CHAPTER VI

ASYMPTOTIC DISTRIBUTION OF THE SUM OF CHANCE VARIABLES

A. Asymptotic Results for Infinite Products. Stirling's Formula. Laplace's Formula (Sections 1 and 2)

1. Product of an Infinite Number of Functions[1]

1.1. *Asymptotic formula for a product.* In this section we shall derive two theorems concerning the product of an infinite number of functions of one real variable. (More general ones concern functions of several variables). Such theorems play an important part in probability theory. They were discovered by Laplace who devoted a considerable part of his *Théorie analytique des probabilités* (1812) to this problem.

Let us consider an infinite sequence of (real or complex) functions

$$f_1(x), \quad f_2(x), \quad f_3(x), \dots$$

of the real variable x; the $f_\nu(x)$ have derivatives up to order three in some interval $|x| \leqslant a$, $a > 0$. We call $g_n(x)$ the product of the first n of them:

$$g_n(x) = f_1(x) f_2(x) \cdots f_n(x). \tag{1}$$

We are interested in the asymptotic behavior of g_n under certain assumptions. The first group of assumptions about the $f_\nu(x)$ will be

$$f_\nu(0) = 1, \quad f_\nu'(0) = 0, \quad f_\nu''(0) = -r_\nu^2, \qquad \nu = 1, 2, \dots, \tag{A}$$

where r_ν^2 is some positive quantity. For real f_ν, these are well-known sufficient conditions for a maximum (a vertex) of magnitude 1.

We make the additional assumptions that for all ν

$$0 < r^2 \leqslant r_\nu^2 \leqslant R^2, \qquad \text{and} \qquad |f_\nu'''(x)| \leqslant A \quad \text{in} \quad |x| \leqslant a, \qquad a > 0; \tag{B}$$

[1] P. S. de Laplace, *Théorie analytique des probabilités*, Vol 7, p. 89. Œuvres, Paris, 1886.

this means that the r_ν^2 lie, for all ν, between two finite positive numbers and that the third derivatives of all $f_\nu(x)$, for all $|x| \leqslant a$, are bounded by a single number A. It is then easily seen [by successive integration from 0 to x of $f_\nu'''(x)$ and use of (B)] that the $f_\nu''(x)$, $f_\nu'(x)$, $f_\nu(x)$ are likewise uniformly bounded, for all ν, in $|x| \leqslant a$; in particular, we can find an interval $|x| \leqslant b$, $0 < b \leqslant a$, such that $|f_\nu(x) - f_\nu(0)|$ is less than $\frac{1}{2}$, say, and therefore

$$\tfrac{1}{2} \leqslant |f_\nu(x)| \leqslant \tfrac{3}{2} \qquad \text{in} \quad |x| \leqslant b. \tag{2}$$

Denote by $F_\nu(x)$ the logarithm of $f_\nu(x)$; if $f_\nu(x)$ is complex, we choose the principal value (i.e., the value of log x which is zero for $x = 1$). The first three derivatives of $F_\nu(x)$ are

$$F_\nu' = \frac{f_\nu'}{f_\nu}, \qquad F_\nu'' = \frac{f_\nu''}{f_\nu} - \frac{f_\nu'^2}{f_\nu^2}, \qquad F_\nu''' = \frac{f_\nu'''}{f_\nu} - 3\frac{f_\nu'' f_\nu'}{f_\nu^2} + 2\frac{f_\nu'^3}{f_\nu^3}$$

and

$$F_\nu(0) = 0, \qquad F_\nu'(0) = 0, \qquad F_\nu''(0) = -r_\nu^2.$$

From the uniform boundedness of the $f_\nu', f_\nu'', f_\nu'''$ in $|x| \leqslant b$ and the first inequality (2), it follows that the above derivatives of F_ν are uniformly bounded there and that, in particular, for $F_\nu'''(x)$ and some positive S:

$$|F_\nu'''(x)| < S, \qquad \text{for all} \quad \nu, \quad \text{in} \quad |x| \leqslant b. \tag{3}$$

We write the Taylor formula of $F_\nu(x)$ in $|x| \leqslant b$ [for the real and imaginary parts of $F_\nu(x)$, respectively] up to terms of order three and obtain

$$F_\nu(x) = 0 + 0 - \frac{x^2}{2} r_\nu^2 + \frac{x^3}{6}\{\mathrm{Re}[F_\nu'''] + i\,\mathrm{Im}[F_\nu''']\}, \tag{4}$$

where the arguments of the real and imaginary parts of F_ν''' are not necessarily the same but are certainly between 0 and b; hence, in $|x| \leqslant b$, and with the abbreviation[2]

$$r_1^2 + r_2^2 + \cdots + r_n^2 = s_n^2, \tag{5}$$

we obtain from (4), using (3) and (1)

$$|F_1(x) + \cdots + F_n(x) + \tfrac{1}{2} x^2 s_n^2| = |\log g_n(x) + \tfrac{1}{2} x^2 s_n^2| \leqslant \frac{|x|^3}{3} nS. \tag{6}$$

[2] In this and the following chapter we use a and s^2 for mean value and variance without suggesting sample mean and sample variance. As long as there is no contrasting of empirical and theoretical distributions, Latin letters are used for simplicity.

If then $|x| < \gamma/n^\beta$, where $\gamma > 0, \beta > \frac{1}{3}$, we have, uniformly for these x:

$$\lim_{n\to\infty} | F_1(x) + F_2(x) + \cdots + F_n(x) + \tfrac{1}{4} x^2 s_n^2 | = 0. \tag{6'}$$

Now, introducing the new variable z by

$$s_n x = z, \qquad x = \frac{z}{s_n}, \tag{7}$$

this becomes, according to the first condition (B) and (3),

$$\left| \log g_n\left(\frac{z}{s_n}\right) + \frac{z^2}{2} \right| < \left| \frac{z^3}{3} \right| \frac{nS}{n\sqrt{nr^3}} < \frac{Z^3}{3} \frac{S}{r^3} \frac{1}{\sqrt{n}} \qquad \text{for} \quad |z| < Z. \tag{6''}$$

Since the last term goes to zero with increasing n, we have

$$\lim_{n\to\infty} \log\left[g_n\left(\frac{z}{s_n}\right) e^{z^2/2}\right] = 0.$$

or (8)

$$\lim_{n\to\infty} g_n\left(\frac{z}{s_n}\right) = e^{-z^2/2},$$

uniformly for all $|z| < Z$. This is the first result.

If in (A) a number α is used instead of zero, (7) is replaced by $x = \alpha + (z/s_n)$ and the result may be worded as follows:

The product of n functions $f_\nu(\alpha + z/s_n)$, *each of which has a regular maximum equal to* 1 *at* $x = \alpha$, *and satisfies conditions* (B) *with x replaced by* $x - \alpha$, *tends, as* $n \to \infty$, *toward the Gaussian* $e^{-z^2/2}$ *in the sense of Eq.* (8).

Note that, from (6) $\lim_{n\to\infty} | \log g_n(x) + \frac{1}{2}x^2 s_n^2 | = 0$ has been proved for $|x - \alpha|^3 n < \gamma_1/n^{\epsilon_1}$, where γ_1 and ϵ_1 are positive, and this inequality holds only in a small neighborhood of $x = \alpha$.

Consider the particular case of n functions $f_\nu(x)$ which are real and identical. The MacLaurin expansion of a function $f(x)$ which has continuous derivatives up to the nth order at $x = 0$ can be written

$$f(x) = \sum_{\nu=0}^{n} \frac{f^{(\nu)}(0)}{\nu!} x^\nu + R_n(x),$$

where

$$R_n(x) = \frac{f^{(n)}(\theta x) - f^{(n)}(0)}{n!} x^n, \qquad 0 < \theta < 1.$$

Because of the continuity of the nth derivative, $f^{(n)}(\theta x) - f^{(n)}(0)$ tends to zero with x, and therefore $R_n(x) = o(x^n)$; i.e., $R_n(x)$ is of a lower order than x^n. If $f(x)$ is complex, the same conclusion holds for the real and the imaginary parts separately and the sum of the remainders is still of the same order of magnitude. If $f(x)$ satisfies condition (A), we have with $r_\nu^2 = r^2$,

$$f(x) = 1 - \frac{r^2}{2} x^2 + o(x^2).$$

Now, $s_n^2 = nr^2$, and with $x = z/s_n = z/r\sqrt{n}$:

$$f\left(\frac{z}{s_n}\right) = 1 - \frac{z^2}{2n} + o\left(\frac{1}{n}\right),$$

$$g_n\left(\frac{z}{s_n}\right) = \left[f\left(\frac{z}{s_n}\right)\right]^n$$

and by (6′) this tends uniformly for finite z toward $e^{-z^2/2}$. Hence, *in the case of equal functions $f_\nu(x) = f(x)$ the result*

$$\lim_{n\to\infty} g_n\left(\frac{z}{r\sqrt{n}}\right) = e^{-z^2/2}, \qquad z = xr\sqrt{n} \tag{8′}$$

holds under the conditions $f(0) = 1, f'(0) = 0, f''(0) = -r^2$ if the second derivative is continuous at $x = 0$.

1.2. *Integration of the product formula.* We return to the case of different $f_\nu(x)$. It follows from (8) that if $\psi(z)$ is a complex continuous and bounded function of z, $|\psi(z)| < \Psi$, and if a, b are any real numbers between $-Z$ and Z, then

$$\lim_{n\to\infty} \int_a^b \psi(z) g_n\left(\frac{z}{s_n}\right) dz = \int_a^b \psi(z) e^{-z^2/2}\, dz, \tag{9}$$

uniformly for all $|z| \leqslant Z$. This is almost obvious since with η_0 an arbitrarily small number, we have, on account of (8),

$$\left| g_n\left(\frac{z}{s_n}\right) - e^{-z^2/2} \right| < \eta_0 \qquad \text{for sufficiently large } n \text{ and } \quad |z| < Z,$$

and therefore

$$\left| \int_a^b \psi(z) \left[g_n\left(\frac{z}{s_n}\right) - e^{-z^2/2}\right] dz \right| < \Psi(b-a)\,\eta_0 = \epsilon. \tag{10}$$

This argument, however, is not applicable if a or b or both are infinite and this is the case that will be needed. To prove (9) for an infinite range of integration, we assume that $\psi(x)$ is bounded in $-\infty < x < +\infty$ and we have to introduce some restrictions on the behavior of the $f_\nu(x)$ over the entire range. [Conditions (A) and (B) apply only to $x = 0$ and its neighborhood.] We now assume that $f_\nu(x)$ never approaches the value 1 at a point other than $x = 0$; that for large $|x|$, $f_\nu(x)$ tends to zero like some negative power of $|x|$; and that these restrictions hold uniformly for all ν. This is expressed as follows: for any $\delta > 0$, there exists a positive η such that

$$|f_\nu(x)| < 1 - \eta \qquad \text{if} \quad |x| > \delta; \qquad \nu = 1, 2, \dots ; \tag{C}$$

and: for some $\lambda > 0$, a (large) value X can be found such that

$$|f_\nu(x)| < |x|^{-\lambda} \qquad \text{if} \quad |x| > X; \qquad \lambda > 0; \quad \nu = 1, 2, \dots . \tag{C'}$$

Under these conditions, we shall prove that

$$\lim_{n\to\infty} \left[\int_{-\infty}^{+\infty} g_n\left(\frac{z}{s_n}\right) \psi(z)\, dz\right] = \int e^{-z^2/2}\, \psi(z)\, dz. \tag{9'}$$

It will be sufficient to prove (9) for $a = 0$, $b \to \infty$. The integration limit $-\infty$ can be treated in the same way.

We first choose a finite value Z of z large enough that, with $\Phi(x)$ denoting as always the Gaussian $\Phi(x) = (1/\sqrt{2\pi}) \int^x e^{-z^2/2}\, dz$:

$$\sqrt{2}\,\sqrt{2\pi}\, \Psi \left[1 - \Phi\left(\frac{Z}{\sqrt{2}}\right)\right] < \frac{\epsilon}{8} \tag{11}$$

for a given, arbitrarily small $\epsilon > 0$. That such a Z can be found for any ϵ follows from the fact that the Gauss integral $\Phi(x)$ approaches 1 as x increases. Then with this finite Z, we take $a = 0$, $b = Z$ in (10), and determine $\eta_0 = (\epsilon/4)\Psi Z$ and n so large that

$$\left| \int_0^Z \psi(z) \left[g_n\left(\frac{z}{s_n}\right) - e^{-z^2/2}\right] dz \right| < \Psi Z \eta_0 = \frac{\epsilon}{4}. \tag{12}$$

Next, we have, using (11),

$$\left| \int_Z^\infty \psi(z)\, e^{-z^2/2}\, dz \right| < \Psi \int_Z^\infty e^{-z^2/2}\, dz = \Psi\, \sqrt{2\pi}[1 - \Phi(Z)] < \frac{\epsilon}{8}. \tag{13}$$

There remains only the $\int_Z^\infty \psi(z)\, g_n(z/s_n)\, dz$. It will be divided into three parts I_1, I_2, I_3: (1) from Z to $s_n\delta$, (2) from $s_n\delta$ to s_nX, (3) from s_nX to infinity.[3] Here X is the quantity appearing in condition (C') and δ the

[3] Note that by choosing n large enough we can always make $s_n\delta > Z$.

one appearing in (C). To find an appropriate δ for I_1 we return to (6):

$$|f_1(x)f_2(x)\cdots f_n(x)| \leqslant \exp[-\tfrac{1}{2}s_n^2x^2 + \tfrac{1}{3}nSx^3]$$

$$= \exp\left[-\tfrac{1}{2}s_n^2x^2\left(1 - \frac{2}{3}\frac{nxS}{s_n^2}\right)\right] = \exp[-\tfrac{1}{2}s_n^2x^2(1-c)],$$

valid in $(0, b)$ where b is the number introduced in (3). We wish to choose x small enough so that $c \leqslant \frac{1}{2}$. Now

$$c = \frac{2}{3}\frac{nxS}{s_n^2} \leqslant \frac{1}{2} \quad \text{if} \quad x \leqslant \frac{3}{4}\frac{s_n^2}{nS},$$

and this holds *a fortiori* if

$$x \leqslant \frac{3}{4}\frac{nr^2}{nS} = \frac{3}{4}\frac{r^2}{S} \equiv \delta.$$

Hence, for $x \leqslant \delta$, or $z \leqslant s_n\delta$, we have $c \leqslant \frac{1}{2}$, or $1 - c \geqslant \frac{1}{2}$, and therefore

$$\left| g_n\left(\frac{z}{s_n}\right) \right| \leqslant e^{-z^2/4}.$$

This supplies by means of (11)

$$|I_1| < \int_Z^\infty |\psi(z)|\, e^{-z^2/4}\, dz < \Psi \cdot 2\sqrt{\pi}\left[1 - \Phi\left(\frac{Z}{\sqrt{2}}\right)\right] < \frac{\epsilon}{8}. \tag{14}$$

In the next interval we make use of condition (C) since here the argument of g_n is $\geqslant \delta$. Thus $|f_\nu| < 1 - \eta$ and $|g_n| < (1-\eta)^n$, which with the use of (B) leads to

$$|I_2| = \left| \int_{s_n\delta}^{s_nX} g_n\left(\frac{z}{s_n}\right)\psi(z)\, dz \right| < (1-\eta)^n\Psi\, s_n\, |X - \delta|$$

$$< \sqrt{n}\,(1-\eta)^n\Psi R\, |X - \delta|. \tag{15}$$

Since $\sqrt{n}(1-\eta)^n$ goes to zero with increasing n this means n can be chosen large enough to have

$$|I_2| < \frac{\epsilon}{4}. \tag{15'}$$

Finally, in the last interval we use condition (C′) since here the argument of g_n is greater than X:

$$\left| g_n\left(\frac{z}{s_n}\right) \right| = |g_n(x)| < |x|^{-n\lambda},$$

and with $x = z/s_n$ as the integration variable, and remembering condition (B):

$$|I_3| < \left| \int_X^\infty \psi(s_n x) x^{-n\lambda} s_n \, dx \right| < \Psi R \frac{\sqrt{n}}{n\lambda - 1} X^{-n\lambda+1}. \tag{16}$$

Again, the right-hand side goes to zero with increasing n, that is,

$$|I_3| < \frac{\epsilon}{4} \tag{16'}$$

for sufficiently large n.

Now, combining (12)–(14), (15′), and (16′), we have

$$\left| \int_0^\infty \psi(z) g_n\left(\frac{z}{s_n}\right) dz - \int_0^\infty \psi(z) e^{-z^2/2} \, dz \right| \leqslant \frac{\epsilon}{4} + \frac{\epsilon}{8} + \frac{\epsilon}{8} + \frac{\epsilon}{4} + \frac{\epsilon}{4} = \epsilon,$$

which proves that the absolute value of the difference under consideration is below any arbitrarily chosen positive number as n increases. The same is true if both integrals are taken from $-\infty$ to $+\infty$.

Moreover, nothing changes if $\psi = \psi_n$ depends on n, *provided that an upper bound Ψ, independent of n, exists for all ψ_n*. The result can be roughly formulated in this way:

Under the conditions (A), (B), (C), (C′) *the product g_n of n functions f_ν is approximated for large n by the Gaussian function $e^{-z^2/2}$ in such a way that for a bounded ψ the integral of ψg_n over an infinite interval converges toward the integral of $\psi e^{-z^2/2}$*:

$$\lim_{n\to\infty} \left| \int \psi(z) g_n\left(\frac{z}{s_n}\right) dz - \int \psi(z) e^{-z^2/2} \, dz \right| = 0. \tag{17}$$

Here, $-s_n^2$ is the sum of the second derivatives of the f_ν at $x = 0$.

