_{\alpha,k} \frac{(it)^k}{k!}.$$

$\chi(a; t)$ is merely a formal power series, which need not converge for any $t \neq 0$; on the other hand,

LEMMA 1. $\chi_\alpha(t) = \exp[(it)^\alpha/\alpha!]$.

Proof.

$$\chi_\alpha(t) = \sum_{k=0}^{\infty} c_{\alpha,k} \frac{(it)^k}{k!} = \sum_{l=0}^{\infty} \frac{(l\alpha)!}{(\alpha!)^l l!} \frac{(it)^{l\alpha}}{(l\alpha)!} = \sum_{l=0}^{\infty} ((it)^\alpha/\alpha!)^l/l! = \exp[(it)^\alpha/\alpha!].$$

Our formal power series $\chi(a; t)$ satisfy some familiar identities; we have

LEMMA 2. *If* $\mathscr{L}(a^m b^n) = \mathscr{L}(a^m)\mathscr{L}(b^n)$ *for all* m, n, *then*

$$\chi(a+b; t) = \chi(a; t)\chi(b; t).$$

Proof. The given hypothesis together with the binomial theorem implies, for all k,

$$\mathscr{L}(a+b)^k = \sum_{j+l=k} \frac{k!}{j!\,l!} \mathscr{L}(a^j)\mathscr{L}(b^l).$$

Therefore,

$$\begin{aligned} \chi(a; t)\chi(b; t) &= \sum_j \mathscr{L}(a^j) \frac{(it)^j}{j!} \sum_l \mathscr{L}(b)^l \frac{(it)^l}{l!} \\ &= \sum_k \sum_{j+l=k} \mathscr{L}(a^j)\mathscr{L}(b^l) \frac{(it)^{j+l}}{j!\,l!} \\ &= \sum_k \mathscr{L}(a+b)^k \frac{(it)^k}{k!} = \chi(a+b; t). \end{aligned}$$

LEMMA 3. *For any real* r *and all* a, $\chi(ra; t) = \chi(a; rt)$.

Proof.

$$\chi(ra; t) = \sum \mathscr{L}(ra)^k \frac{(it)^k}{k!} = \sum \mathscr{L}(a^k) \frac{(irt)^k}{k!} = \chi(a; rt).$$

The obvious thing to prove now is that $\chi(S_{N,\alpha}; t)$ converges to χ_α as $N \to \infty$. To do this, however, we need the existence of $\chi(S_{N,\alpha}; t)$ in some interval. So we add

HYPOTHESIS III. *There exists a positive number* M *such that*

$$|\mathscr{L}(a_n^k)| \leqq (n^{1/\alpha}M)^k k! \quad \textit{for all } k > \alpha \textit{ and } n > 0.$$

It now follows immediately that $\chi(a_n; t)$ is analytic if $|t| < 1/MN^{1/\alpha}$ for $n \leqq N$. Therefore,

THEOREM 2a. *Under Hypotheses* I, II *and* III, $\chi(S_{N,\alpha}; t) \to \chi_\alpha(t)$ *as* $N \to \infty$ *uniformly in any compact subset of* $|t| < 1/M$.

Proof. By Lemmas 2 and 3, $\chi(S_{N,\alpha}; t) = \prod_{n=1}^{N} \chi(a_n; t/N^{1/\alpha})$. By Hypothesis I the Taylor expansions at the origin are identical for all the $\chi(a_n)$ up to the αth term, and we have for each $\chi(a_n; t)$ the finite expansion

$$\chi(a_n; t) = 1 + \frac{(it)^\alpha}{\alpha!} + \frac{(it)^{\alpha+1}}{(\alpha+1)!} \chi^{(\alpha+1)}(a_n; \theta_n t)$$

with $0 \leqq \theta_n \leqq 1$ and $|t| < 1/MN^{1/\alpha}$. Therefore

$$\chi(S_{N,\alpha}; t) = \prod_{n=1}^{N} \left(1 + \frac{1}{N}\left[\frac{(it)^\alpha}{\alpha!} + \frac{(it)^{\alpha+1}\chi^{(\alpha+1)}(a_n; \theta_n t N^{-1/\alpha})}{(\alpha+1)!\,N^{1/\alpha}}\right]\right)$$

for $|t| < 1/M$. It is an exercise in elementary calculus to show that as $N \to \infty$ this product goes to $\exp[(it)^\alpha/\alpha!]$, provided that the error terms $\chi^{(\alpha+1)}(a_n; \theta_n t N^{-1/\alpha})$ are bounded, for large N, uniformly with respect to n and t. Now, since $|\theta_n| \leqq 1$, this expression is uniformly close to $\chi^{(\alpha+1)}(a_n; 0) = i^{\alpha+1}\mathscr{L}(a_n^{\alpha+1})$ for N large and t bounded. An appeal to Hypothesis II, which we evidently now need only for the case $k = \alpha + 1$, completes the proof.