Problem 1. Let all $f_\nu(x)$ be equal to $f(x)$ with a regular maximum $f(\alpha)$ at the point $x = \alpha$. Prove that (8) is replaced by the asymptotic formula

(a) $\quad g_n\left(\alpha + \dfrac{z}{s_n}\right) \sim [f(\alpha)]^n e^{-z^2/2f(\alpha)} \qquad$ where $\qquad s_n^2 = -nf''(\alpha)$,

or, in terms of the variable $x = \alpha + z/s_n$:

(b) $\quad g_n(x) \sim [f(\alpha)]^n e^{-s_n^2(x-\alpha)^2/2f(\alpha)}$.

Problem 2. With the notations of Problem 1, let α be an inner point of the finite or infinite interval (a, b). Prove that, asymptotically with s_n^2 as above

$$\text{(c)} \qquad \int_a^b g_n(x)\,dx \sim \frac{\sqrt{2\pi}}{s_n}[f(\alpha)]^{n+\frac{1}{2}}.$$

Problem 3. Let α be an inner point of (a, b) and consider n functions $f_1(x)$, $f_2(x)$, ..., $f_n(x)$, each having a regular maximum $f_\nu(\alpha)$ at $x = \alpha$. Prove that, under the hypotheses of the text, formula (c), Problem 2, is replaced by the asymptotic formula

$$\text{(d)} \qquad \int_a^b f_1(x)f_2(x)\cdots f_n(x)\,dx \sim \frac{\sqrt{2\pi}}{s_n} f_1(\alpha)f_2(\alpha)\cdots f_n(\alpha),$$

with

$$s_n^2 = -\sum_{\nu=1}^{n} \frac{f_\nu''(\alpha)}{f_\nu(\alpha)}.$$

Problem 4. Using the formula $\int_0^{\pi/2} \sin^{2n}x\,dx = \frac{1\cdot3\cdot5\cdots(2n-1)}{2\cdot4\cdots2n}\pi$ show that

$$\frac{1\cdot3\cdot5\cdot\cdots\cdot(2n-1)}{2\cdot4\cdot6\cdots2n} \sim \frac{1}{\sqrt{n\pi}}.$$

Problem 5. Let $\lambda > 1$, real, and

$$P_n(\lambda) = \frac{1}{2\pi}\int_{-\pi}^{\pi} (\lambda + \sqrt{\lambda^2 - 1}\cos x)^n\,dx.$$

Prove that, asymptotically for large n

$$P_n(\lambda) \sim \frac{1}{\sqrt{2n\pi}}\,\frac{(\lambda + \sqrt{\lambda^2-1})^{n+\frac{1}{2}}}{\sqrt[4]{\lambda^2 - 1}}.$$

Problem 6. Find an asymptotic expression for large n for

$$\int_{-k}^{k} (k^2 - x^2)^n\,dx.$$

Problem 7. How would the result (8) be modified if for all f_ν the maximum at $x = 0$ were such that

$$f_\nu(0) = 1,\ f_\nu'(0) = f_\nu''(0) = f_\nu'''(0) = 0; \qquad f_\nu^{(\text{iv})}(0) = -r_\nu^4?$$

2. Application of the Product Formulas

2.1. *Stirling's formula.* As an application of the integral formula for an infinite product developed in the foregoing section, we shall derive an asymptotic expression for $n! = 1 \cdot 2 \cdot \cdots \cdot n$.

If integration by parts is applied to the integral

$$I_n = \int_0^\infty t^n e^{-t}\,dt, \tag{18}$$

one obtains

$$I_n = -t^n e^{-t}\Big|_0^\infty + n\int_0^\infty t^{n-1}e^{-t}\,dt = nI_{n-1}\,. \tag{18'}$$

Since $I_0 = \int_0^\infty e^{-t}\,dt = 1$, the repeated application of (18′), with n a positive integer or zero, yields

$$I_n = n!\,. \tag{19}$$

The introduction of the integration variable x by

$$t = n(1+x), \qquad dt = n\,dx$$

leads to

$$I_n = \int_{-1}^\infty n^n(1+x)^n\, e^{-n(1+x)}\, n\,dx = n^{n+1}e^{-n}\int_{-1}^\infty [(1+x)\,e^{-x}]^n\,dx. \tag{20}$$

The new integrand $[(1+x)\,e^{-x}]^n$ has the form of a product of n identical functions and permits us to apply the integral formula (17) provided that the conditions (A), (B), (C), and (C′) are fulfilled. In the notation of Section 1, we identify for all ν

$$f_\nu(x) = f(x) = \begin{cases}(1+x)e^{-x}, & x \geqslant -1\\ 0, & x < -1\end{cases}$$

and find

$$f' = -xe^{-x}, \qquad f'' = (x-1)e^{-x}, \qquad f''' = -(x-2)e^{-x}; \qquad x \geqslant -1 \tag{21}$$

At $x = 0$ we have

$$f(0) = 1, \qquad f'(0) = 0, \qquad f''(0) = -1, \qquad r_\nu^2 = 1 \qquad \text{for all } \nu. \tag{22}$$

Condition (A) and the first condition (B) is therefore fulfilled. The third derivative is seen to be bounded and the second condition (B) is fulfilled.

Condition (C) is fulfilled, because f' is negative for all positive x and positive for all negative x. Finally, for large negative x the function is identically zero and for $x = +\infty$ it vanishes exponentially, i.e., faster than any negative power of x.

Formula (17) may therefore be applied to the integral in (20), the function ψ being identically 1, and $s_n = \sqrt{\sum_{\nu=1}^{n} r_\nu^2} = \sqrt{n}$. Writing, with $s_n x = z$,

$$\int_{-1}^{\infty} [(1 + x)e^{-x}]^n \, dx = \int_{-1}^{\infty} [f(x)]^n \, dx = \frac{1}{\sqrt{n}} \times \int_{-\infty}^{\infty} \left[f\left(\frac{z}{\sqrt{n}}\right)\right]^n dz = \frac{1}{\sqrt{n}} \int_{-\infty}^{\infty} g_n\left(\frac{z}{\sqrt{n}}\right) dz, \tag{23}$$

we obtain from (17)

$$\lim_{n\to\infty} \sqrt{n} \int_{-1}^{\infty} [(1 + x)e^{-x}]^n \, dx = \int_{-\infty}^{\infty} e^{-z^2/2} \, dz. \tag{24}$$

Since the last integral has the value $\sqrt{2\pi}$, and the integral to the left, according to (19) and (20), equals $n!/(n^{n+1}e^{-n})$, Eq. (24) is equivalent to

$$\lim_{n\to\infty} \frac{n!\sqrt{n}}{n^{n+1}e^{-n}} = \sqrt{2\pi}\,. \tag{25}$$

Thus, an asymptotic formula has been obtained, which approximates $n!$ for large values of n by functions well suited to logarithmic computations. One may write Eq. (25), *Stirling's formula*, in the form

$$n! \sim \sqrt{2\pi} \cdot n^{n+\frac{1}{2}} e^{-n}, \qquad \text{or} \qquad \log n! \sim n \log n - n + \log\sqrt{2\pi n} \tag{26}$$

but the exact meaning of (26) is given by (25).[1]

2.2. *Asymptotic expression for the Bernoulli distribution.* The principal use made of the Stirling formula in probability calculus consists of transforming the Newton binomial formula, (11) in Chapter IV, which was the solution of the Bernoulli problem, into a form suitable for computation for large values of n. The following derivation makes use of the particular analytic form of $p_n(x)$. Writing now $q_n(x)$ instead of $p_n(x)$, we have for the probability of x events in n trials (sum x in n alternatives of zero or one):

$$q_n(x) = \frac{n!}{x!(n - x)!} p^{n-x} q^x, \qquad p + q = 1; \qquad p, q \neq 0. \tag{27}$$

[1] Stirling's formula is remarkably accurate. For $10! = 3{,}628{,}800$ the approximation is 3,598,600, an error of 0.8%.

Using the logarithmic form of (26) for $n!$, $x!$, and $(n - x)!$, and assuming that not only n, but also x and $n - x$ go to infinity, we find

$$\log q_n(x) \sim n \log n - x \log \frac{x}{q} - (n - x) \log \frac{n - x}{p} + \frac{1}{2} \log \frac{n}{x(n - x) \cdot 2\pi}$$

$$= -x \log \frac{x}{qn} - (n - x) \log \frac{n - x}{pn} + \frac{1}{2} \log \frac{n}{2\pi x(n - x)} . \tag{28}$$

The mean value nq and the variance npq of $q_n(x)$ both tend to infinity with n. If we wish to find a limit of $q_n(x)$ we must use a new variable with respect to which both mean value and variance have finite values. We introduce a new variable u which will first be considered bounded, $(|u| < U)$, connected with x by

$$x = nq + u\sqrt{npq} = nq\left(1 + \frac{u}{\sqrt{n}}\sqrt{\frac{p}{q}}\right). \tag{29}$$

This new variable u is the standardized variable with mean value zero and variance one. The origin is thus translated by nq and the unit of abscissas is diminished so that to the unit interval between two x-values there corresponds an interval $h = 1/\sqrt{npq}$ in u. This change of unit is illustrated in Fig. 9 of Chapter IV for the binomial distributions with $q = \frac{1}{3}$ and $n = 4, 9, 16, 36$, and 100, respectively.

From (29) it follows that

$$n - x = np - u\sqrt{npq} = np\left(1 - \frac{u}{\sqrt{n}}\sqrt{\frac{q}{p}}\right); \tag{29'}$$

it is seen that for $|u| < U$ both x and $n - x$ and also $|x - nq|$ become infinite with n while, at the same time, the last terms in the parentheses (29) and (29') tend toward zero as $1/\sqrt{n}$. If we apply the development of $\log(1 + z) = z - z^2/2 + \cdots$ and use (29) we have

$$\log \frac{x}{nq} = \log\left[\left(1 + \frac{u}{\sqrt{n}}\sqrt{\frac{p}{q}}\right)\right] \sim \frac{u}{\sqrt{n}}\sqrt{\frac{p}{q}} - \frac{u^2}{2n}\frac{p}{q}, \tag{30}$$

where we omit only a remainder term of order $(u/\sqrt{n})^3$. Multiplying (30) by x, we arrive at

$$x \log \frac{x}{nq} \sim u\sqrt{npq} + \frac{u^2}{2} p, \tag{31}$$

where a term has been neglected that vanishes as $u^3/\sqrt{n}$. In the same way $\log[(n-x)/p]$ and its product by $n-x$ can be computed:

$$(n-x)\log\frac{n-x}{np} \sim -u\sqrt{npq} + \frac{u^2}{2}q. \tag{31'}$$

The sum of (31) and (31′) which appears in expression (28) with the negative sign is seen to be simply $(p+q)(u^2/2) = u^2/2$, and thus from (28)

$$\log q_n(x) \sim \frac{-u^2}{2} + \frac{1}{2}\log\frac{n}{2\pi x(n-x)}, \qquad q_n(x) \sim \sqrt{\frac{n}{2\pi x(n-x)}}\, e^{-u^2/2}. \tag{32}$$

We obtain from (32)

$$\lim_{n\to\infty}\sqrt{\frac{2\pi x(n-x)}{n}}\, q_n(x) = e^{-u^2/2} \qquad \text{for} \quad |u| < U. \tag{33}$$

In our derivation, we had to assume that $u^3/\sqrt{n} \to 0$, i.e., $(x-nq)^3/n^2 \to 0$. Then, *a fortiori*, $u/\sqrt{n} = (x-nq)/n \to 0$. Now, from (29) and (29′) we have $x(n-x)/n^2 \to pq$, and therefore (33) can be written as the limit formula

$$\lim_{n\to\infty}\sqrt{2\pi npq}\, q_n(nq + u\sqrt{npq}) = e^{-u^2/2}, \tag{34}$$

or, reintroducing x instead of u,

$$q_n(x) \sim \frac{1}{\sqrt{2\pi npq}} \exp\left(-\frac{(x-nq)^2}{2npq}\right). \tag{35}$$

This can be used as an approximation to $q_n(x)$ for large values of n, x, and $n-x$, if u is bounded or even if only

$$\frac{u^3}{\sqrt{n}} = \frac{(x-nq)^3}{n^2} \to 0. \tag{36}$$

The statement expressed in Eq. (34) is that the right-hand limit is approached uniformly for all $|u| \leqslant U$, and that (34) holds even for u satisfying (36). Formula (35) was first given by Laplace (1812) and is known as *Laplace's solution of the Bernoulli problem.*

We give a numerical example. With an unbiased die we make $n = 1200$ throws and consider the probability $q_n(x)$ of obtaining the six 180 times.

Here $q = \frac{1}{6}$, $nq = 200$. The Bernoulli formula gives $q_{1200}(180) = \binom{1200}{180}(\frac{5}{6})^{1020}(\frac{1}{6})^{180} = 0.009387$. The approximation formula (35) gives

$$q_{1200}(180) \sim \frac{\exp\left(-\dfrac{400 \cdot 360}{5 \cdot 2400}\right)}{\sqrt{2\pi \cdot 1200 \cdot 5/36}} = \frac{0.3}{\sqrt{30\pi}} e^{-1.2} = 0.009308.$$

The error is less than 1 %.

In many cases, however, it is much more interesting to know the probability that a result be between two bounds x_1 and x_2; in our example we may ask for the probability of obtaining a result between 180 and 220. This probability is equal to

$$\sum_{x=180}^{220} q_{1200}(x),$$

hence to a sum of 41 values. To simplify this computation we introduce $h = 1/\sqrt{npq}$ and write (34) in the form

$$q_n(x) \sim \frac{1}{\sqrt{2\pi npq}} e^{-u^2/2} = h\phi(u). \tag{34'}$$

Replacing $q_n(x)$ by $h\phi(u)$ we then approximate the sum of the $h\phi(u)$ by a definite integral. We see that with $h = 1/s_n$, $u_1 = h(x_1 - nq)$, $u_2 = h(x_2 - nq)$ we have

$$\sum_{x_1}^{x_2} q_n(x) \sim \sum_{u_1}^{u_2} h\phi(u) \sim \int_{u_1-(h/2)}^{u_2+(h/2)} \phi(t)\, dt.$$

Here $\Sigma_{x_1}^{x_2} q_n(x)$ was first replaced, by means of (34′), by $\Sigma_{u_1}^{u_2} h\phi(u)$, the sum of 41 rectangles of basis h and height $\phi(u)$; next the sum of these 41 rectangles was approximated by the area under the $\phi(u)$ curve, given by the integral to the right; the limits are $u_1 - (h/2)$ and $u_2 + (h/2)$ [since the abscissas of midpoints of the rectangles are u_1, $u_1 + h$, ..., u_2, respectively], and therefore the first rectangle is between $u_1 - (h/2)$ and $u_1 + (h/2)$, etc. (Fig. 14). We have thus obtained the *de Moivre-Laplace* theorem in the form

$$\sum_{x_1}^{x_2} q_n(x) \sim \int_{u_1-(h/2)}^{u_2+(h/2)} \phi(u)\, du = \Phi\left(u_2 + \frac{h}{2}\right) - \Phi\left(u_1 - \frac{h}{2}\right). \tag{37}$$

Since for large n, h is small compared to u, the $h/2$ in the arguments of Φ is often neglected; in this case, we may make up for the error by adding to $\Phi(u_2) - \Phi(u_1)$ the term $(h/2)(\phi(u_1) + \phi(u_2))$. At any rate, if h is not very small compared to u, then neither (34′) [derived under the condition $(u^3/\sqrt{n}) \to 0$] nor replacement of the sum by the integral is justified. The problem of the error has been carefully investigated.[2]

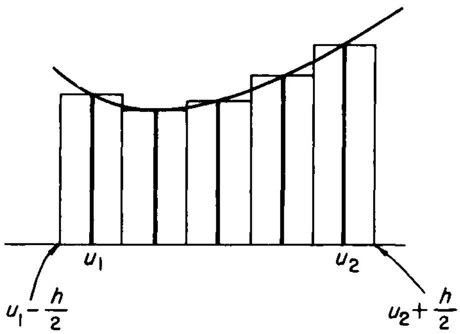

FIG. 14. Numerical integration.

The result in our numerical example is $\Sigma_{x_1}^{x_2} q_n(x) \sim 0.879$. We defer further discussion to the next section where it will be seen that our results have a much wider scope.

Problem 8. Compute the probability of $x = 10$ heads in twenty tossings of an unbiased coin from the Newton formula, and compare it with the approximation given by the Laplace formula; do the same for $x = 1$ heads. (log 10! = 6.55976, log 20! = 18.38612).

B. Limit Distribution of the Sum of Independent Discrete Random Variables (Sections 3 and 4)

3. *Arithmetical Probabilities*

3.1. *Local limit theorem.* Assume an infinite set of collectives with arithmetical probabilities $p_\nu(x)$, $\nu = 1, 2, \dots$. The label set of each of the $p_\nu(x)$ is assumed to be the set of integers $0, \pm 1, \pm 2, \dots$. Of course, the

[2] See for example S. N. BERNSTEIN, "Rückkehr zu der Frage der Genanigkeit der Laplace'schen Grenzwertformel." *Isw. Akad. Nauk. SSSR* **7** (1943); W. FELLER, "On the normal approximation to the binomial distribution." *Ann. Math. Statist.* **16** (1945), pp. 319–29; and J. PRATT, "Approximating the binomial cumulative distribution." to appear.

probabilities for certain labels may vanish for some $p_\nu(x)$. Denote the c.f. of $p_\nu(x)$ by $f_\nu{}^*(u)$, so that

$$f_\nu{}^*(u) = \sum_x e^{iux} p_\nu(x),$$

and by Eq. (62), Chapter V:

$$p_\nu(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-iux} f_\nu{}^*(u)\, du.$$

We now write x instead of u and use t as the summation variable, in order to have a notation similar to that of Section 1, and define the functions $f_\nu(x)$ by

$$\begin{aligned} f_\nu(x) &= e^{-ia_\nu x} f_\nu{}^*(x) = e^{-ia_\nu x} \sum_t e^{ixt} p_\nu(t) \\ &= \sum_t e^{ix(t-a_\nu)} p_\nu(t), \end{aligned} \tag{38}$$

with a_ν denoting the mean value of t relative to $p_\nu(t)$. (The f_ν are the c.f.'s for the chance variables $t - a_\nu$.)

With $q_n(x)$ denoting again the probability of drawing the sum x from the first n collectives, the inversion formula, V, Eq. (62) together with the multiplication theorem, V, Eq. (55′), gives

$$\begin{aligned} q_n(x) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ixt} \prod_{\nu=1}^{n} f_\nu{}^*(t)\, dt \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ixt+ib_n t} \prod_{\nu=1}^{n} f_\nu(t)\, dt, \end{aligned} \tag{39}$$

where b_n has been written for $a_1 + a_2 + \cdots + a_n$.

In order to obtain an asymptotic expression for $q_n(x)$, we apply to (39) the integral formula (17). The integral from $-\pi$ to π in (39) can be formally written with the limits $-\infty$ and ∞ if we define functions $f_\nu(x)$ by (38) for $-\pi \leqslant x \leqslant \pi$, and set $f_\nu(x) \equiv 0$ outside that interval, and identify $e^{it(b_n - x)}$ with the function ψ appearing in (17), where $e^{it(b_n-x)}$ is certainly bounded.[1] From (38) we obtain, with $|x| \leqslant \pi$,

$$\begin{aligned} f_\nu(x) &= \sum_t e^{ix(t-a_\nu)} p_\nu(t), & f_\nu(0) &= \sum_t p_\nu(t) = 1; \\ f_\nu{}'(x) &= \sum_t i(t-a_\nu)\, e^{ix(t-a_\nu)} p_\nu(t); & f_\nu{}'(0) &= i\sum_t (t-a_\nu) p_\nu(t) = 0; \\ f_\nu''(x) &= -\sum_t (t-a_\nu)^2\, e^{ix(t-a_\nu)} p_\nu(t); & f_\nu''(0) &= -\sum_t (t-a_\nu)^2 p_\nu(t) = -r_\nu{}^2. \end{aligned} \tag{40}$$

[1] Here ψ depends on n (see the remark on p. 269) and on x as parameter, but each ψ is integrable and all of them have the upper bound 1.

One sees that condition (A) of Section 1 is fulfilled. To comply with the first condition (B), the assumption

$$r^2 \leqslant r_\nu^2 \leqslant R^2 \tag{41}$$

must be made about the variances r_ν^2 of the distributions $p_\nu(x)$. As a consequence of (41), $\lim_{n\to\infty} s_n^2/n$ is finite, where s_n^2 again stands for $r_1^2 + r_2^2 + \cdots + r_n^2$. The second condition (B) will be fulfilled if all $f_\nu'''(x)$ have an upper bound independent of ν. Since

$$f_\nu'''(x) = -i \sum_t (t - a_\nu)^3 \, e^{ix(t-a_\nu)} \, p_\nu(t), \tag{42}$$

it follows that $| f_\nu'''(x) | \leqslant \Sigma \, | \, t - a_\nu \, |^3 \, p_\nu(t)$. The assumption will therefore be made that the third absolute moments of the distributions have a common upper bound:

$$T_\nu = \sum_t | \, t - a_\nu \, |^3 p_\nu(t) \leqslant T, \qquad \nu = 1, 2, 3, \ldots . \tag{43}$$

Condition (C′) is obviously fulfilled ($f_\nu = 0$ for $| \, x \, | > \pi$). Condition (C) requires some investigation which will be taken up later.

Assuming for the present that all conditions (A), (B), (C), (C′) are fulfilled, we replace t in (39) by z/s_n . The integral formula (17) gives then, if we use again the abbreviation $\Pi_{\nu=1}^n f_\nu(x) = g_n(x)$:

$$\lim_{n\to\infty} \left[\int_{-\infty}^{\infty} \exp\left(-i \frac{z}{s_n} (x - b_n)\right) \cdot g_n\left(\frac{z}{s_n} - e^{-z^2/2}\right) dz \right], \tag{44}$$

uniformly for all x. By (39) and the definition of the $f_\nu(x)$, the (44) integral is equal to $2\pi \lim_{n\to\infty} q_n(x) \cdot s_n$; the second integral in first becomes with $(x - b)_n/s_n = u$:

$$\int_{-\infty}^{+\infty} \exp\left(-\frac{z^2}{2} - iz \frac{x - b_n}{s_n}\right) dz = e^{-u^2/2} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}(z+iu)^2} \, dz = \sqrt{2\pi} \, e^{-u^2/2}, \tag{45}$$

and we obtain

$$\lim_{n\to\infty} s_n \cdot q_n(b_n + us_n) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}, \qquad \text{uniformly for} \quad | \, u \, | < U. \tag{46}$$

To the left stands a function defined for all u and identical with the distribution of the sum $x_1 + \cdots + x_n$ for integral values of the argument. Written in terms of the chance variable x, as an asymptotic approximation for large n, the formula reads:

$$q_n(x) \sim \frac{1}{\sqrt{2\pi} s_n} \exp\left(-\frac{(x - b_n)^2}{2s_n^2}\right). \tag{46'}$$

Equation (46) expresses a fundamental theorem of probability calculus. *The distribution of the sum of independent chance variables* $x_1, x_2, \ldots, x_n$, *each of which is subject to an arithmetic distribution, tends, for large n, toward a Gaussian the mean value and variance of which are found by summing the individual mean values and variances, respectively.*[2]

The conditions that have to be met by the individual distributions have all been listed in the foregoing derivation with one exception. We have still to account, in terms of the p_ν, for condition (C) of p. 267 that is, $| f_\nu(x) |$ less than one for $| x | > \delta$.

Now, $f_\nu(x)$ is the sum of the complex numbers

$$e^{-ixa_\nu}p_\nu(0), \quad e^{ix(1-a_\nu)}p_\nu(1), \quad e^{ix(2-a_\nu)}p_\nu(2), \ldots . \tag{47}$$

If this sum is represented as a vector sum in the complex plane (see Fig. 15), the vector corresponding to the label value k is $p_\nu(k)e^{ix(k-a_\nu)}$;

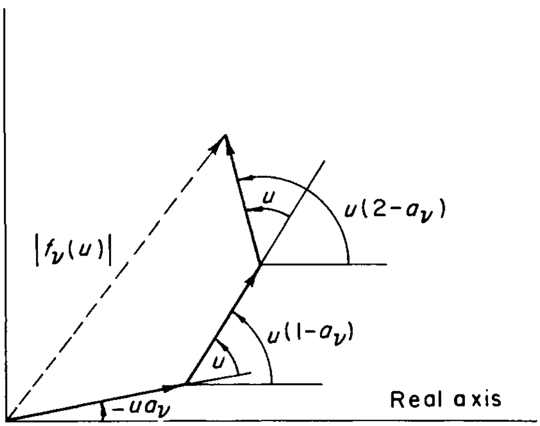

FIG. 15.

hence, it has the length $p_\nu(k)$ and forms the angle $x(k - a_\nu)$ with the real axis. Any two successive vectors form the angle $[(k + 1)x - a_\nu x] - [kx - a_\nu x] = x$. The length of the side that closes the vector polygon (broken line in Fig. 15) is $| f_\nu(x) |$. It is seen that, in general, $f_\nu(x)$ will be smaller than $p_\nu(0) + p_\nu(1) + \cdots = 1$. If $f_\nu(x)$ is to reach 1, all vectors (47) must have the same direction. This could happen if x were a multiple of 2π; but this is impossible since $| x | \leqslant \pi$. [For $| x | > \pi$ our $f_\nu(x)$ are

[2] R. v. MISES, "Fundamentalsätze." *Math. Z.* **4** (1919), pp. 1–96, see p. 30, also v. Mises [21], pp. 200–212.

zero.] But, if, for example, every second $p_\nu(k)$ is zero [say, $p_\nu(\pm 1) = p_\nu(\pm 3) = p_\nu(\pm 5) = \cdots = 0$), then the sum could be *one* for $x = \pi$ or $x = -\pi$. This case and similar ones are not covered by our proof given in Section 1.2, for which condition (C) was essential. To make sure that $|f_\nu|$ cannot become 1, except for $x = 0$, it is sufficient to require, for example, that each distribution $p_\nu(x)$ contains at least one pair of successive integral label values, with non-vanishing probabilities. Since the f_ν have to satisfy the condition $|f_\nu| < 1$ uniformly, we stipulate that for any ν there is an x such that $p_\nu(x) > \epsilon$ and $p_\nu(x+1) > \epsilon$. This condition is a little stronger than necessary, but we do not strive here for the greatest generality.[3]

That *some* restriction must be imposed on the $p_\nu(x)$ in addition to those introduced earlier in this section is seen by considering the above mentioned case where all $p_\nu(x)$ vanish for odd values of x. The probability $q_n(x)$ of drawing an odd sum x is then obviously zero, and consequently, the asymptotic function in (46) should reflect this property of the q_n. But this is not the case, since, whether x is odd or even, formula (46) approximates $q_n(x)$ by the bell-shaped normal distribution curve. Thus (46) loses its validity for such $p_\nu(x)$. On the other hand, in this particular example, the additional condition is of a formal nature, since one need only introduce another variable, for example, $2x' = x$, to obtain an asymptotic formula that holds.

3.2. *Discussion. Integration. Complete statement.* The asymptotic Laplace solution (34) and (35) for the Bernoulli problem, that was derived by applying the Stirling formula, is a special case of formulas (46) and (46'). If all $p_\nu(x)$ are identified with the alternative $p(0) = p$, $p(1) = q$, the mean value and variance of which are q and pq, respectively, the quantities b_n and $s_n{}^2$ become nq and npq. Equation (46') becomes identical with (35) upon making these substitutions.

The remarks of p. 275 are true also of more general applications: one is less interested in an approximate determination of $q_n(x)$ for an individual value of x, than in finding the probability that x falls in a given interval (x_1, x_2). Since x_1 and x_2 are often large numbers,

[3] Compare: R. v. MISES, "Generalizzazione di una teorema sulla probabilità della somma di un numero illimitate di variabili casuali." *Ist. Ital. Attuar.* 5 (1934), pp. 1–15, where the above condition is replaced by a less restrictive one; other generalizations have been given by G. BAWLY and by T. CONSOLI, *Rév. Fac. Sci. Univ. Istanbul* (1937) and (1940).

We see that the assumption of the text could be abated by dropping our stipulation [of the existence of two non-vanishing $p_\nu(x)$, $p_\nu(x+1)$ for each ν] for any finite number of distributions; even if $f_\nu = 1$ for a finite number of these functions, the product of all of them would still tend to zero like the nth power of a proper fraction.

this would require computing and summing a large number of single terms each of the form of the right-hand side of (46′). In the limit for infinite n, however, the summation can again be replaced by the much simpler process of integration. Let $h = 1/s_n$ and take for $q_n(x)$ its asymptotic approximation; in terms of the variable $u = h(x - b_n)$, we have, with $u_i = h(x_i - b_n)$, $i = 1, 2$,

$$\sum_{x=x_1}^{x_2} q_n(x) \sim \sum_{u=u_1}^{u_2} h\phi(u). \tag{48}$$

Here the second summation refers to the increments h in u that correspond to increments of magnitude 1 in x. Since h tends to 0 as $n \to \infty$, it becomes apparent that (48) converges toward the definite integral

$$\int_{u_1}^{u_2} \phi(u)\, du = \Phi(u_2) - \Phi(u_1), \tag{48′}$$

with the values of u_1, u_2 given above [see comments p. 275 and Eq. (37)].

On the other hand, if we introduce the c.d.f. of the distribution $q_n(x)$, namely,

$$Q_n(x) = \sum_{-\infty}^{x} q_n(x)$$

the probability $\sum_{x=x_1}^{x=x_2} q_n(x)$ can be written as $Q_n(x_2) - Q_n(x_1)$. Thus, we have found, as an approximation for large n, that

$$Q_n(x) \sim \Phi\left(\frac{x - b_n}{s_n}\right), \tag{49′}$$

under the conditions that led to (46′). The more exact statement of this asymptotic law is

$$\lim_{n\to\infty} Q_n(b_n + us_n) = \Phi(u), \qquad \text{for} \quad |u| < U. \tag{49}$$

If, instead of Φ, the function $G(x) = \int_0^x \phi(t)\, dt$ (introduced in Chapter III) is used Eq. (49′) reads, with $u = (x - b_n)/s_n$,

$$\begin{aligned} Q_n(x) \sim \Phi(u) &= \tfrac{1}{2} + \int_0^u \phi(t)\, dt, \\ &= \tfrac{1}{2} + G(u) \end{aligned} \tag{49″}$$

and the probability that the sum x of n chance variables lies between x_1 and x_2 becomes

$$\Pr\{x_1 \leqslant x \leqslant x_2\} = Q_n(x_2) - Q_n(x_1) \sim G(u_2) - G(u_1). \tag{50}$$

If, in particular, the interval (x_1, x_2) has center b_n and width $2X$, one obtains

$$Q_n(b_n + X) - Q_n(b_n - X) = \frac{2}{\sqrt{2\pi}} \int_0^{X/s_n} e^{-t^2/2}\, dt = 2G\left(\frac{X}{s_n}\right). \tag{51}$$

We summarize our results as follows:

(a) *For $(x - b_n)/s_n$ bounded, the probability $q_n(x)$ of drawing the sum x from n collectives with arithmetical distributions is, for large n, given by the approximation*

$$q_n(x) \sim \frac{1}{\sqrt{2\pi} s_n} \exp\left(-\frac{(x - b_n)^2}{2s_n^2}\right).$$

(b) *For the cumulative probability $Q_n(x)$ of drawing the sum $\leqslant x$,*

$$Q_n(x) \sim \Phi\left(\frac{x - b_n}{s_n}\right).$$

These formulas have been proved under the conditions

(1) *the variances r_ν^2 of the individual distributions are bounded*

$$r^2 \leqslant r_\nu^2 \leqslant R^2,$$

(2) *the absolute moments of third order have an upper bound*

$$\sum_x | x - a_\nu |^3 p_\nu(x) < T,$$

(3) *in every individual distribution (or in all except a finite number) there is one pair of consecutive integers with probabilites larger than some* $\epsilon > 0$.

Statement (b) will be proved later under much more general conditions. Note that statement (a) which may be denoted as a *local* limit theorem does not follow from (b) or from other more general theorems of type (b). The de Moivre-Laplace theorem proved in Chapter V, p. 261 by means of the continuity theorem is a particular case of the present result.

4. Examples

4.1. Numerical examples. Let us conclude this group of results with a few examples for the approximation formulas (46′), (50), and (51). We start with a simple numerical example. Assume a given probability, $q = 0.01$ for an individual of a certain population to die during a certain period (1 year). Let the number of individuals in the population be $n = 140{,}000{,}000$. The probability that x people out of these n die in one year can then be considered as the $q_n(x)$ in the Bernoulli problem, x being the number of events in n trials, and q the probability of a single event. Since n is so large, the direct application of the Bernoulli formula becomes impossible. However, if we apply (46′) to find $q_n(x)$, we need only know the mean value b_n, which equals $nq = 1.4 \times 10^6$, and the variance $s_n^2 = npq = 1.39 \times 10^6$, $s_n = 1180$. With these figures, one obtains from (46′) the probability of x deaths in one year:

$$q_n(x) \sim 3.38 \times 10^{-4} \exp\left(-[(x - 1.4 \times 10^6)^2/2.77 \times 10^6]\right)$$

Let us ask for the probability that the number of deaths in a given year be within a given fraction η of its expected value 1.4×10^6. With $\eta = 0.1\%$, formula (51) gives immediately the answer $2\int_0^{X/s_n} \phi(t)\,dt$, where $X = 0.001 \times 1.4 \times 10^6 = 1.4 \times 10^3$ and $s_n = 1180$. This gives the probability 0.77. Taking $\pm 0.2\%$ instead of $\pm 0.1\%$, we obtain $X/s_n = 2.4$ and a probability as high as 0.98. One sees that the distribution $Q_n(x)$ is very strongly concentrated at its mean value. This is due to the high value of n in our example.