In case Hypothesis III holds true for arbitrary M, the convergence is uniform in any compact set.

In Theorem 2a we have used moments higher than the $(\alpha+1)$th only in order to establish the existence of χ_n in some interval. If we *assume* $\chi_n(t)$ exists, we can prove a sharper theorem, which does not require that χ have an $(\alpha+1)$th derivative:

THEOREM 2b. *Let* $\chi_n(t)$ *be a sequence of functions defined and having* α *equicontinuous derivatives in a* $(1/Mn^{1/\alpha})$*-neighborhood of the origin, with all derivatives of order between* 1 *and* $\alpha - 1$ *vanishing at the origin, and with* $\chi_n(0) = 1$, $\chi_n^{(\alpha)}(0) = i^\alpha$, *for all n. Then* $\prod_{n=1}^{N} \chi_n(t/N^{1/\alpha})$ *converges to* $\exp[(it)^\alpha/\alpha!]$ *as* $N \to \infty$, *uniformly in any compact subset of* $|t| < 1/M$.

Proof. By the (extended) mean value theorem of differential calculus,

$$\chi_n(t) = 1 + \frac{\chi_n^{(\alpha)}(\theta_n t)(it)^\alpha}{\alpha!},$$

which can be rewritten as

$$\chi_n(t) = 1 + \frac{(it)^\alpha}{\alpha!} + \frac{(it)^\alpha}{\alpha!}\left(\chi_n^{(\alpha)}(\theta_n t) - \chi_n^{(\alpha)}(0)\right),$$

where $|\theta_n| \leqq 1$ and $\chi_n^{(\alpha)}(0) = 1$. Therefore,

$$\chi_n\left(\frac{t}{N^{1/\alpha}}\right) = 1 + \frac{1}{N}\left(\frac{(it)^\alpha}{\alpha!} + \chi_n^{(\alpha)}\left(\frac{\theta_n t}{N^{1/\alpha}}\right) - \chi_n^{(\alpha)}(0)\right)\cdot$$

Because of the equicontinuity of $\chi_n^{(\alpha)}$ near the origin, $\chi_n^{(\alpha)}(\theta_n t/N^{1/\alpha})-\chi_n^{(\alpha)}(0)$ is small for large N uniformly in n for $|t|<1/M$, and the conclusion follows as in Theorem 2a.

Discerning readers will recognize that for $\alpha=2$ these proofs are virtually the standard Liapounoff approach to the central limit theorem; see, e.g., [4]. Equicontinuity of $\chi_n^{(\alpha)}$ near $t=0$ is our version of the Lindeberg condition. It remains to convert uniform convergence of χ into weak convergence of its inverse Fourier transform.

Convergence of distributions. The subtitle is ambiguous; the word "distribution" nowadays means one thing to an analyst and another to a probabilist. But while the objects constructed in this section are Schwartz distributions, they also generalize the notion of probability distributions, so the term is appropriate in whichever sense it is understood. (To avoid confusion, in the following we will refer to Schwartz distributions as generalized functions or functionals, and to probability distributions as c.d.f.'s.)

In probability theory the objects of principal interest are not the moments or the characteristic functions $\chi(t)$ but measures $m(x)$ of which the χ are Fourier transforms. One proves uniform convergence of pointwise products of $\chi_n(t)$ in order to obtain weak convergence of convolution products of m_n (or of their canonical representatives, the c.d.f.'s $F_n(x)=m_n(-\infty, x])$.

Our principal question is, to what extent can the classical limit theorems of probability be generalized from positive measures to more general functionals, and from $\alpha=1$, 2 to $\alpha\geqq 3$? With this in mind, we now focus our attention on a class of linear functionals which are much more general than measures, but on which the operations of convolution ($\ast$) and Fourier transformation ($\mathscr{F}$) are well defined. The characteristic functions of probability theory satisfy $|\chi(t)|\leqq 1$. So far we have had to make assumptions about our $\chi(t)$ only near $t=0$. Suppose we add the mild hypothesis that near ∞ they are smaller than $K|t|^r$ for some r and some positive K. If r is negative, we are assuming more than in the classical case, and we will obtain correspondingly stronger conclusions. If r is positive, we are assuming less; we will still get a convergence theorem, in a suitably weakened sense. In fact, *for α not a multiple of* 4, we will obtain convergence theorems for convolution products which in the cases $\alpha=1$ and 2 generalize the weak law of large numbers and the central limit theorem.