The probability that in n drawings from a Bernoulli distribution with the probabilities p and q, a sum between $b_n(1 - \eta)$ and $b_n(1 + \eta)$ is obtained, becomes, according to (51),

$$2\int_0^{\eta\sqrt{nq/p}} \phi(t)\,dt. \tag{51′}$$

Since for any small η whatever, the argument of the Gaussian goes to ∞ with increasing n, the probability approaches 1 that the number x falls in a neighborhood of its expected value, whose width is a given fraction 2η of that value. The statement is, of course, only another (more precise) expression of the Bernoulli theorem of Chapter IV. The behavior of (51) with increasing n may be seen as follows. With $q = 0.01$, $\eta = 0.1\%$, that is, considering an interval centered on the mean value and extending a total length of 20% of this mean value, (51′) gives, for $n = 10^2, 10^3, 10^4, 10^5$, the probabilities 0.08, 0.25, 0.68, 0.9985, respectively.

As another example, take the repeated throwing of two unbiased dice. Since mean value and variance of the uniform distribution of the labels 1, 2, ..., 6 are 3.5 and 35/12, respectively, the mean value and variance for the sum of spots of two dice can be found from the summation theorems, Chapter V, (13) and (17), as 7 and 35/6, respectively. Suppose the pair of dice is thrown $n = 100$ times and one asks for the probability that the sum of all spots is 700, that is, exactly equal to its expected value. This probability becomes, since the exponential now has the value 1,

$$q_{100}(700) \sim \frac{1}{\sqrt{2\pi}\sqrt{\frac{3500}{6}}} = 0.0165.$$

Small though this probability is, it is still the maximum of the $q_{100}(x)$ for the 1001 possible values of x. But the probability of obtaining a sum between 680 and 720 is quite sizable. According to (51) we have

$$Q_{100}(720) - Q_{100}(680) \sim 0.59.$$

4.2. *Galton board. Simplest random walk.* As a more general example, we consider *Galton's board* or Galton's quincunx. In this apparatus small balls of equal diameter run down a slightly inclined plane, passing through n horizontal rows of pins arranged in such a way that a network of equilateral triangles is formed. All balls are released at the same spot, $x = 0$, opposite the central pin of the first row (Fig. 16). The diameter of the balls is slightly smaller than the distance between two neighboring pins; the latter distance may be taken as two units. We assume that each ball when hitting a nail (of any row) will deviate by one unit to the right or to the left with equal probabilities $p = q = \frac{1}{2}$. Each ball, after passing through the n rows of pins, will come to rest in one of the compartments (the width of each being equal to two units) at the bottom of the board. The probability q_n that a ball arrives in the compartment the abscissa of whose midpoint is x, is equal to the probability that in n alternatives with probabilities $p = q = \frac{1}{2}$, the positive result appears $\frac{1}{2}(n + x)$ times, the negative one $\frac{1}{2}(n - x)$ times, where x is a positive or negative number equal to one of the following:

$$-n, \quad -n + 2, \quad -n + 4, \cdots, n - 4, \quad n - 2, \quad n.$$

This probability is given by

$$q_n(y) = \binom{n}{y} p^{n-y} q^y = \binom{n}{y}\left(\frac{1}{2}\right)^n, \quad y = \frac{n + x}{2}, \quad y = 0, 1, \ldots, n. \qquad (52)$$

If a large number N of balls are observed, then approximately $Nq_n((n + x)/2)$ balls will be found in the xth compartment. In fact, consider for a fixed x, the number $Z(x)$ of events if an alternative with

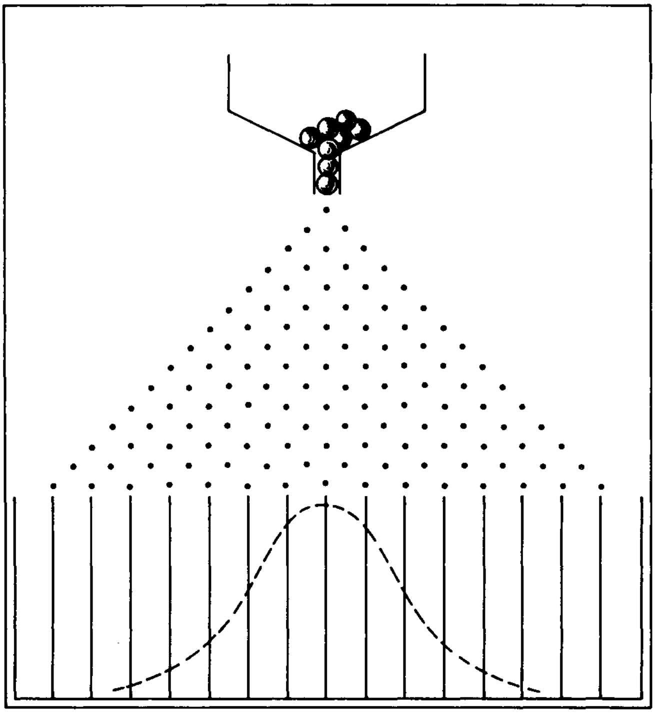

FIG. 16. Galton board.

probabilities $q_n(y)$ and $1 - q_n(y)$ is observed N times; the expected value of $Z(x)$ is

$$E[Z(x)] = Nq_n(y), \qquad y = \frac{n + x}{2}.$$

If the number n of rows is large, the right-hand side of $q_n(y)$ is approximated by the normal distribution with mean value and variance

$$b_n = n[\tfrac{1}{2} \cdot 1 + \tfrac{1}{2} \cdot (-1)] = 0, \qquad s_n^2 = n[1^2 \cdot \tfrac{1}{2} + (-1)^2 \cdot \tfrac{1}{2}] = n;$$

hence

$$E[Z(x)] = Nq_n(y) \sim \frac{N}{\sqrt{2n\pi}} e^{-x^2/2n}, \qquad x = -n, \quad -n + 2, \ldots, n - 2, \quad n. \tag{53}$$

The actual experiment for fairly large N and n yields indeed approximately the shape of a normal curve.[1]

It is of interest to mention another interpretation of the same problem; we mean its consideration in terms of the simplest *one-dimensional random walk*. Consider a long street extending in the positive and in the negative x-direction. A person starting at a point o throws a (true) penny and walks one unit length to the right if heads turns up, and a unit to the left if tails appears; then he repeats the game, and so on. We assume that his velocity is constant and that he spends a negligible time in throwing his penny and turning (if necessary). Our variables are the distance x and the time t spent in this walk; the unit of time is the time that elapses between two decisions (hence his velocity is equal to 1). Figure 17 shows a possible x-t line. Each path which joins the point O

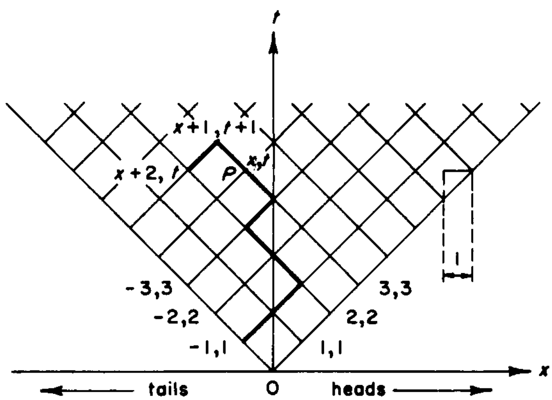

FIG. 17. Simplest random walk.

to a given point P has the same probability, namely, $(\frac{1}{2})^n$ if n is the number of throws between O and P (seven in the figure). We denote now by t, rather than by n, the number of throws which, with our choice of units, is equal to the time needed for the motion determined by these throws; x is the (positive or negative) number of units he moved to the right after t trials [having walked $\frac{1}{2}(t + x)$ times to the right and $\frac{1}{2}(t - x)$ times to the left.] We ask for *the probability $u(x, t)$ that in this one-dimensional random walk the person* (starting at the origin) *reaches the abscissa x at time t.*

The solution is given by (52) with $q_n(y)$ replaced by $u(x, t)$. We wish,

[1] The observed variance of this curve is rather consistently larger than the computed variance; see end of the section.

however, to consider (52) from a different angle. From $\binom{n}{y} + \binom{n}{y+1} = \binom{n+1}{y+1}$ it follows that

$$\tfrac{1}{2}[q_n(y) + q_n(y+1)] = q_{n+1}(y+1),$$

or , in terms of x and t, with $n = t$, $y = \frac{1}{2}(t + x)$:

$$u(x+1, t+1) = \tfrac{1}{2}u(x, t) + \tfrac{1}{2}u(x+2, t). \tag{54}$$

This formula has an immediate interpretation: in order to arrive at time $t + 1$ at the point with abscissa $x + 1$, the person may have stood at point x at time t and thrown heads or he may have stood at point $x + 2$ at time t and thrown tails. The recurrence equation (54) is a partial difference equation and we know the "initial values," that means the values of $u(x, t)$ along the two border lines, the 45° lines of Fig. 17. The line to the right contains the points $x = t = 0$, $x = t = 1$, $x = t = 2$, ..., that to the left: $-x = t = 1$, $-x = t = 2$, ... We obtain the table

$$\begin{array}{c|cccc} & (0,0) & (1,1) & (2,2) & (3,3)\ldots \\ & & (-1,1) & (-2,2) & (-3,3)\ldots \\ \hline u(x,t) = & 1 & \frac{1}{2} & \frac{1}{4} & \frac{1}{8},\ldots \end{array} \tag{54'}$$

Thus, by (54) and (54′) $u(x, t)$ is determined in the region of the figure.

We assume now that the chosen units of distance and time are very small in terms of some macroscopic measurements; or, in other words, x and t are large compared to these units. We write (54) in the form

$$u(x+1, t+1) - u(x+1, t) = \tfrac{1}{2}\,[u(x+2, t) - 2u(x+1, t) + u(x, t)]. \tag{54''}$$

Considering the increments small compared to x and t corresponds to the passage from the difference equation (54″) to a differential equation. On the left-hand side we have, at the position $x + 1$, the first difference with respect to t, and on the right-hand side we have at time t the second difference with respect to x. We obtain in the limit

$$\frac{\partial u}{\partial t} = \frac{1}{2}\frac{\partial^2 u}{\partial x^2}, \tag{55}$$

the "heat equation."[2] The reader may verify that this equation has the solution

$$u = \frac{1}{\sqrt{2\pi t}}\, e^{-x^2/2t}. \tag{55'}$$

[2] Note that the t in Eq. (55′) is continuous while the n in the original Galton board problem is discontinuous.

Equation (55) is also called the *diffusion equation*; $u(x, t)$ is then the probability density of x for a "particle" to be at a point with abscissa x at time t.

We return once more to the original Galton board. In experiments, a variance has been observed which is considerably larger than the variance computed in Eq. (53). One possible explanation is indicated in Problem 11. A more satisfactory theory is based on Markov chains.[3] The notation is now as in Chapter IV, Section 14.2.

It is plausible to assume and has actually been observed that the successive "decisions" of a ball are not independent of each other but that after a "decision" toward the right (left) it is rather more probable that the next one is again toward the right (left) than otherwise. Such a tendency would clearly lead to a wider curve (greater variance) than in the case of independent decisions.

In the application of a Markov chain to this problem, α is the probability of "left after left," β that of "left after right," $1 - \beta$ that of "right after right," etc. We assume symmetry, i.e., $\beta = 1 - \alpha$. From that it follows that the p and q in Eq. (98), Chapter IV, both equal $\frac{1}{2}$.

We take $\alpha > \frac{1}{2}$, which means that "left after left" and "right after right" are each more probable than either of the two other cases, in accordance with our intuitive starting point. Then $\delta = \alpha - \beta > 0$ and $(1 + \delta)/(1 - \delta) > 1$. Hence, according to Eq. (102) of Chapter IV, compared to the setup for independent trials, the variance is augmented *and strongly, if* $1 - \delta$ *is small.*

Problem 9. If an unbiased die is thrown $n = 100$ times, what is the probability of the total sum of points lying between $3.4n$ and $3.6\,n$? Compare the result with the estimate of the lower bound for the probability given by the Tchebycheff inequality. Do the same problem for $n = 1000$.

Problem 10. How many times must a pair of unbiased dice be thrown for the probability of the sum of points falling in the interval $6.9n$ to $7.1n$ to be about 0.99?

Problem 11. Suppose that in an experiment on the Galton board the observation of the number of balls landing in each compartment did not agree satisfactorily with the simple theory of Section 4.2. We assume then that in passing from one row to the next, the abscissa of a ball may change by some small multiple of ± 1 the unit being the half-distance between two nails and, in particular, we assume that the only changes occurring are $\Delta x = \pm 1, \pm 3$. Let $\frac{1}{2} - \eta$ be the probability of $\Delta x = \pm 1$

[3] G. Schulz, "Zur Theorie des Galtonschen Brettes." *Z. Physik* **92** (1934), pp. 747–754.

and η the probability of $\Delta x = \pm 3$. With $n = 100$ rows, it was found that 80% of the balls landed in the 21 center compartments at the bottom row with abscissas 0, ± 2, ± 4, ..., ± 20. Determine a value of η consistent with this result.

C. Probability Density. Central Limit Theorem. Lindeberg's and Liapounoff's Conditions (Sections 5-7)

5. The Summation Problem in the General Case

5.1. *Limit theorem in the case of densities.* The theorem derived in Section 3 includes two statements: (a) the probabilities of the individual values of the sum x of n chance variables (with arithmetical distributions) tend, as n increases, toward the values of a Gaussian function; (b) the sum of such probabilities for x lying between x_1 and x_2 tends toward a Gaussian integral $\Phi(u)$. The two analogous results can be reached, under certain conditions, when the given distributions have densities $p_\nu(t)$. We give a brief account of these results.

Let $p_1(t)$, $p_2(t)$, $p_3(t)$, ... be a sequence of probability densities. The mean values and variances will again be called a_ν and r_ν^2, respectively. The c.f. of the variables $t - a_\nu$ are

$$f_\nu(x) = \int e^{ix(t-a_\nu)}\, p_\nu(t)\, dt, \qquad \nu = 1, 2, 3, \ldots \tag{56}$$

It can immediately be seen that for every ν

$$f_\nu(0) = 1, \qquad f_\nu'(0) = 0, \qquad f_\nu''(0) = -r_\nu^2,$$
$$f_\nu'''(x) = -i \int (t - a_\nu)^3\, e^{ix(t-a_\nu)}\, p_\nu(t)\, dt. \tag{57}$$

Thus, conditions (A) of Section 1 are fulfilled and if the assumptions

$$r^2 \leqslant r_\nu^2 \leqslant R^2, \qquad T_\nu = \int |\, t - a_\nu\,|^3 p_\nu(t)\, dt < T \tag{58}$$

are made [see Eqs. (41) and (43)], it follows from the first part of Section 1 that

$$g_n\left(\frac{z}{s_n}\right) = f_1\left(\frac{z}{s_n}\right) f_2\left(\frac{z}{s_n}\right) \cdots f_n\left(\frac{z}{s_n}\right) \to e^{-z^2/2}\,. \tag{59}$$

Here, as p. 277 ff, g_n is the c.f. of the sum of the chance variables $t_\nu - a$. or, of the sum of the t's minus b_n. In addition to (59) we need the stronger theorem expressed in Eq. (17). This requires that conditions (C) and (C′) of Section 1 be fulfilled by the $f_\nu(x)$. It can be shown[1] that the present $f_\nu(x)$ fulfill the conditions if $p_\nu(x)$ is of uniformly bounded total variation[2] δ_ν in $(-\infty, +\infty)$, that is, if a number Δ exists such that $\delta_\nu < \Delta$ for all ν.

Consider condition (C): Since $e^{ix(t-a_\nu)}$ has absolute value one, $|f_\nu(x)| \leqslant \int p_\nu(t)\,dt = 1$, if, as we assume, $\int p_\nu\,dt$ exists. The bound 1 cannot be reached except for $x = 0$; in fact, it could be reached only if in (56) the values of $t - a_\nu$ differ by multiples of 2π, and that is not possible for a probability density $p_\nu(x)$. The validity of (C′) is proved by means of the second mean value theorem of integral calculus applied to both the real and imaginary parts of $f_\nu(x)$. The conditions (A), (B), (C), (C′) are thus satisfied, and formula (17) holds. It follows that we can use the inversion theorem of Chapter V in the form (62″) and we obtain

$$q_n(x) = \frac{1}{2\pi}\int e^{it(b_n-x)}\,g_n(t)\,dt = \frac{1}{2\pi s_n}\int \exp\left(-i\frac{z}{s_n}(x-b_n)\right) g_n\left(\frac{z}{s_n}\right) dz. \quad (60)$$

In this formula we may now replace $g_n(z/s_n)$ by its asymptotic value according to (59). Then with $(x - b_n)/s_n = u$, we obtain from (60) by the same computation as that carried out in Section 3

$$\lim_{n\to\infty} s_n q_n(b_n + s_n u) = \frac{1}{2\pi}\int e^{-(z^2/2)-iuz}\,dz = \frac{1}{\sqrt{2\pi}}\,e^{-u^2/2}. \quad (60')$$

Likewise $Q_n(x) = \int^x q_n(t)\,dt$ tends toward the d.f. $\Phi(u)$ of a normal distribution. The parameters of this normal distribution, its mean value and variance, are known to be the sums of the a_ν and r_ν^2, respectively. The resulting local limit theorem may be stated as follows:

If for a sequence of probability densities $p_\nu(x)$

(1) *the variances are bounded* $r^2 \leqslant r_\nu^2 \leqslant R^2$,

(2) *the absolute moments of third order have an upper bound* $\int |x - a_\nu|^3 p_\nu(x)\,dx < T$,

(3) *the total variations* δ_ν in $(-\infty, +\infty)$ of the $p_\nu(x)$ *are uniformly bounded, then* (a) *Eq.*) (60′) *holds,* or as an approximation statement: *the probability density* $q_n(x)$ *of the sum of the first* n *chance variables is*

[1] Cf. R. v. Mises, "Fundamentalsätze." *Math. Z.* 4 (1919), see pp. 31 and 32.

[2] See, e.g. W. Rudin, *Principles of Mathematical Analysis*, p. 99 ff, New York, 1953, or de la Vallée Poussin, *Cours d'Analyse Infinitésimale*, Vol. II, p. 1 ff, p. 45 ff, 1925.

approximated by a normal density with mean b_n and variance s_n^2; (b) *the c.d.f. of the sum tends toward the c.d.f. of a normal distribution when n increases indefinitely, the parameters b_n and s_n^2 of the resultant distribution being the sums of the a_ν and r_ν^2.*

As an example consider a uniform distribution

$$p(x) = \frac{1}{b-a}, \qquad a \leqslant x \leqslant b$$
$$= 0, \qquad x < a \quad \text{or} \quad x > b$$

The mean value is $(a+b)/2$, the variance $(b-a)^2/12$. Taking $a = -\frac{1}{2}$, $b = +\frac{1}{2}$ the variance is $\frac{1}{12}$ and the characteristic function (Chapter V, Problem 12) is $f(u) = (2/u) \sin (u/2)$. We consider the distribution of the sum of the independent variables

$$x = x_1 + x_2 + \cdots + x_n$$

all uniformly distributed in $(-\frac{1}{2}, +\frac{1}{2})$. The mean value of x is zero, the variance $n/12$. The reader will verify immediately that the conditions (1)–(3) of the above theorem are satisfied.

Now we compute directly: Let

$$\begin{aligned}
q_2(x) &= x + 1 && \text{for} \quad -1 \leqslant x \leqslant 0 \\
&= (x+1) - 2x = 1 - x && \text{for} \quad 0 \leqslant x \leqslant 1 \\
q_3(x) &= \tfrac{1}{2}(x + \tfrac{3}{2})^2 && \text{for} \quad -\tfrac{3}{2} \leqslant x \leqslant -\tfrac{1}{2} \\
&= \tfrac{1}{2}[(x + \tfrac{3}{2})^2 - 3(x + \tfrac{1}{2})^2] && \text{for} \quad -\tfrac{1}{2} \leqslant x \leqslant \tfrac{1}{2} \\
&= \tfrac{1}{2}[(x + \tfrac{3}{2})^2 - 3(x + \tfrac{1}{2})^2 + 3(x - \tfrac{1}{2})^2] && \text{for} \quad \tfrac{1}{2} \leqslant x \leqslant \tfrac{3}{2}.
\end{aligned}$$

Figure 18 shows $q_2(x)$, $q_3(x)$, $q_4(x)$; $q_n(x)$ consists of n parabolas of order $n-1$ between $-n/2$ and $+n/2$. Even $q_3(x)$ looks similar to a Gauss density.

As a counter example, remember the Gauchy distribution $F(x) = \frac{1}{2} + (1/\pi) \arctan x$ (Chapter V, Problem 7). If each x_i has this distribution, the same holds for the sum and the limit theorem does not hold true. We know that for this distribution neither mean value nor variance exist. The result which we shall find in Eq. (64) for the case of equal distributions cannot hold either.

5.2. *Preparatory study of the general case. Case of equal distributions.* Let us now turn to the *general case* where the original distributions are given by their d.f.'s $P_\nu(x)$, $\nu = 1, 2, 3, \ldots$, without our knowing that a derivative (density) exists or that the distributions are of the discontinuous type. The only question here can be how the d.f.'s $Q_n(x)$ behave for large values of n. We may assume that all individual mean values a_ν, and therefore b_n too, are zero. This involves no loss of generality since in the result, $x - b_n$ can be substituted for x if the a_ν are different from zero.

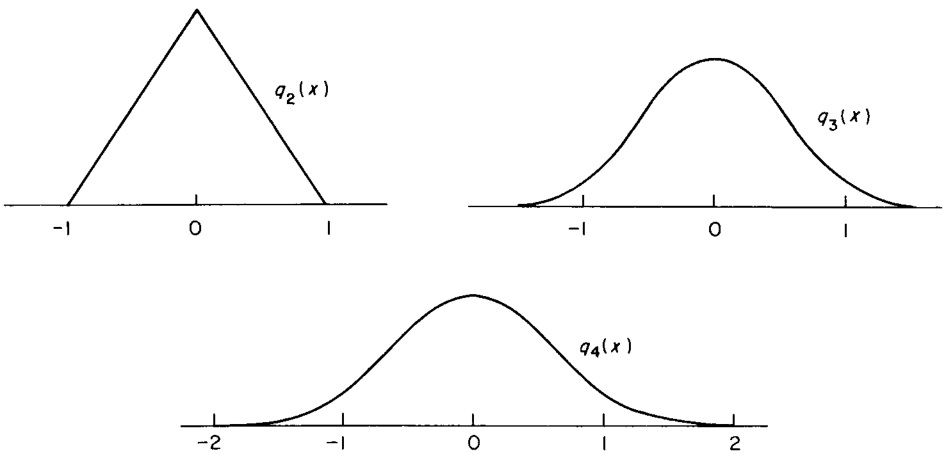

FIG. 18. Convolutions of uniform distribution.

The result (49), proved for the arithmetical case, or the result (60′) for densities can then be written in either of the two forms

$$\lim_{n\to\infty} Q_n\, s_n u) = \Phi(u), \qquad |u| < U, \qquad \text{or} \qquad Q_n(x) \sim \Phi\left(\frac{x}{s_n}\right), \tag{61}$$

where $\Phi(u)$ is the Gauss integral $\Phi(u) = (1/\sqrt{2\pi}) \int_{-\infty}^{u} e^{-x^2/2}\, dx$. We ask: under what conditions is (61) valid if no assumption about the type of the original distributions $P_\nu(x)$ is made?

In Chapter V, Section 8 we proved the far-reaching "continuity theorem" which we repeat here: *If a sequence of characteristic functions converges, in some interval of u, toward a continuous function $g(u)$, the corresponding distribution functions converge toward a d.f. whose c.f. is $g(u)$.* If we use this theorem, our problem reduces to finding conditions for $P_\nu(x)$ such that, with the usual notation

$$f_\nu(x) = \int e^{ixt}\, dP_\nu(t), \qquad g_n(x) = f_1(x) f_2(x) \cdots f_n(x), \tag{62}$$

the relation

$$\lim_{n\to\infty} g_n(x) = \lim_{n\to\infty} g\left(\frac{u}{s_n}\right) = e^{-u^2/2} \equiv g(u) \tag{63}$$

holds. In fact, we know that the c.f. of a normal distribution with mean zero and variance s^2 is $f(u) = e^{-\frac{1}{2}s^2u^2}$; hence $f(u/s) = e^{-\frac{1}{2}u^2}$. Since a c.f. determines the corresponding distribution uniquely we can conclude from (63) that the convolution $Q_n(x)$ whose c.f. is $g_n(u)$, converges toward $\phi(u)$.[3]

Now Eq. (63) has been shown to hold in our very first result (8), under conditions (A) and (B) of Section 1; the latter amount to requiring uniformly bounded variances r_ν^2 and uniformly bounded absolute moments of third order of the $P_\nu(x)$. The sufficient conditions can be replaced by much less restrictive ones of which the most remarkable are due to Liapounoff and to Lindeberg. Lindeberg's condition has even been proved to be in a certain sense both necessary and sufficient for (63). The next section will deal with this problem. However, before returning to these more difficult considerations, we shall prove a theorem which follows almost immediately, if we apply the continuity theorem to *the case of identical distributions* $P_1(x) = P_2(x) = \cdots = P(x)$.

Let the mean value of $P(x)$ be zero and denote by r^2 and by $f(t)$ its variance and its c.f. The distribution of the sum $x_1 + x_2 + \cdots + x_n$ has mean value zero, variance $s_n^2 = nr^2$, and c.f. $g_n(t) = [f(t)]^n$, and we have, as in (8′)

$$\lim_{n\to\infty} g_n\left(\frac{u}{r\sqrt{n}}\right) = \lim_{n\to\infty} f\left(\frac{u}{r\sqrt{n}}\right)^n = e^{-u^2/2}, \tag{62′}$$

which is the characteristic function of $\Phi(x)$. We conclude that the d.f. $Q_n(x)$ which corresponds to g_n tends to $\Phi(u)$ for every u as $n \to \infty$ (see Cramér [5], p. 52). Thus all that was needed here was that the variance r^2 of $P(x)$ be finite. We formulate this result for non-vanishing mean value.

Let $P(x)$ with mean value a and finite variance r^2 be the common d.f. of n independent chance variables $x_1, x_2, \ldots, x_n$; then, for the d.f. $Q_n(x)$ of the sum of these variables

$$\lim_{n\to\infty} Q_n(na + ur\sqrt{n}) = \Phi(u), \tag{64}$$

holds, or, in other words, *the distribution* $Q_n(x)$ of $x_1 + x_2 + \cdots + x_n = x$ *is approximately normal with mean value na and variance* nr^2; the *arithmetical mean* $(x_1 + x_2 + \cdots + x_n)/n$ *is approximately normal with mean value a and variance* r^2/n.[4]

---

[3] Note that the continuity theorem would not suffice to cover our two earlier statements on discrete distributions and densities since it relates only to convergence of the c.d.f.'s.

[4] This theorem generalizes in a simple way to k dimensions, namely, under the only

Problem 12. The probability density for placing a mark at a point x is ce^{-cx} for $x > 0$, $c > 0$. Compute the probability that the abscissa of the centroid of 1000 marks is not greater than 1.05 times its expected value.

Problem 13. If the probability density for placing a mark between the points $x = 0$ and $x = 1$ is constant, how many marks must be placed at random if one wants to have 95% probability that the centroid of the marks falls in the interval 0.4 to 0.6? What is the probability for the centroid lying between $x = 0.3$ and $x = 0.4$?

6. The Central Limit Theorem. Necessary and Sufficient Conditions

6.1. *Sufficiency of the Lindeberg condition.* Our problem is now to find conditions for the $P_\nu(x)$ under which the limit formula

$$\lim_{n\to\infty} f_1\left(\frac{u}{s_n}\right) f_2\left(\frac{u}{s_n}\right) \cdots f_n\left(\frac{u}{s_n}\right) = e^{-u^2/2} \tag{63}$$

holds, and, if possible, to find conditions which are also necessary for a sequence of distributions whose characteristic functions $f_\nu(u)$ satisfy (63).

We assume that the distributions of the individual chance variables are given by their d.f.'s

$$P_1(x), \quad P_2(x) \ldots, P_\nu(x), \ldots,$$

which have finite mean values and variances. We wish to show that the following conditions due to Lindeberg are sufficient for the asymptotic normality of $Q_n(x)$. For every $\epsilon > 0$, the distribution functions $P_\nu(x)$ satisfy the condition that, with $s_n^2 = r_1^2 + r_2^2 + \cdots + r_n^2$:

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{\nu=1}^{n} \int_{|x-a_\nu|>\epsilon s_n} (x - a_\nu)^2 dP_\nu(x) = 0. \tag{L}$$

It is seen that this condition requires that the sum of the outer contributions $|x - a_\nu| > \epsilon s_n$ to the variance integral becomes small compared to the variance s_n^2. This is certainly fulfilled for bounded distributions, that is, when all $dP_\nu(x) = 0$ for $|x| >$ some X.

assumption that for $P(x_1, x_2, \ldots, x_k)$ the vector of the mean value is finite and the matrix of the variances positive definite (see the last section of Chapter VIII for formal expression of a k-dimensional Gaussian).

Denote now by x_ν the r.v. with c.d.f. $P_\nu(x)$. To understand the meaning of (L) better, we consider[1] the probability

$$Pr\{\max_{1\leqslant\nu\leqslant n} |x_\nu - a_\nu| > \epsilon s_n\},$$

that the greatest of the n deviations $|x_\nu - a_\nu|/s_n$ be greater than ϵ. Denote by E_ν the event $|x_\nu - a_\nu| > \epsilon s_n$, $\nu = 1, 2, ..., n$; since

$$Pr\{\max_{1\leqslant\nu\leqslant n} |x_\nu - a_\nu| > \epsilon s_n\} \leqslant P(E_1 + E_2 + \cdots + E_n)$$

and

$$P(E_1 + E_2 + \cdots + E_n) \leqslant \sum_{\nu=1}^{n} P(E_\nu),$$

we find, with

$$P(E_\nu) = \int_{|x-a_\nu|>\epsilon s_n} dP_\nu(x) \leqslant \frac{1}{\epsilon^2 s_n^2} \int_{|x-a_\nu|>\epsilon s_n} (x - a_\nu)^2\, dP_\nu(x),$$

the inequality

$$Pr\{\max_{1\leqslant\nu\leqslant n} |x_\nu - a_\nu| > \epsilon s_n\} \leqslant \frac{1}{\epsilon^2 s_n^2} \sum_{\nu=1}^{n} \int_{|x-a_\nu|>\epsilon s_n} (x - a_\nu)^2\, dP_\nu(x)$$

But this last expression goes to zero for any ϵ, as $n \to \infty$, if (L) holds. Thus, (L) is seen to ensure that the probability that any *single* deviation $|x_\nu - a_\nu|/s_n$ be greater than ϵ, goes to zero: they must all be uniformly small.

Without loss of generality we assume again that the mean values $a_\nu = \int x\, dP_\nu(x)$ vanish. Then

$$r_\nu^2 = \int x^2\, dP_\nu(x); \qquad s_n^2 = r_1^2 + r_2^2 + \cdots + r_n^2.$$

Since the characteristic functions are defined by $f_\nu(u) = \int e^{ixu}\, dP_\nu(x)$, we have

$$f_\nu\left(\frac{u}{s_n}\right) = \int e^{ixu/s_n}\, dP_\nu(x). \tag{65}$$

For any integer $n \geqslant 1$ and any real a we have

$$e^{ia} = \sum_{\nu=0}^{n-1} \frac{(ia)^\nu}{\nu!} + \theta \frac{a^n}{n!}$$

where θ denotes any (real or complex) quantity of absolute value not greater than one. We have then

$$e^{ixu/s_n} = 1 + \frac{ixu}{s_n} + \frac{1}{2}\frac{x^2u^2}{s_n^2}\theta_1, \qquad |\theta_1| \leqslant 1. \tag{66}$$

and

$$e^{ixu/s_n} = 1 + \frac{ixu}{s_n} - \frac{1}{2}\frac{x^2u^2}{s_n^2} + \frac{1}{6}\frac{x^3u^3}{s_n^3}\theta_2, \qquad |\theta_2| \leqslant 1. \tag{66'}$$

[1] Compare Gnedenko [10], p. 227.

The range of integration in (64) will now be divided into the inner and outer domains $|x| \leqslant \epsilon s_n$ and $|x| > \epsilon s_n$, where ϵ is a positive quantity to be fixed later. With (66) and (66′) introduced into the integrals over the outer and inner domains, respectively, $f_\nu(u/s_n)$ appears as the sum of the following seven terms:

$$f_\nu\left(\frac{u}{s_n}\right) = \int_{|x|\leqslant \epsilon s_n} \left(1 + \frac{ixu}{s_n} - \frac{1}{2}\frac{x^2u^2}{s_n^{\,2}} + \frac{1}{6}\frac{x^3u^3}{s_n^{\,3}}\,\theta_2\right) dP_\nu(x)$$
$$+ \int_{|x|>\epsilon s_n} \left(1 + \frac{ixu}{s_n} + \frac{1}{2}\frac{x^2u^2}{s_n^{\,2}}\,\theta_1\right) dP_\nu(x). \tag{67}$$

Here, the sum of the second and sixth terms vanishes because $a_\nu = 0$, and the sum of the first and fifth terms is 1. Calling the sum of the three remaining terms η_ν, we have

$$f_\nu\left(\frac{u}{s_n}\right) = 1 + \eta_\nu\,, \tag{67′}$$

with

$$\eta_\nu = \int_{|x|\leqslant \epsilon s_n} \left(-\frac{1}{2}\frac{x^2u^2}{s_n^{\,2}} + \frac{1}{6}\frac{x^3u^3}{s_n^{\,3}}\,\theta_2\right) dP_\nu(x) + \int_{|x|>\epsilon s_n} \frac{1}{2}\frac{x^2u^2}{s_n^{\,2}}\,\theta_1\, dP_\nu(x). \tag{68}$$

If in the last term θ_1 is replaced by $1 + (\theta_1 - 1)$, η_ν can be written

$$\eta_\nu = -\frac{1}{2}\frac{u^2}{s_n^{\,2}}\int x^2\, dP_\nu(x) + \frac{1}{6}\frac{u^3}{s_n^{\,3}}\int_{|x|\leqslant \epsilon s_n} \theta_2 x^3\, dP_\nu(x)$$
$$- \frac{u^2}{2s_n^{\,2}}\int_{|x|>\epsilon s_n} (\theta_1 - 1)x^2\, dP_\nu(x)$$
$$= -\frac{u^2}{2s_n^{\,2}}\, r_\nu^{\,2} + R_\nu\,. \tag{69}$$

The expression R_ν consists of two terms for which upper bounds can be found. Since $|\theta_2| \leqslant 1$ and $|\theta_1 - 1| \leqslant 2$, we have for $|u| < U$

$$|R_\nu| \leqslant \frac{1}{6}\frac{U^3}{s_n^{\,3}}\int_{|x|\leqslant \epsilon s_n} |x|\, x^2\, dP_\nu(x) + \frac{2U^2}{2s_n^{\,2}}\int_{|x|>\epsilon s_n} x^2\, dP_\nu(x). \tag{70}$$

Replacing in the first integral $|x|$ by its maximum ϵs_n we obtain

$$|R_\nu| < \frac{1}{6}\frac{U^3}{s_n^{\,2}}\,\epsilon\int_{|x|\leqslant \epsilon s_n} x^2\, dP_\nu(x) + \frac{U^2}{s_n^{\,2}}\int_{|x|>\epsilon s_n} x^2\, dP_\nu(x). \tag{71}$$

Since each of these two integrals is smaller than r_ν^2 it is seen that an upper bound of $|R_\nu|$ can be written in the form r_ν^2/s_n^2 times a certain factor that is independent of ν and n. From (69) it follows that an upper bound of the same form holds for the quantities η_ν:

$$|\eta_\nu| < K\frac{r_\nu^2}{s_n^2}, \qquad K \text{ independent of } \nu, n. \tag{72}$$

Now let $\lim_{n\to\infty} s_n \to \infty$. The quotient r_ν^2/s_n^2 must then go to zero with $n \to \infty$ for fixed ν. In addition, we assume that the convergence of the quotients r_ν^2/s_n^2 is *uniform* with regard to ν, that is, to each given ϵ', however small, an n' can be found such that

$$\frac{r_\nu^2}{s_n^2} < \epsilon' \qquad \text{for} \quad \nu = 1, 2, \ldots, n, \quad \text{if} \quad n > n'. \tag{73}$$

It then follows from (72) that η_ν, which depends on ν and n, tends to zero in the same way. We have then established that

$$\lim_{n\to\infty} \eta_\nu = 0, \tag{74}$$

uniformly for $\nu = 1, 2, \ldots, n$, if $|u| < U$.