First we must introduce some facts from Fourier analysis. (A good general reference for this material is Yosida [10], the first two chapters of Hörmander [7], or Bers and Schechter [11], pp. 167–168.) We write $\hat{f}(t)$ for the Fourier transform of $f(x)$, $\hat{f}=\int \exp(ixt)f\,dx$. For any real number r we define G_r as the completion in the $\|\quad\|_r$-norm of the space of smooth functions f such that

$$\int [(1+|t|)^r\hat{f}(t)]^2\,dt = \|f\|_r^2$$

is finite.

G_r is a Hilbert space. G_0 is just L_2. We write $(\cdot,\ \cdot)_r$ for the inner product in G_r. If $\hat{f}$ is bounded, $f \in G_r$ for $r < -1/2$. G_r and G_{-r} are dual Hilbert spaces, in the sense that to each bounded linear functional l on G_r corresponds an element v in G_{-r}, as follows: $l(u) = (v, u)_0$ for all u in G_r. The correspondence $l \leftrightarrow v$ is biunique, and $\|v\|_{-r} = \|l\|_0$. If $a \leqq b$, then $G_b \subset G_a$, and $f_n \to f$ in G_b implies $f_n \to f$ in G_a.

If r is a positive integer, $\|f_n\|_r \to 0$ iff $\langle f_n \rangle$ and its derivatives of order up to r go to zero in L_2.

We say $f_n \to f$ weakly in G_{-r} if, for every g in G_r, $(f_n, g) \to (f, g)$, i.e., $\int \hat{f}_n \hat{g}\, dt \to \int \hat{f}\hat{g}\, dt$. For weak convergence, $a \leqq b$ and $f_n \to f$ in G_{-a} implies $f_n \to f$ in G_{-b}.

We define convolution for functionals in G_r by the formula $\mathscr{F}(f * g) = \hat{f} \cdot \hat{g}$. We define similarity transformations with a real parameter c by the formula $(f(cx), g(x)) = (f(x), g(x/c))/c$ or, equivalently, by $\mathscr{F}[f(cx)] = \hat{f}(t/c)/c$.

If $\hat{f}$ is of class C_k near $t=0$, we define the kth moment μ_k of f as $(-i)^k \hat{f}^{(k)}(0)$.

Needless to say, if $f(cx)$, $\mathscr{F}[f]$, $f * g$, or $\int x^k f(x)\, dx$ exist in the classical sense, these definitions are equivalent to the classical ones.

Our goal is to prove that certain convolution products of functionals converge weakly in G_r to a limit $\mathscr{N}_\alpha$ which is the Fourier transform of $\chi_\alpha(t) = \exp((it)^\alpha/\alpha!)$. For this purpose we need

THEOREM 4. *For all* α *not equal to* 1 *and not a multiple of* 4,

$$\mathscr{N}_\alpha(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp\left[-itx + \frac{(it)^\alpha}{\alpha!}\right] dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-itx)\chi_\alpha(t)\, dt$$

is a convergent improper Riemann integral and defines an entire function of x.

Since the proof of Theorem 4 requires a detour from the main thread of our argument, we postpone it to the end of this paper. Here we content ourselves with four remarks:

REMARK 1. $\mathscr{N}_1 = \delta(x-1)$, Dirac's delta function with mass concentrated at $x=1$, since

$$\mathscr{N}_1[f] = \frac{1}{2\pi}\int \chi_1(t)\hat{f}(t)\, dt = \frac{1}{2\pi}\int e^{i1\cdot t}\hat{f}(t)\, dt = f(1)$$
$$= \int \delta(x-1) f(x)\, dx.$$

REMARK 2. Since our numbers $\mathscr{L}(a_n)^k$ could in particular be moments of a sequence of probability distributions satisfying Markov's condition, it follows from the classical weak law of large numbers what $\mathscr{N}_1$ must be. For similar reasons, now referring to the central limit theorem, $\mathscr{N}_2$ is, and has to be, the Gaussian normal density function.

REMARK 3. $\mathscr{N}_3$ comes as a surprise; it equals

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} \exp\left(-itx - i\frac{t^3}{6}\right) dt = \frac{1}{\pi}\int_0^{\infty} \cos\left(tx + \frac{t^3}{6}\right) dt = \sqrt[3]{2}\, Ai(x\sqrt[3]{2}),$$

where $Ai(x)=(1/\pi)\int_0^\infty \cos(xt+t^3/3)\,dt$ is the well-known Airy's integral of geometrical optics. $Ai(x)$ is expressible in terms of a pair of Bessel functions of order 1/3 (see [9]). It is interesting to note that $\mathcal{N}_{\alpha+1}$ satisfies the simple differential equation $(d/dx)^\alpha \mathcal{N}_{\alpha+1}+(\alpha!)x\mathcal{N}_{\alpha+1}=0$. For $\alpha=2$ this is "Airy's equation." It is shown in [9] that $Ai(x)$ is positive for $x>0$, but changes sign infinitely often for $x<0$, so, in particular, it is not a probability density. This is no accident, since

REMARK 4. For $\alpha\geqq 3$ $\mathcal{N}_\alpha$ is not a probability density. If it were, it would have mean zero and variance zero. This is possible only for the Dirac density with unit mass at the origin, and then all higher moments must be 0, not $c_{\alpha,k}$.

With all this preparation out of the way, we are now almost ready to state our theorem. We suppose given a sequence $\mathcal{N}_n$ of tempered distributions (each $\mathcal{N}_n$ is in some G_r). We define the sequence of convolution products

$$W_N = \mathop{\ast}_{n=1}^{N} N^{1/\alpha}\mathcal{N}_n(N^{1/\alpha}x).$$

We introduce

HYPOTHESIS IV. *The $\mathcal{N}_n$ have moments μ_k for $k\leqq\alpha$, and, for all $\mathcal{N}_n$, $\mu_0=\mu_\alpha=1$, $\mu_k=0$ for $0<k<\alpha$. The αth order derivatives of all the Fourier transforms $\hat{\mathcal{N}}_n=\chi_n$ are equicontinuous in some fixed neighborhood of $t=0$.*

HYPOTHESIS V. *The products W_N, for N sufficiently large, all lie in some fixed G_{-r}, and are uniformly bounded in G_{-r}.*

THEOREM 3a. *Let* $\alpha\not\equiv 0 \bmod 4$, $r>1/2$. *Then, under Hypotheses* IV *and* V, *$W_N\to\mathcal{N}_\alpha$ weakly in G_{-r}.*

THEOREM 3b. *Let* $\alpha\equiv 2 \bmod 4$. *Then, under Hypotheses* IV *and* V *with any real r, $W_N\to\mathcal{N}_\alpha$ weakly in G_{-r}.*

Proof. Note first of all that since χ_α is bounded for $\alpha\not\equiv 0 \bmod 4$, $\mathcal{N}_\alpha$ is in G_{-r} for $r>1/2$, and since χ_α for $\alpha\equiv 2 \bmod 4$ vanishes at $|t|=\infty$ faster than any negative power, $\mathcal{N}_{4p+2}$ is in G_{-r} for any real r.

Now, we must prove that for any $f\in G_r$, $(W_N,f)\to(\mathcal{N}_\alpha,f)$. It is well known that a real sequence approaches a limit L if from every subsequence of the given sequence can be chosen a sub-subsequence converging to L. Suppose then that we confront an arbitrary subsequence of W_N. Since the W_N are uniformly bounded in G_{-r}, the weak compactness of the unit sphere in a Banach space implies that we can find a sub-sequence W_{N_k} which converges weakly to *something* in G_{-r}, say to H. What we must prove is that $H=\mathcal{N}_\alpha$. Now $(W_{N_k},f)=2\pi\int \hat{W}_{N_k}\hat{f}\,dt$, so we have

$$2\pi\int \hat{W}_{N_k}\hat{f}\to(H,f) = 2\pi\int \hat{H}\hat{f}$$

which means that $[\chi_{N_k}(1+|t|)^{-r}]$ converges weakly in L_2 to $H(1+|t|)^{-r}$. At the same time, by Hypothesis IV and Theorem 2, we know that the sequence $\hat{W}_N$, and so also the subsequence $\hat{W}_{N_k}$, converges uniformly on bounded sets to χ_α. But from this it is immediate that $\hat{H}=\chi_\alpha$, and therefore $H=\mathscr{N}_\alpha$. Q.E.D.

This proof is a simple adaptation of the standard one in probability theory; we have merely replaced the Helly selection principle with the weak compactness of the unit sphere in Banach space (Alaoglu's theorem).