For small η it is known that

$$\log(1+\eta) = \eta(1+\epsilon) \tag{75}$$

where ϵ is a function of η such that $\epsilon \to 0$ as $\eta \to 0$. From (75) it follows

$$\sum_{\nu=1}^{n} \log(1+\eta_\nu) = \sum_{\nu=1}^{n} \eta_\nu + \sum_{\nu=1}^{n} \epsilon_\nu\eta_\nu \tag{76}$$

or, in view of (72),

$$\left|\sum_{\nu=1}^{n} \log(1+\eta_\nu) - \sum_{\nu=1}^{n} \eta_\nu\right| \leqslant \sum_{\nu=1}^{n} |\epsilon_\nu|\,|\eta_\nu| \leqslant \bar{\epsilon}_n \sum_{\nu=1}^{n} |\eta_\nu|$$

$$\leqslant \bar{\epsilon}_n K \sum_{\nu=1}^{n} \frac{r_\nu^2}{s_n^2} = \bar{\epsilon}_n K, \tag{77}$$

where $\bar{\epsilon}_n$ is the largest among the quantities $|\epsilon_\nu|$, for $\nu = 1, 2, \ldots, n$.

Let n increase indefinitely in (77). Since then $\bar{\epsilon}_n \to 0$ because of the uniformity of the convergence (74), it follows that

$$\lim_{n\to\infty} \sum_{\nu=1}^{n} \log(1 + \eta_\nu) = \lim_{n\to\infty} \sum_{\nu=1}^{n} \eta_\nu , \tag{77'}$$

provided $\lim_{n\to\infty} \sum_{\nu=1}^{n} \eta_\nu$ exists.

To investigate $\sum_{\nu=1}^{n} \eta_\nu$, we sum (69) for $\nu = 1, 2, ..., n$, obtaining

$$\sum_{\nu=1}^{n} \eta_\nu = -\frac{u^2}{2} + \sum_{\nu=1}^{n} R_\nu . \tag{78}$$

and

$$\sum_{\nu=1}^{n} | \eta_\nu | < \frac{U^2}{2} + \sum_{\nu=1}^{n} | R_\nu | . \tag{78'}$$

We obtain an estimate for $\sum | R_\nu |$ by going back to (71). Replace $\int x^2 \, dP_\nu(x)$ which is $\leqslant r_\nu^2$, by r_ν^2, and sum from $\nu = 1$ to n: $| x | \leqslant \epsilon s_n$

$$\sum_{\nu=1}^{n} | R_\nu | \leqslant \epsilon \frac{U^3}{6} + U^2 \frac{1}{s_n^2} \sum_{\nu=1}^{n} \int_{|x| > \epsilon s_n} x^2 \, dP_\nu(x). \tag{79}$$

The first term is independent of n and can be made arbitrarily small for any given U by choosing ϵ sufficiently small. The second term is the crucial one. We see that the coefficient of U^2 is the left-hand side of condition (L). We thus see that $\sum | R_\nu | \to 0$, no matter how large U, if condition (L) holds, viz.,

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{\nu=1}^{n} \int_{|x| > \epsilon s_n} x^2 \, dP_\nu(x) = 0 \qquad \text{for any given } \epsilon. \tag{L}$$

From (L), (79), and (78) it follows

$$\lim_{n\to\infty} \sum_{\nu=1}^{n} \eta_\nu = -\frac{u^2}{2} , \tag{80}$$

and, if (67′), (77′), and (80) are considered, we obtain

$$\sum_{\nu=1}^{n} \log(1 + \eta_\nu) = \sum_{\nu=1}^{n} \log f_\nu\left(\frac{u}{s_n}\right) = \log \prod_{\nu=1}^{n} f_\nu\left(\frac{u}{s_n}\right) \to -\frac{u^2}{2} . \tag{81}$$

The result has been reached under the assumptions: (1) that $s_n \to \infty$ as $n \to \infty$; (2) that the convergence toward zero of r_ν^2/s_n^2 for $\nu = 1, 2, ..., n$

is uniform; (3) that the limit condition (L) holds. It can be seen, however, that (2) and, therefore, (1) are included in (L). In fact, (L) states that for $n > n_1$ the sum of n positive terms $[\int_{|x|>\epsilon s_n} x^2\, dP_\nu(x)]/s_n{}^2$ and, therefore, each of these terms is smaller than an arbitrary ϵ_1. On the other hand,

$$\begin{aligned}\frac{r_\nu{}^2}{s_n{}^2} &= \frac{1}{s_n{}^2}\int_{|x|\leqslant\epsilon s_n} x^2\, dP_\nu(x) + \frac{1}{s_n{}^2}\int_{|x|>\epsilon s_n} x^2\, dP_\nu(x) \\ &< \epsilon^2 + \frac{1}{s_n{}^2}\int_{x>\epsilon s_n} x^2\, dP_\nu(x) < \epsilon^2 + \epsilon_1, \quad \text{for} \quad \nu = 1, 2, \ldots, n;\end{aligned} \tag{82}$$

hence, (2) follows if (L) holds; hence (1) follows, since, if $s_n{}^2$ were bounded, (2) could hold only if all $r_\nu{}^2$ were zero.

Condition (L) was given by J. W. Lindeberg in 1922.[2] Referring to what was said in the last section, our result can be summarized in the statement:

The distribution of the sum of n independent chance variables with finite variances (not necessarily bounded, not all vanishing) *tends toward the Gaussian distribution if the Lindeberg condition is fulfilled.*

6.2. *Comments. Lindeberg condition as necessary.* Note that this condition is certainly fulfilled if all distributions are equal with finite variance (c.f. theorem, p. 293). In this case $s_n = r\sqrt{n}$ and

$$\sum_{\nu=1}^{n}\frac{1}{s_n{}^2}\int_{|x|>\epsilon s_n} x^2\, dP_\nu(x) = \frac{1}{nr^2}\, n\int_{|x|>\epsilon s_n} x^2\, dP(x) = \frac{1}{r^2}\int_{|x|>r\epsilon\sqrt{n}} x^2\, dP(x).$$

Since $\int x^2\, dP(x)$ converges, the last integral tends to zero no matter how small ϵ as $n \to \infty$. Hence, (L) holds.

Another situation where (L) clearly holds is the case of *bounded* random variables x_ν (that means that any possible value of any x_ν is less than a constant C) if $s_n{}^2$ diverges toward infinity.

We want now to show that, in a certain sense, the Lindeberg condition is also a *necessary condition*[3] for the convergence toward a normal distribution.

---

[2] J. W. LINDEBERG, "Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung." *Math. Z.* **15** (1922), p. 211; see also quotations Section 7.1. A condition about equivalent to Lindeberg's has been given by P. Lévy (cf. three notes in *R. Acad. Sci.* (*Paris*) (1922), and a paper in *Bull. Soc. Math.* **52** (1924), pp. 49-84; see also [19a], p. 234.

[3] This has been proved in 1935 independently by Feller and P. Lévy.

Assume that

$$\lim_{n\to\infty} s_n = \infty \qquad \text{and} \qquad \lim_{n\to\infty} \frac{r_\nu^2}{s_n^2} = 0 \qquad \text{uniformly in } \nu, \tag{83}$$

and that $Q_n(x)$, the distribution of the sum, converges toward the Gaussian distribution, $\Phi(u)$. It follows from Chapter V, p. 257, (1), that the sequence of the corresponding characteristic functions $g_1, g_2, \ldots$ converges toward the characteristic function of $\Phi(u)$; hence (63) holds, and since it follows from (83) and (72) that the η_ν converge uniformly toward zero, we see, using (67′), and (80) that (81) is equivalent to

$$\sum_{\nu=1}^{n} \eta_\nu = \sum_{\nu=1}^{n} \left(f_\nu\left(\frac{u}{s_n}\right) - 1\right) \to -\frac{u^2}{2}. \tag{84}$$

Let us write (84) in the form

$$\frac{u^2}{2} - \sum_{\nu=1}^{n} \left[1 - f_\nu\left(\frac{u}{s_n}\right)\right] \to 0. \tag{85}$$

This relation must hold for both real and imaginary parts separately. Since the real part of $f_\nu(u/s_n)$ is $\int \cos xu/s_n \, dP_\nu(x)$ and $\int dP_\nu = 1$, relation (85) yields for the real part

$$\frac{u^2}{2} - \sum_{\nu=1}^{n} \int \left(1 - \cos\frac{xu}{s_n}\right) dP_\nu(x) \to 0, \tag{86}$$

which we write in the form

$$\lim_{n\to\infty} \left\{\frac{u^2}{2} - \sum_{\nu=1}^{n} \int_{|x|\leqslant\epsilon s_n} \left(1 - \cos\frac{xu}{s_n}\right) dP_\nu(x)\right\}$$
$$= \lim_{n\to\infty} \sum_{\nu=1}^{n} \int_{|x|>\epsilon s_n} \left(1 - \cos\frac{xu}{s_n}\right) dP_\nu(x). \tag{86′}$$

On the right-hand side, the integrand $1 - \cos xu/s_n$ is replaced by its maximum value, 2. For the resulting expression we obtain

$$2 \int_{|x|>\epsilon s_n} dP_\nu(x) \leqslant 2\,\frac{r_\nu^2}{\epsilon^2 s_n^2}, \tag{87}$$

where Tchebycheff's inequality has been used; hence,

$$\sum_{\nu=1}^{n} 2 \int_{|x|>\epsilon s_n} dP_\nu(x) \leqslant \frac{2}{\epsilon^2 s_n^2} \sum_{\nu=1}^{n} r_\nu^2 \leqslant \frac{2}{\epsilon^2}. \tag{88}$$

Introducing the abbreviation

$$S = \sum_{\nu=1}^{n} \int_{|x| \leqslant \epsilon s_n} \left(1 - \cos \frac{xu}{s_n}\right) dP_\nu(x),$$

we have thus found that the limit of $|(u^2/2) - S|$, as $n \to \infty$, if it exists, is certainly not larger than $2/\epsilon^2$, or

$$\overline{\lim_{n\to\infty}} \left| \frac{u^2}{2} - S \right| \leqslant \frac{2}{\epsilon^2}. \tag{89}$$

Using the inequality

$$\cos t \geqslant 1 - \frac{t^2}{2} \qquad \text{for all values of } t,$$

we find for S

$$S \leqslant \frac{u^2}{2s_n^2} \sum_{\nu=1}^{n} \int_{|x| \leqslant \epsilon s_n} x^2\, dP_\nu(x). \tag{90}$$

Replacing the inner integral by r_ν^2 minus the outer integral, we obtain

$$S \leqslant \frac{u^2}{2} - \frac{u^2}{2s_n^2} \sum_{\nu=1}^{n} \int_{|x| > \epsilon s_n} x^2\, dP_\nu(x), \tag{91}$$

or

$$\frac{u^2}{2s_n^2} \sum_{\nu=1}^{n} \int_{|x| > \epsilon s_n} x^2\, dP_\nu(x) \leqslant \frac{u^2}{2} - S. \tag{91'}$$

The left-hand side of this inequality is positive; hence $(u^2/2) - S = |(u^2/2) - S|$ and, using (89), we have finally

$$0 \leqslant \overline{\lim_{n\to\infty}} \frac{u^2}{2s_n^2} \sum_{\nu=1}^{n} \int_{|x| > \epsilon s_n} x^2\, dP_\nu(x) \leqslant \frac{2}{\epsilon^2}. \tag{92}$$

Since u may take any value and the bound is independent of u, the limit of the coefficient of $u^2/2$ must be zero, which is exactly Lindeberg's condition. Thus, it is seen that *for sequences of distributions fulfilling the conditions* (83), *condition* (L) *is also necessary for the convergence of the distribution* $Q_n(x)$ *toward a Gaussian.*

We have seen (pp. 298-299) that the Lindeberg condition implies $s_n^2 \to \infty$, $r_\nu^2/s_n^2 \to 0$ uniformly. Therefore, if either of these conditions does not hold, (L) cannot hold either. Cramér has proved[4] that in the case where $\lim_{n\to\infty} s_n = s$ (s finite) *the necessary and sufficient condition for the convergence of* $Q_n(x)$ *toward* Φ *is that each chance variable* x_ν *is normally distributed.*

7. Liapounoff's Sufficient Condition

7.1. *Derivation from Lindeberg's condition.* In practical applications, it is inconvenient to use the Lindeberg condition (L) in order to decide whether or not, for a sequence of collectives, the asymptotic distribution of the sum is normal. It will be found much easier to check the following sufficient (not necessary) condition for the existence of a Gaussian limit, which was given by Liapounoff twenty years before Lindeberg's work.

Theorem of Liapounoff.[1] Let $M_{|k|}^{(\nu)} = E[|\, x_\nu\, |^k]$ be the absolute moment about the mean value of order k of the νth distribution $P_\nu(x)$ and $s_n{}^2$ the sum of the individual variances $r_\nu{}^2$, as above. *The distribution of the sum of independent chance variables* x_ν *tends toward the normal distribution if, as* $n \to \infty$,

$$\frac{\sum_{\nu=1}^{n} M_{|k|}^{(\nu)}}{s_n{}^k} \to 0 \tag{LL}$$

for some $k > 2$ (k not necessarily integral).

Here, it is only necessary to check whether, in addition to the variances, the absolute moments of some order higher than the second exist and do not increase too strongly.[2] Since, in general (i.e., if all variances $r_\nu{}^2$ are of the same order of magnitude) $s_n{}^2$ increases with the first power of n, it is sufficient in most actual cases to show that the moments $M_{|k|}^{(\nu)}$ for some $k > 2$ are bounded, $M_{|k|}^{(\nu)} < M$.

To prove Liapounoff's theorem, we have to show that the condition (L) is satisfied if (LL) is true. In fact, with

$$|\, x\, |^k = x^2 |\, x\, |^{k-2}, \qquad M_{|k|}^{(\nu)} = \int |\, x\, |^{k-2} x^2 \, dP_\nu(x),$$

[4] Cramér [5], p. 59 ff. Further investigations concerning cases in which (L) is not satisfied are in the same work, p. 62 ff.

[1] A. Liapounoff, "Sur une proposition de la théorie des probabilités." *Bull. Acad. Sci., St. Petersbourg* **13** (1900).

[2] We mention without proof the useful fact that if (LL) holds for any $k > 2$, it will hold for all smaller $k > 2$. See Uspensky [26], p. 289.

one sees that

$$M_{|k|}^{(\nu)} = \int_{|x| \leqslant \epsilon s_n} |x|^{k-2} x^2 \, dP_\nu(x) + \int_{|x| > \epsilon s_n} |x|^{k-2} x^2 \, dP_\nu(x) > (\epsilon s_n)^{k-2} \times \int_{|x| > \epsilon s_n} x^2 \, dP_\nu(x) \tag{93}$$

and, for the sum of the $M_{|k|}^{(\nu)}$, one has

$$\sum_{\nu=1}^{n} M_{|k|}^{(\nu)} \geqslant \epsilon^{k-2} s_n^{k-2} \sum_{\nu=1}^{n} \int_{|x| > \epsilon s_n} x^2 \, dP_\nu(x). \tag{93'}$$

Dividing by the positive number $\epsilon^{k-2} s_n{}^k$ we obtain

$$\frac{1}{\epsilon^{k-2} s_n{}^k} \sum_{\nu=1}^{n} M_{|k|}^{(\nu)} \geqslant \frac{\sum_{\nu=1}^{n} \int_{|x| > \epsilon s_n} x^2 \, dP_\nu(x)}{s_n{}^2}. \tag{94}$$

If (LL) holds, the left-hand side goes to zero with increasing n, no matter how small the ϵ is; Therefore, the right-hand side goes to zero. Thus condition (LL) implies condition (L).