Notice that the equicontinuity condition in Hypothesis IV is automatically fulfilled in the important special case when the $\mathscr{N}_n$ are all the same (the "identically distributed" case).

Unlike the usual statements of the limit theorems of probability, Hypothesis V explicitly restricts the partial products W_N, not merely the factors $\mathscr{N}_n$. We could remove this blemish by requiring the $\mathscr{N}_n$ to belong to a convolution algebra—for example, to L_1, as is customary in probability theory. (See Corollaries 1 and 2 below.) But we prefer to make it clear that the conclusion actually holds good in greater generality. Some of the factors can be very singular if there are enough smooth ones to counteract them. Moreover, it now becomes evident that the better-behaved are the factors $\mathscr{N}_n$, the stronger is the norm in which we can assert convergence.

In case the functions being convolved are in C^k for some $k \geqq 0$, we might ask what conclusions can be drawn about convergence in C^k. A similar question could be asked about functionals on C^k; for $k=0$ these are signed measures.

By the one-dimensional special case of Sobolov's lemma (Theorem 2.2.7, [7]; Lemma 3, p. 167, [11]) functions in G_r have continuous derivatives of order up to s if $r-s>1/2$, and convergence in G_r ("convergence in norm") implies uniform convergence of derivatives of order up to s ("convergence in the C_s topology").

From this we see that a linear functional continuous in the C_s topology is necessarily continuous on G_r if $r-s>1/2$. This means that C_s', the dual of C_s, is embedded in G_{-r}, the dual of G_r, if $r-s>1/2$. In particular, C_0', the space of signed measures, is in G_{-r} for $r>1/2$.

If a sequence of measures converges to a measure in the sense of weak convergence in G_{-r}, $r>1/2$, then it converges in the weak* topology of C_0'. This is merely a restatement of the well known and often used fact that a measure is determined by its action on smooth (say, C_2) functions.

By the same token, if $f_n \to f$ weakly in G_{-r} and f_n and f are in C_s', then $f_n \to f$ in the weak* topology of C_s'.

For $s=0$, this would be equivalent, for finite positive measures, to convergence of cumulative distribution functions at continuity points of the limit c.d.f. ("convergence in law").

We now define $C_s(K)$ to mean the subspace of C_s with support in $K \subset R_1$. For s a positive integer and compact K, weak* convergence in $C_s'(K)$ is weak convergence of sth order distributional derivatives of measures.

Just as the c.d.f. of probability is an indefinite integral of a measure, we can introduce a generalized c.d.f. as the s-fold iterated indefinite integral of our element of C_s'. In probability theory the lower limit of integration is taken at $-\infty$; since we need not have finite absolute mass in our general case, we take the fixed limit of integration at a finite point, say the origin. Then weak convergence in G_{-r} implies pointwise convergence of the generalized c.d.f. for $r > s+1/2$.

It follows immediately from Sobolev's lemma that if $r > s+1/2$, weak convergence in G_r implies weak convergence in C_s.

Weak convergence in C_0 implies pointwise (not uniform) convergence, since the delta function in particular is in the dual of C_0. Thus weak convergence in G_r for $r > 1/2$ implies pointwise convergence; weak convergence in G_r for $r > n+1/2$ implies pointwise convergence of all derivatives up to the nth.

Thus we have the easy

COROLLARY 1. *If the* $\mathcal{N}_n$ *are measures, satisfying Hypothesis* IV *for* $\alpha = 1$ *or* 2, *then* $W_N \to \mathcal{N}_\alpha$ *as measures in the weak* topology.*

Proof. For measures, the condition $\mu_0 = 1$ implies $|\chi_n| \leqq 1$, so that

$$\hat{W}_N = \prod (\chi_n(tN^{-1/\alpha})) \leqq 1.$$

Therefore we automatically have Hypothesis V with any $r > 1/2$ and $W_N \to \mathcal{N}_\alpha$ weakly in G_{-r}. In particular, if F_N is the c.d.f. of W_N, then for any compactly supported $f \in C_2 \subset G_1$,

$$\int f\, dF_N = (W_N, f) = 2\pi \int \hat{W}_N \hat{f}\, dt \to (\mathcal{N}_\alpha, f) = \int e^{-x^2/2} f(x)\, dx \qquad (\alpha = 2)$$
$$= f(1) \qquad (\alpha = 1)$$

and since C_2 is dense in C_0, the corollary is proved.

This means that our result is the right one in the probabilistic case. For $\alpha = 1$, Corollary 1 is just Khinchin's theorem—i.e., the weak law of large numbers, but without assuming finite variance.

If $\alpha \equiv 2 \bmod 4$, we can state more refined results under suitable hypotheses. For example, the following is apparently a new result:

COROLLARY 2. *If the* $\mathcal{N}_n$ *are probability densities of independent random variables with* $\mu_1 = 0$, $\mu_2 = 1$, *and* χ_n'' *equicontinuous near* $t = 0$, *and if for some* $\varepsilon > 0$ *all but a finite number of the* $\mathcal{N}_n$ *belong to* G_ε *and satisfy* $\|\mathcal{N}_n\|_\varepsilon \leqq M$ *for some fixed* M, *then for any positive integer* k *the* k*th derivative of the partial product* W_N *converges as* $N \to \infty$ *to the* k*th derivative of* $\mathcal{N}_2 = (1/2\pi) \exp(-x^2/2)$, *uniformly on compact sets.*

Proof. Now the χ_n are not only bounded by 1 but also decay at $|t| = \infty$ faster than $|t|^{-\varepsilon}$. Therefore $\hat{W}_N$, the partial product, is less than 1 and decays at ∞ faster than $|t|^{-N\varepsilon+K}$, where K estimates the rate of growth of the product of the finite number of χ_n which may not be in G_ε. Thus for N large enough W_N is in G_r for arbitrary r. Moreover, max $|\chi_n| \leqq 1$ implies

$$\|W_N * \mathcal{N}_n\|_r = \|(1+|t|)^r \chi_N \chi_n\|_{L_2} \leqq \|(1+|t|)^r \chi_N\|_{L_2} = \|W_N\|_r$$

so that once we are far enough out in the sequence of partial products we are not only in G_r but we stay on a bounded set in G_r.

The conclusion now follows from Theorem 3 and Sobolev's lemma.

REMARK. Theorem 3 has a physical interpretation which has nothing to do with probability. In the theory of "linear systems" in electrical engineering, one uses the "impulse-response" function of a system. If this function is called G, then to any input f corresponds the output $G * f$. G is the Fourier transform of the "transfer function." Now suppose that N linear subsystems are connected in series (in cascade) to constitute a single large system. Both inputs and outputs are defined for all $-\infty < t < \infty$, not just $t > 0$. Suppose also that, for some given positive integer α not a multiple of 4, each subsystem satisfies $\alpha+1$ normalizing conditions prescribing the output values at $t=0$ for the inputs t^k, $0 \leqq k \leqq \alpha$. Then Theorem 3 says the following: if N is sufficiently large, and if a suitable change in time scale depending on N is made in each subsystem, then, independently of any further properties of the subsystems, the system as a whole will behave like that special linear system specified by the transfer function $\exp[(it)^\alpha/\alpha!]$.

Existence and regularity of $\mathcal{N}_\alpha$. We still have to prove Theorem 4.

Before proving $\mathcal{N}_\alpha$ entire for $\alpha \not\equiv 0 \bmod 4$ and $\neq 1$, we mention that if α is a multiple of 4, $\mathcal{N}_\alpha$ can be defined as a functional on the space of functions whose Fourier transforms decay more rapidly than $\exp(-t^\alpha)$. (See [5] for the theory of such functionals.)

On the other hand, it is obvious that if $\alpha \equiv 2 \bmod 4$, then $\chi_\alpha = \chi_{4p+2}$ will have a genuine classical inverse Fourier transform $\mathcal{N}_{4p+2}$. Since χ_{4p+2} is infinitely differentiable, $\mathcal{N}_{4p+2}$ dies out at ∞ faster than $|x|^{-r}$ for any r; since

$$\int \exp(-itx)\chi_{4p+2}(t)\,dt = \int \exp\left(-itx - \frac{(t)^{4p+2}}{(4p+2)!}\right) dt$$

converges uniformly for x in any bounded region of the complex plane, $\mathcal{N}_{4p+2}$ is an entire function. It is less obvious what happens when α is odd. Suppose then that α is odd and $\geqq 3$. To see that the integral for $\mathcal{N}_\alpha$ converges, we note that

$$\mathcal{N}_\alpha = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp\left(-itx \pm i\frac{t^\alpha}{\alpha!}\right) dt = \frac{1}{\pi}\int_0^\infty \cos\left(tx \pm \frac{t^\alpha}{\alpha!}\right) dt$$

is the sum of an alternating series. The terms are the areas under the arches of a cosine curve whose frequency grows, and whose wavelength goes monotonically to zero, as $t \to \infty$. Since the amplitude is 1 for all t, Leibnitz's test for alternating series tells us the integral converges.