In most cases, one will first examine the moments of third order ($k = 3$). Note that the condition $\sum_{\nu=1}^{n} M_{|3|}^{(\nu)}/s_n{}^3 \to 0$ appeared in the theorem on discrete probabilities in the arithmetical case (Section 3) as well as in that on probability densities (Section 5). Once more, however, it may be emphasized that the local theorems obtained in these two particular cases are not included in the central limit theorem. This theorem deals merely with the cumulative distribution function of the sum.[3]

---

[3] In addition to the basic papers quoted above we mention the following work relevant to our subject.

A. Markov, "Démonstration du second théorème limite du calcul des probabilités par la méthode des moments." *Bull. Acad. Sci., St. Petersbourg* (1913).

G. Pólya, "Über den zentralen Grenzwertsatz d. Wahrscheinlichkeitsrechnung." *Math. Z.* **8** (1920).

Lévy [19a].

A presentation of a direct proof of Liapounoff's theorem is given in Uspensky's book [26], Chapter XIV. An estimate of the error terms involved in the approximation is also quoted there.

The problem of the most general conditions of convergence (which is connected with the problem of necessary conditions) has been investigated by P. Lévy, "Propriétés asymptotiques des sommes de variables aléatoires...." *J. Math. Pures Appl.* **14** (1935), p. 347; W. Feller, "Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung." *Math. Z.* **40** (1935), p. 521; **42** (1937), p. 301.

The following important books are relevant to the subject of this chapter: P. Lévy

7.2. *Examples.* We give some examples.

(1) Consider the simple alternative with probabilities p_ν, q_ν, $p_\nu + q_\nu = 1$, $\nu = 1, 2, \ldots$. Here $a_\nu = q_\nu$, $r_\nu^2 = p_\nu q_\nu$, $M_{|3|}^{(\nu)} = (1 - q_\nu)^3 q_\nu + |0 - q_\nu|^3 p_\nu = p_\nu q_\nu (p_\nu^2 + q_\nu^2) < p_\nu q_\nu$. Thus, $\sum_{\nu=1}^n M_{|3|}^{(\nu)}/s_n^3 < (\sum_{\nu=1}^n p_\nu q_\nu)^{-1/2}$. The Liapounoff condition is satisfied if $\sum p_\nu q_\nu$ is divergent. If $\sum p_\nu q_\nu$ is convergent we can use Cramér's result (p. 302); $Q_n(x)$ does not converge toward $\Phi(x)$ since the x_ν are not normally distributed. In this example the distribution is bounded but the variance does not tend toward infinity (see statement p. 299).

(2) Consider the discrete distributions

$$p_\nu(-\nu^k) = \frac{1}{2\nu^{2k}}, \qquad p_\nu(0) = 1 - \frac{1}{\nu^{2k}}, \qquad p_\nu(\nu^k) = \frac{1}{2\nu^{2k}}.$$

Here, the distributions $p_\nu(x)$ are not bounded, as in the preceding example. We have $a_\nu = 0$, $r_\nu^2 = 2\nu^{2k}$, $1/2\nu^{2k} = 1$, $s_n^2 = n$, $r_\nu^2/s_n^2 \to 0$, $M_{|3|}^{(\nu)} = \nu^k$; from Euler's formula, for large n, we estimate $\sum_1^n \nu^k$ as $n^{k+1}/(k+1)$ if $k > -1$ and have approximately $M_{|3|}^{(\nu)}/s_n^3 \sim [1/(k+1)]\, n^{k-\frac{1}{2}}$. This tends to zero for $k < \frac{1}{2}$ and (LL) holds.[4] Now consider $k \geqslant \frac{1}{2}$. We shall show that (L) cannot hold. Let $k = \frac{1}{2}$ and $\epsilon < 1$; then as long as ν is such that $\sqrt{\nu} < \epsilon s_n = \epsilon\sqrt{n}$ the term corresponding to this ν will make no contribution to the sum in (L); from $\nu = \epsilon^2 n$ on, each distribution $P_\nu(x)$ contributes 1 to this sum; between $\nu = \epsilon^2 n$ and $\nu = n$, there are $n - \epsilon^2 n = n(1 - \epsilon^2)$ terms each contributing unity and $n(1-\epsilon^2)/n = 1 - \epsilon^2$ does not tend to zero, and similarly for $k > \frac{1}{2}$.[5] [The reader should think over the (L) condition for a value of $k < \frac{1}{2}$]. The central limit theorem holds for our distribution if $k < \frac{1}{2}$.

(3) Let $p_\nu(-\nu^k) = \frac{1}{2}$, $p_\nu(\nu^k) = \frac{1}{2}$, $\nu = 1, 2, 3, \ldots$, k, a constant. Here $a_\nu = 0$, $r_\nu^2 = \nu^{2k}$, $M_{|3|}^{(\nu)} = \nu^{3k}$; let $k > -\frac{1}{2}$; then, estimating $\sum M_{|3|}^{(\nu)}/s_n^3$ as before, we have approximately $n^{3k+1}/(n^{k+\frac{1}{2}})^3$ and this tends to 0 as $n \to \infty$. For $k = -\frac{1}{2}$, we have s_n^2 of the order of $\log n$, and $\sum M_{|3|}^{(\nu)} = \sum \nu^{-3/2}$ is convergent; hence (LL) holds true. For $k < -\frac{1}{2}$, s_n^2 remains finite and Cramér's result shows that the central limit theorem cannot hold.

[19b], 1937; and B. W. Gnedenko and A. N. Kolmogoroff, *Grenzverteilungen von Summen unabhaengiger Zufallsgroessen.* Akademie Verlag, Berlin, 1959.

[4] Here, the theorem of Section 3 holds also.

[5] For $k > \frac{1}{2}$ there is a probability greater than $\prod_{\nu=2}^{\infty}(1 - \nu^{-2k}) > 0$ that any sum $x_2 + \cdots + x_n$ assumes the value zero.

Problem 14. The independent chance variables $x_1, x_2, \ldots, x_n$ are subject to a uniform distribution in the interval from zero to c. Find, for large n, the probability that $(x_1 + x_2 + \cdots + x_n)$ lies in the limits $\frac{1}{2}nc - nX$ and $\frac{1}{2}nc + nX$, or $(x_1 + x_2 + \cdots + x_n)/n$ in the limits $c/2 - X$ and $c/2 + X$.

Problem 15. If the individual distributions for the independent chance variables $x_1, x_2, \ldots, x_n$ have the density ce^{-cx}, $c > 0$, $x \geqslant 0$, find the distribution of $x_1 + x_2 + \cdots + x_n$ for large n.

Problem 16. The chance variable x_ν is uniformly distributed in the interval zero to $c = A_\nu{}^k$, $\nu = 1, 2, \ldots, n$. Prove that the central limit theorem applies if $k > -\frac{1}{2}$ and give, in this case, the asymptotic distribution of the sum $x_1 + x_2 + \cdots + x_n$. [Note (see p. 304) that the sum $1^m + 2^m + 3^m + \cdots + n^m$ increases as n^{m+1} if $m > -1$.

Problem 17. Each of the chance variables $x_1, x_2, \ldots, x_n$ varies in the interval 0 to 1. There is a point probability at $x_\nu = 0$ equal to $c_\nu > 0$, while otherwise there is a constant probability density in the interval 0, 1. Prove that the central limit theorem applies if the c_ν have an upper bound smaller than 1 and give, in this case, the asymptotic distribution of $x_1 + x_2 + \cdots + x_n$.

D. Probability of the Sum of Rare Events. Compound Poisson Distribution (Sections 8-10)

8. *Asymptotic Distribution of the Sum of n Discrete Random Variables in the Case of Rare Events*

8.1. *The Problem.* In Chapter IV we derived the Poisson distribution

$$\psi(x; a) = \frac{a^x}{x!} e^{-a}, \qquad a > 0, \quad x = 0, 1, 2, \ldots, \tag{95}$$

from the binomial distribution by letting $n \to \infty$, while $nq = a$ is fixed: this is the case of "rare" events. The result is parallel to the Laplace result of Section 2.2.

The application of the distribution (95) is not limited to the statistics of unusual and very rare events like suicides of children, births of triplets, deaths due to the kick of a horse, etc.; its great role has been recognized in problems arising in the statistics of telephone traffic, in many problems of theoretical physics, like Brownian motion, radioactive disintegration, etc. In the following we wish to generalize the limit result that led to (95) and to find in this way a more general theorem and a new more general distribution—of a similar character—which plays a rule in applications as well as in the theory. In Appendix Four we shall reconsider the Poisson distribution and its generalization, from another point of view.

Consider n distributions $p_\nu(x)$, $\nu = 1, 2, \ldots, n$; $x = 0, 1, 2, \ldots, m$. In each of these distributions all probabilities except that for $x = 0$ are "small"; we introduce the notation

$$r_\nu = \sum_{\mu=1}^{m} p_\nu(\mu), \qquad p_\nu(0) = p_\nu\,, \qquad r_\nu + p_\nu = 1, \qquad \nu = 1, 2, \ldots, n; \tag{96}$$

r_ν denotes the probability of a non-zero result in the νth trial and we assume that r_ν, which depends on n, is of order $1/n$; more precisely, letting R_n be greater than all r_ν, $R_n > r_\nu$, $\nu = 1, 2, \ldots, n$, we assume that

$$R_n < \frac{C}{n}\,, \tag{97}$$

where C is a positive constant. The case of $m = 1$, i.e., n alternatives $p_\nu(0) = p_\nu\,, p_\nu(1) = q_\nu\,, p_\nu + q_\nu = 1$, and q_ν small in the above sense, has been settled by v. Mises.[1] We put

$$\begin{aligned} a_\mu &= \sum_{\nu=1}^{n} p_\nu(\mu), \qquad \mu = 0, 1, \ldots, m; \\ \sum_{\mu=1}^{m} a_\mu &= \sum_{\nu=1}^{n} r_\nu = a, \end{aligned} \tag{98}$$

and assume as in the classical case that, as $n \to \infty$, the a_μ and their sum, a, remain fixed. We also set for $\mu = 1, 2, \ldots, m$

$$\alpha_\mu = \frac{a_\mu}{a}\,, \qquad \sum_{\mu=1}^{m} \alpha_\mu = 1. \tag{98'}$$

We denote by $q_n(x)$ the probability of obtaining a sum x in n trials when the distribution $p_\nu(x)$ previals for the νth trial, $\nu = 1, 2, \ldots, n$. We wish to find $q_n(x)$ and, in particular, its limit as $n \to \infty$, for the above described

[1] R. v. Mises, Über die Wahrscheinlichkeit seltener Ereignisse." *Z. Angew. Math. Mech.* 1 (1921), pp. 121–124. The general problem has been investigated by H. Pollaczek Geiringer, "Über die Poissonche Verteilung und die Entwicklung willkurlicher Verteilungen." *Z. Angew. Math. Mech.* **8** (1928), pp. 292–309, and in H. Geiringer, "On a limit theorem leading to a compound Poisson distribution." *Math. Z.* **72** (1960), pp. 229–234.

situation of "rare" events. The solution is simple if we assume the n distributions equal (see Problem 19). We wish to consider, however, the general case.

8.2. *Limit theorems.* Introducing generating functions we have

$$\sum_{x=0}^{mn} q_n(x)t^x = \prod_{\nu=1}^{n} \left(\sum_{\mu=0}^{m} p_\nu(\mu)t^\mu\right). \tag{99}$$

Using notation (96) we obtain

$$\begin{aligned}\sum_{\mu=0}^{m} p_\nu(\mu)t^\mu &= p_\nu(0) + \sum_{\mu=1}^{m} p_\nu(\mu)t^\mu \\ &= 1 - r_\nu + \sum_{\mu=1}^{m} p_\mu(\mu)t^\mu.\end{aligned} \tag{100}$$

Put $-r_\nu + \Sigma_{\mu=1}^{m} p_\nu(\mu)\, t^\mu = \eta_\nu$. Since $\Sigma_{\mu=1}^{m} p_\nu(\mu)\, t^\mu = O(r_\mu)$, it is seen from (97) that $\lim_{n\to\infty} \eta_\nu = 0$, uniformly for $\nu = 1, 2, \ldots, n$.

As in Eq. (75) of the present chapter we use for small η_ν

$$\log(1 + \eta_\nu) = \eta_\nu(1 + \epsilon_\nu), \tag{75'}$$

where ϵ_ν is a function of η_ν which tends to zero with η_ν. Writing as a temporary abbreviation A_n for the left-hand side of (99) we have

$$\log A_n = \sum_{\nu=1}^{n} \log(1 + \eta_\nu) = \sum_{\nu=1}^{n} \eta_\nu + \sum_{\nu=1}^{n} \epsilon_\nu \eta_\nu$$

and

$$\lim_{n\to\infty} \sum_{\nu=1}^{n} \log(1 + \eta_\nu) = \lim_{n\to\infty} \sum_{\nu=1}^{n} \eta_\nu\,,$$

provided this sum exists.

Now, by (98)

$$\lim_{n\to\infty} \log A_n = \lim_{n\to\infty} \sum_{\nu=1}^{n} \eta_\nu = -a + \sum_{\mu=1}^{m} a_\mu t^\mu \tag{101}$$

and, with $\lim_{n\to\infty} q_n(x) = q(x)$:

$$\lim_{n\to\infty} A_n = \sum_{x=0}^{\infty} q(x)t^x = \exp\left(-a + \sum_{\mu=1}^{m} a_\mu t^\mu\right) = h(t). \tag{102}$$

Introducing with notation (98′)

$$g(t) = \sum_{\mu=1}^{m} \alpha_\mu t^\mu \tag{103}$$

we see that $h(t)$ can be written

$$h(t) = e^{-a[1-g(t)]} . \tag{102′}$$

Thus with definition (102′):

$$\sum_{x=0}^{\infty} q(x) t^x = h(t). \tag{104}$$

A probability whose generating function is of the form (102′) is called a *compound Poisson distribution.* Some study of such distributions will follow in Section 9.2. We state: *The* $\lim_{n\to\infty} q_n(x) = q(x)$ *is a compound Poisson distribution.*

We now write $h(t)$ of (102) in the form

$$\begin{aligned} h(t) &= e^{-a}\Big[1 + (a_1 t + a_2 t^2 + \cdots + a_m t^m) + \frac{1}{2!}(a_1 t + \cdots + a_m t^m)^2 + \cdots\Big] \\ &= e^{-a} \sum_{k=0}^{\infty} \frac{1}{k!} (a_1 t_1 + a_2 t_2 + \cdots + a_m t_m)^k. \end{aligned} \tag{105}$$

In (105), in the last equality, t^μ has been replaced by t_μ , $\mu = 1, 2, \ldots, m$ $t^0 = t_0 = 1$; this is permissible since the proof of the result (102) would have worked in the same way with t_μ instead of t^μ. Using the last form of the right-hand side of (105) we obtain, since $\sum_{\mu=1}^{m} a_\mu = a$,

$$\begin{aligned} h(t) &= e^{-a} \cdot e^{a_1 t_1 + a_2 t_2 + \cdots + a_m t_m} \\ &= e^{-a_1(1-t_1)} \cdot e^{-a_2(1-t_2)} \cdot \cdots \cdot e^{-a_m(1-t_m)} \\ &= h_1(t_1) h_2(t_2) \cdots \cdot h_m(t_m), \end{aligned} \tag{106}$$

where

$$h_\mu(t) = e^{-a_\mu(1-t)} . \tag{106′}$$

Here, $h_\mu(t)$ is the generating function of $\psi(x; a_\mu) = (a_\mu{}^x/x!)\, e^{-a_\mu}$ [see Chapter V, Eq. (48)], i.e., of the limit probability of obtaining x $(=n_\mu)$ times the result μ in the continued drawing from urns which each contain two types of lots only, designated by 0 and μ, respectively, the latter being the rare event. Since $h(t)$, the generating function of $q(x)$, equals $h_1(t)\, h_2(t) \cdots h_m(t)$ we conclude from the multiplication theorem of generating functions that $q(x)$ must be equal to the probability of obtaining the sum x if successive drawings are made from m groups of urns, where the μth group contains the numbers 0 and μ only (in compositions which may vary from urn to urn), and "μ" is a "rare event."

We have thus obtained the theorem.

Consider n urns with the distributions $p_\nu(\mu)$, $\nu = 1, 2, ..., n$ $\mu = 0, 1, ..., m$ and assume that, as n tends toward infinity, the upper bound of all $r_\nu = \sum_{\mu=1}^m p_\nu(\mu)$ decreases as $1/n$. Then, the probability of obtaining x as the sum of the results of n drawings tends to

$$q(x) = \lim_{n\to\infty} q_n(x) = \sum \psi(n_1\,;\, a_1) \cdots \psi(n_m\,;\, a_m), \tag{107}$$

where the sum is taken over all those combinations of the $n_1, ..., n_m$ for which $\sum_{\mu=1}^m \mu n_\mu = x$. The a_μ are defined by (98).

The generating function of $q(x)$ is a compound Poisson distribution,[2] *as defined in* (102′).