On the other hand, it is also clear that χ_α is not integrable in the Lebesgue sense, for each arch, being convex, has area greater than that of the inscribed triangle, ABC, where the vertices A, B are at two adjacent zeros of χ, and the vertex C is at the maximum (or minimum) of χ in $A < t < B$. Since each triangle has height 1, the sum of the absolute values of their areas is clearly 1/2 the length of the positive real axis, so

$$\int_0^\infty |\cos (tx \pm t^\alpha/\alpha!)|\, dt \geqq \infty/2 = \infty.$$

So far, we know only that $\mathcal{N}_\alpha(x)$ exists, for all real x. It is not in L_1 or L_2, since χ_α neither vanishes at ∞ nor belongs to L_2. To see that $\mathcal{N}_\alpha$ is entire, we change the path of integration from the real t-axis to a pair of rays $\mathcal{R}_-$ and $\mathcal{R}_+$. If $\alpha=4p-1$, $\mathcal{R}_-$ is $\arg t=\pi+\pi/2(4p-1)$ and $\mathcal{R}_+$ is $\arg t=-\pi/2(4p-1)$. If $\alpha=4p+1$, $\mathcal{R}_-$ is $\arg t=\pi-\pi/2(4p+1)$ and $\mathcal{R}_+$ is $\arg t=\pi/2(4p+1)$. (For $\alpha=3$, this is 'Stokes transformation.') The rays $\mathcal{R}_-$ and $\mathcal{R}_+$ are so defined that as we go along an arc of fixed modulus ρ from the positive real axis to $\mathcal{R}_+$ or from the negative real axis to $\mathcal{R}_-$, Re $(it)^\alpha$ decreases monotonically from 0 to $-\rho^\alpha$.

Integrating from $-\infty$ to ∞ along $\{\mathcal{R}_- \cup \mathcal{R}_+\}$, we have a uniformly and absolutely convergent integral for any bounded complex x, so Theorem 4 is proved once we justify the change in path of integration. To do this, it is enough to show that the integral along a circular arc Γ of radius ρ between the real axis and the ray $\mathcal{R}_+$ (or $\mathcal{R}_-$) vanishes as $\rho\to\infty$. (Note that Jordan's lemma does *not* apply, Watson [9] to the contrary notwithstanding!) Break up Γ into two parts, Γ_1 with $0\leqq|\theta|\leqq 1/\rho^{3/2}$, Γ_2 with $1/\rho^{3/2}\leqq|\theta|\leqq$ arg of $\mathcal{R}_+$ or $\mathcal{R}_-$.

The integral over Γ_1 clearly goes to zero, since on Γ_1 the absolute value of the integrand is less than 1, and the arc-length less than $\rho\cdot 1/\rho^{3/2}=1/\sqrt{\rho}$. Over Γ_2, the arc length is less than ρ, and the integrand is less in absolute value than

$$\exp\left[\max_\theta\left(\operatorname{Re}\left(-ix\rho e^{i\theta}+\frac{(i\rho e^{i\theta})^\alpha}{\alpha!}\right)\right)\right] \leqq \exp\left[|x|\rho+\frac{\rho^\alpha}{\alpha!}\max \operatorname{Re}\,(ie^{i\theta})^\alpha\right].$$

Since $\mathcal{R}_\pm$ were so defined that on Γ Re $(ie^{i\theta})^\alpha$ decreases from 0 monotonically as $|\theta|$ or $|\pi-\theta|$ increases, on Γ_2 Re $(ie^{i\theta})^\alpha$ is negative, and

$$|\operatorname{Re}\,(ie^{i\theta})^\alpha| = |\operatorname{Im}\, e^{i\alpha\theta}| \leqq |\operatorname{Im}\, e^{i\alpha/\rho^{3/2}}|,$$

which is asymptotic for large ρ to $\alpha/\rho^{3/2}$. Thus the integrand over Γ_2 is less than

$$\exp\,[|x|\rho-(\rho^\alpha/2\alpha!)(\alpha/\rho^{3/2})]$$

for ρ sufficiently large. Since $\alpha\geqq 3$, we see that for any fixed complex x this is less than $\exp(-\rho/2\alpha!)$ for large enough ρ, and so, even when multiplied by an arc length of order ρ, it vanishes as $\rho\to\infty$. End of proof.

We can obtain the power series expansion of $\mathcal{N}_\alpha(x)$ by expanding $\exp(-ixt) = \sum(-ixt)^n/n!$ under the integral sign and integrating term by term. After making the change of variable $t'=t^\alpha/\alpha!$, the coefficients are obtained explicitly in terms of Euler's Γ-function evaluated at the points $1/\alpha, 2/\alpha, \ldots, (\alpha-1)/\alpha$. For $\alpha\equiv 2 \bmod 4$ this is legitimate without any prior manipulation; for α odd, it can be done after the path of integration is shifted to $\{\mathscr{R}_- \cup \mathscr{R}_+\}$. When $\alpha=3$, this leads to the representation of $Ai(z)$ or $\mathscr{N}_3(x)$ in terms of Bessel functions. For $\alpha>3$, the resulting series can be expressed in terms of generalized hypergeometric series; the greater α, the greater the number of parameters needed. For our present purposes, there is no need to make the $\mathscr{N}_\alpha(x)$ more explicit.

In conclusion, we mention that functions essentially identical to these attracted a (slight) flurry of interest in the '20's. They arose in the work of Hardy and Littlewood [6] on Waring's problem. Elaborate asymptotic analyses were made by Burwell [3] and Bakhoom [1], using the Riemann-Debye method of steepest descent. Burwell shows that for α odd, $\mathscr{N}_\alpha$ has infinitely many zeroes, which lie asymptotically on the real and imaginary axes. For α even, Pólya [8] and F. Bernstein [2] prove $\mathscr{N}_\alpha$ has infinitely many real zeroes; Pólya, that it has no nonreal zeroes. The motivation for such studies was of course Riemann's conjecture on the zeroes of the zeta function.

Postscript. While this paper was at the printer's, I was fortunate to find several Russian studies related to this work. The point of view and most of the results of the present paper are new, but Theorem 2b and Remark 3 are essentially contained in the work of Zhukov [12]. Subsequent brief announcements by Krylov [13], [14] and Studnev [15] follow up and extend Zhukov's work. The remarkable thing about [12] is that the functions $\mathscr{N}_\alpha$ arise from a concrete problem in numerical analysis. Let $F(\tau)$ be the solution operator for Cauchy's problem for a constant-coefficient partial differential equation of the form $u_t=P(\partial/\partial x)u$. F is approximated by a difference operator G, so that the exact solution is $u(n\tau, x)=F^n u^0(x)$, and the computed solution is $u^{*n}(x)=G^n u^0(x)$. Then, if $Gf=Ff$ for arbitrary polynomial $f(x)$ of degree $\leqq \alpha-1$, it turns out that u^{*n} converges weakly, as $n\to\infty$, to $\mathscr{N}_\alpha * u(n\tau, \sigma x)$, where σ depends on G, α, n and τ, and is given explicitly by Zhukov. The proof is a simple application of a limit theorem which is essentially the same as our Theorem 2b. This result of Zhukov, which should be of considerable computational interest, seems to have gone unnoticed in the U.S. up to the present.

Bibliography

1. N. G. Bakhoom, Proc. London Math. Soc. (2) **35** (1933), 83.
2. F. Bernstein, Math. Ann. (3) **79** (1918), 265.
3. W. R. Burwell, Proc. London Math. Soc. (2) **22** (1924), 57.
4. W. Feller, *Introduction to probability theory and applications*, Vol. 2, Wiley, New York, 1966.
5. I. M. Gel'fand and G. E. Šilov, Amer. Math. Soc. Transl. (2) **5** (1957), 221.
6. G. H. Hardy and J. E. Littlewood, Ges. Wiss. Göttingen (1920), 33.
7. L. Hörmander, *Linear partial differential operators*, Academic Press, New York, 1963.
8. G. Pólya, J. Reine Angew. Math. **158** (1927), 6.
9. G. N. Watson, *A treatise on the theory of Bessel functions*, Cambridge, 1962.
10. K. Yosida, *Functional analysis*, Academic Press, New York, 1965.

11. L. Bers, F. John and M. Schechter, *Partial differential equations*, Interscience, New York, 1964.

12. A. I. Zhukov, Uspehi Mat. Nauk (3) **14** (1959), 129

13. V. Yu. Krylov, Dokl. Akad. Nauk SSSR. **139** (1961), 20=Soviet Math. Dokl. **2** (1961), 855.

14. ———, Dokl. Akad Nauk, SSSR. **132** (1960), 1254=Soviet Math. Dokl. **1** (1960), 760.

15. Yu. P Studnev, Teor. Verojatnost. i Primenen. **12** (1967), 668.

University of New Mexico,
Albuquerque, New Mexico