Next, denote by $q_n(n_1, n_2, ..., n_m)$ the probability of obtaining in n draws n_1 ones, n_2 twos, ..., n_m m's, where $n_1 + n_2 + \cdots + n_m \leqslant n$. The relation between $q_n(n_1, ..., n_m)$ and $q_n(x)$ is given by the mixing

$$q_n(x) = \sum q_n(n_1, ..., n_m), \tag{108}$$

where the sum is over all those combinations of $n_1, n_2, ..., n_m$ for which $\sum_{\mu=1}^m \mu n_\mu = x$. Obviously, this relation holds also between $q(x)$ and $q(n_1, ..., n_m) = \lim_{n\to\infty} q_n(n_1, ..., n_m)$; hence

$$q(x) = \sum q(n_1, ..., n_m), \tag{108′}$$

with the summation rule as in (107).
Hence, from (107)

$$\lim_{n\to\infty} q_n(n_1, n_2, ..., n_m) = \psi(n_1; a_1) \cdots \psi(n_m; a_m). \tag{109}$$

If we want a direct deriviation of the result (109), the definition (99) is to be replaced by

$$\sum_{n_1=0}^{n} \cdots \sum_{n_m=0}^{n} q_n(n_1, ..., n_m) t_1^{n_1} \cdots t_m^{n_m} = \sum_{\nu=1}^{n} \left[\sum_{\mu=0}^{m} p_\nu(\mu) t_\mu\right], \tag{110}$$

[2] See H. Pollaczek Geiringer (1928), *loc. cit.* p. 306 and the paper by H. Geiringer, *loc. cit.*, p. 306. The present simplified proof is due in part to John Pratt.

where, on the left-hand side $n_1 + \cdots + n_m \leqslant n$. From this definition, one may compute the coefficient of $t_1^{n_1} \cdots t_m^{n_m}$ appearing on the right-hand side of (110), and then let $n \to \infty$ [as done in the paper of Geiringer (1928), *loc. cit.* p. 306]. Then, $q(x)$ follows by (108′) and the result is (107).

The main result, contained in (107) and (109), is that we have proved (not assumed as is often done) that, under our very general conditions, rare events become *asymptotically independent*. That means that the situation described in Section 8.1 amounts for large n, to that of m groups of independent alternatives. It is not surprising (see more in Section 8.3) that very rare events do not influence one other. A quite different result holds in the Gaussian case, where the Gaussian limit of $q_n(n_1, n_2, \ldots, n_m)$ contains $m(m-1)/2$ correlation terms (see Chapter VIII).

We would indeed not conjecture statistical independence of "frequent events" like, for example, deaths due to cancer and deaths due to heart disorder, while in the case of two rare causes of death the probability is quite small for any single individual to be exposed to both; this probability is of the order of the product of two small numbers.

We mention still another interpretation of the result (109). Write

$$\lim_{n\to\infty} q_n(n_1, \ldots, n_m) = \frac{(n - {}_0n)!}{n_1! \cdots n_m!} \alpha_1^{n_1} \cdots \alpha_m^{n_m} \frac{a^{n-n_0}}{(n-n_0)!} e^{-a},$$

or, with $\lim_{n\to\infty} q_n(n_1, \ldots, n_m) = q(n_1, \ldots, n_m)$:

$$\frac{1}{\psi(n-n_0; a)} q(n_1, \ldots, n_m) = \frac{(n_1 + \cdots + n_m)!}{n_1! \cdots n_m!} \alpha_1^{n_1} \cdots \alpha_m^{n_m}. \tag{111}$$

The joint limit distribution $q(n_1, \ldots, n_m)$, given that $(n - n_0)$ rare events have happened, equals the multinomial distribution of $n_1, \ldots, n_m$ with $\alpha_\mu = a_\mu/a$, $\mu = 1, 2, \ldots, m$ as probabilities.

Problem 18. Derive the result (109) for the particular case of n equal distributions $p_1(\mu) = p_2(\mu) = \cdots = p(\mu)$, $\mu = 0, 1, \ldots, m$. Use the polynomial distribution $q_n(n_1 \cdots n_m) = \dfrac{n!}{n_0! \cdots n_m!} p(0)^{n_0} \cdots p(m)^{n_m}$.

8.3. *Discussion and applications.* In applications we use (107) and (109) as approximation formulas when n is large and the expected values $a_1, \ldots, a_m$ are comparatively small. An example is the statistics of multiple births. For a certain population one observes over N months the number x of sets of twins born per month; assume that the numbers $x = 0, 1, \ldots, k$ appear with frequencies $f_0, f_1, \ldots, f_k, f_0 + f_1 + \cdots + f_k = N$. From these one computes in the usual way $(1 \cdot f_1 + 2 \cdot f_2 + \cdots$

$+ k \cdot f_k)/N = a_1$, and finds $\psi(n_1; a_1)$ which compares for $n_1 = 0, 1, 2, \ldots$ with $f_0, f_1, f_2, \ldots$. In the same way $\psi(n_2; a_2)$ is computed for triplets, then $\psi(n_3; a_3)$ for quadruplets. From this material, $q(x)$ is computed by (107) for certain values of x. On the other hand, a statistic of *all multiple births* for the same population gives relative frequencies for the occurrence of x children due to multiple births; these figures are then compared with the $q(x)$ found above.

In this procedure we actually check the statistical independence of multiple births; this independence seems plausible anyway. One expects that the probabilities of "a" twins and of "b" triplets in the same months do not influence each other.

Similar concepts seem adequate in a statistic of accidents, say, automobile accidents, involving 1 or 2 or 3 $\cdots$ people (or 1 or 2, ... cars). The $\psi(n_1; a_1)$, $\psi(n_2; a_2)$, ... are again computed from the separate statistics of these types of accidents and $q(x)$ follows from Eq. (107) and is compared in the same way as just explained with observed relative frequencies.

Another application of our results is suggested by consideration of mean value and variance of $q(x)$. For the distribution $p_\nu(\mu)$ the mean value is $b_\nu = 1p_\nu(1) + 2p_\nu(2) + \cdots + mp_\nu(m)$ and $s_\nu^2 = 1p_\nu(1) + 4p_\nu(2) + \cdots + m^2p_\nu(m) - b_\nu^2$. Now by (97) the b_ν^2 are of order $1/n^2$ and $\sum_{\nu=1}^{n} b_\nu^2$ of order $1/n$. As $n \to \infty$, mean value and variance of $q_n(x)$ approach

$$a = a_1 + 2a_2 + \cdots + ma_m$$

$$s^2 = a_1 + 4a_2 + \cdots + m^2a_m,$$

with the a_μ given by (98). Hence, in contrast to the ordinary Poisson distribution for which mean value and variance are equal, we see that, for $m > 1$, i.e., for $q(x)$ a compound Poisson distribution, *the variance of $q(x)$ is greater than the mean value*. Thus, if for data which suggest the possibility of being a sum of rare events, the variance is found appreciably greater than the mean, one may try the hypothesis that the basic distributions are not alternatives but multivalued distributions. In population statistics, cases of "supernormal dispersion" are very frequent.

Consider the Polya-Eggenberger investigation of mortality due to

[3] As against an observed mean $a = 5.5$ monthly deaths, a variance of $s^2 = 83.6$ was observed in a community.

[4] An urn contains a white and b black balls. A ball is drawn at random, the result is noted; the ball is then replaced and c balls of the color drawn are added. A new drawing is made from the $a + b + c$ balls, and so on. F. EGGENBERGER and G. PÓLYA, "Über die Statistik verketteter Vorgange." *Z. Angew. Math. Mech.* **3** (1923), pp. 279–289.

smallpox.[3] It is well known that the scheme of urn[4] indicated by these authors, which is such that each "success" increases the probability of further success, leads to a satisfactory description of the observations. An alternative model could, however, be based on the following reasoning: there are smallpox occurrences of various types—a type by which one single person is afflicted, others which befall a couple, a larger family, all the inhabitants of a house, etc. Each of these occurrences is a rare event; there is a distribution $p(x)$, where the $p(\mu)$, $\mu = 1, 2, ..., m$ are small, $p(0)$ large. Of course, such a theory makes sense only if values of the $p(1)$, $p(2)$, ... may be estimated from the observations and if the computation leads to a satisfactory over-all picture.

We finally mention the "mixed" case[5] where of the $(m + 1)$ lots in the original "urns" there are j "rare" ones and $k + 1$ "frequent" ones, $k \geqslant 1$, and we ask for the limit value of $q_n(x_1, ..., x_j, z_1, ..., z_k)$, viz., the probability of obtaining in n drawings $x_1, ..., x_j, z_1, ..., z_k$ lots of the various kinds. It is assumed that the upper bound of $r_\nu = \sum_{\mu=1}^{j} p_\nu(\mu)$ decreases as $1/n$ while the other $p_\nu(\mu)$ remain fixed.

Denote by $\phi(z_1, z_2, ..., z_k)$ the k-dimensional Gauss distribution. The result is

$$q_n(x_1, ..., x_j, z_1, ..., z_k) \sim \psi(x_1; a_1) ..., \psi(x_j; a_j)\phi(z_1, ..., z_k).$$

There is no link between the $\phi(z_1, ..., z_k)$ and the joint distribution of the j rare events.

Problem 19. Assume that for $\nu = 1, 2, ...$ there are five lots, with "small" probabilities s_1, s_2, and "large" probabilities p, q, and $r' = 1 - p - q - s_1 - s_2$; set $ns_1 = a_1$, $ns_2 = a_2$, and $r = 1 - p - q$. Prove that the mixed probability $q_n(x_1, x_2, x, y)$ (where x, y correspond to p and q) is given by $q_n(x_1, x_2, x, y) \sim \psi(x_1; a_1)\, \psi(x_2; a_2)\, \phi(x, y)$, with

$$\phi(x, y) = \frac{1}{2\pi n \sqrt{pqr}} \cdot$$

$$\exp\left[\frac{-q(r+p)(x-np)^2 + 2pq(x-np)(y-nq) - p(r+q)(y-nq)^2}{2npqr}\right].$$

---

[5] F. Bernstein suggested the consideration of a "mixed" case, since it occasionally occurs in heredity theory.

As a statistical illustration of the mixed case one could study statistics of suicides where men, women, boys, and girls are considered separately with the first two as frequent, the last two as rare events.

9. *Limit Probability of the Sum of Rare Events as a Compound Poisson Distribution*

9.1. *Additions and modifications.* We remember (104) and the first equality (105) and note that $q(x)$, the limit probability of the sum x, can be written in the following way. Denote by $\{\alpha\}_x^{(k)}$ the coefficient of t^x in $(\alpha_1 t + \cdots + \alpha_m t^m)^k$, i.e., the k-fold convolution of m constants α_1, α_2, ..., α_m with sum of subscripts equal to x. If α_i is the probability of the result "i," then $\{\alpha\}_x^{(k)}$ is the probability of obtaining the sum x in k trials with probabilities α_1, α_2, ..., α_m:

$$q(x) = e^{-a} \sum_{k=1}^{x} \frac{a^k}{k!} \{\alpha\}_x^{(k)} . \tag{112}$$

Of course, any $q_n(x)$ (probability of a sum x in n trials) is, by definition, a convolution—but generally it is an n-fold convolution, where n may tend toward infinity. In (112), however, k goes merely from 1 to x, and x is finite, actually rather small in a problem of rare events.

From (112) we compute, for example, $q(6)$[1]:

$$q(6) = e^{-a}\Big[a\alpha_6 + \frac{a^2}{2!}(2\alpha_5\alpha_1 + 2\alpha_4\alpha_2 + \alpha_3{}^2) + \frac{a^3}{3!}(3\alpha_4\alpha_1{}^2 + 6\alpha_1\alpha_2\alpha_3 + \alpha_2{}^3)$$
$$+ \frac{a^4}{4!}(4\alpha_3\alpha_1{}^3 + 6\alpha_2{}^2\alpha_1{}^2) + \frac{a^5}{5!}\cdot 5\alpha_2\alpha_1{}^4 + \frac{a^6}{6!}\alpha_1{}^6\Big] .$$

The result (112) admits also the following interpretation. Consider the distribution of a sum $x_1 + x_2 + \cdots + x_n$, *where n is not given, but is a random variable*, independent of the x_i and having Poisson distribution with mean value "a":

$$\Pr\{n = k\} = \frac{a^k}{k!} e^{-a}.$$

Each of the x_i has the arithmetical distribution (98′). Then

$$\Pr\{x_1 + \cdots + x_n = x\} = \sum_{k=1}^{\infty} \Pr\{n = k\} \Pr\{x_1 + \cdots + x_k = x\}$$
$$= \sum_{k=1}^{\infty} \frac{a^k}{k!} e^{-a} \{\alpha\}_x^{(k)} .$$

[1] $x = 6$ is a fairly large value for a problem where the $p_\nu(0)$ are the only appreciable probabilities.

Since $\{\alpha\}_x^{(k)}$ is the probability of the sum x in k trials, it is seen that the above sum goes actually from 1 to x only; indeed there must be at least *one* trial and there cannot be more than x trials, since in (98′) there is no probability of obtaining zero. Thus we recover (112).

In the case of n identical distributions a direct derivation of (112) is as follows. Let $1 - (a/n)$ and a/n be the probabilities of "zero" and "non-zero," respectively, where "non-zero" corresponds to the m rare events. Then $(a/n)^k(1 - (a/n))^{n-k}$ is the probability that a certain group of k results (for example the first, second, ... and kth one) are different from zero, while $(n - k)$ results are zero, and $\{\alpha\}_x^{(k)}$ is the probability that the sum of these k results, which we know is different from zero, equals x. Now all groups of k variables out of n have the same $(a/n)^k(1 - (a/n))^{n-k}$ and the same $\{\alpha\}_x^{(k)}$; hence, in order to take care of all possible groups of k variables out of n, we have only to multiply by $\binom{n}{k}$. Finally, by the addition rule, the sum over k is to be taken so that

$$q_n(x) = \sum_{k=1}^{x} \left(\frac{a}{n}\right)^k \left(1 - \frac{a}{n}\right)^{n-k} \binom{n}{k} \{\alpha\}_x^{(k)} . \tag{112′}$$

Then, in the limit $n \to \infty$ we obtain (112). If the n distributions $p_\nu(x)$ are not equal these simple considerations do not hold.

It is of interest to rewrite (112) in terms of the cumulative distribution $Q(x)$. Denote by $\Phi(x)$ the d.f. of the probability distribution[2] $\{\alpha_\mu\}$, $\mu = 1, 2, \ldots, m$, introduced in (98′). We use the unit step function $\epsilon(x)$ [(Chapter II, Section 2)], where

$$\epsilon(x) = 0 \quad \text{for} \quad x \leqslant 0, \qquad \epsilon(x) = 1 \quad \text{for} \quad x > 0$$

and set

$$\Phi_0(x) = \epsilon(x) \tag{113}$$

$$\Phi(x) = \Phi_1(x) = \alpha_1\epsilon(x - 1) + \cdots + \alpha_m\epsilon(x - m) \tag{113′}$$

and

$$\Phi_k(x) = \int \Phi_{k-1}(x - t)\, d\Phi(t), \qquad k = 1, 2, \ldots, \tag{114}$$

$\Phi_k(x)$ is the d.f. of the sum of k independent random variables each having d.f. $\Phi_1(x) = \Phi(x)$. In the present case $\Phi_1(x)$ is given by (113′) and

$$\Phi_2(x) = \alpha_1{}^2\epsilon(x - 2) + 2\alpha_1\alpha_2\epsilon(x - 3) + \cdots + \alpha_m{}^2\epsilon(x - 2m). \tag{113″}$$

[2] The $\Phi(x)$ here and in the remainder of this chapter have nothing to do with a normal distribution.

Now consider $Q(x)$, where we set $[x] = i$. We have

$$Q(x) = q(0)\epsilon(x) + q(1)\epsilon(x-1) + \cdots + q(i)\epsilon(x-i).$$

Here $q(0) = e^{-a}$ and each $q(\mu)$ is of the form (112). We rearrange $Q(x)$ by collecting first all terms with only α-factor [they are $\alpha_1\epsilon(x-1) + \alpha_2\epsilon(x-2) + \cdots + \alpha_i\epsilon(x-i) = \Phi(x)$, and they come from $q(1)$, from $q(2)$, ..., from $q(i)$]; then those with two such factors whose sum is equal to (113″), and which have been collected from $q(2)$, $q(3)$, ..., $q(i)$, those with three factors, collected from $q(3)$, $q(4)$, ..., $q(i)$, ..., and the only term with i factors coming from $q(i)$. These groups add up to $\Phi_1(x)$, $\Phi_2(x)$, ... and we obtain the formula, corresponding to (112),

$$Q(x) = e^{-a}\sum_{k=0}^{[x]} \frac{a^k}{k!}\Phi_k(x); \tag{115}$$

it allows the same interpretations as (112).

We may write in a formal way

$$Q(x) = e^{-a}\sum_{k=0}^{\infty} \frac{a^k}{k!}\Phi_k(x), \tag{115$'$}$$

where for $k > [x]$ the series will break off.

The expansions (112) and (115) have been proved here for the general case of non-identical distributions $p_\nu(\mu)$, $\nu = 1, 2, ..., \mu = 0, 1, ..., m$.

The results of this subsection and of Section 8 may be reviewed as follows:

The probability $q_n(x)$ of obtaining the sum x in the situation of rare events, described in Section 8.1., *converges as $n \to \infty$ toward a compound Poisson distribution $q(x)$ whose generating function $h(t) = \Sigma_{x=0}^{\infty} q(x)\,t^x$, is given by the expressions* (102′) *and* (103) *with the α_μ defined in Eq.* (98′). *Explicit expressions for $q(x)$ are* (107), (112), *and* (115).

Problem 20. Prove that the negative binomial distribution (see Chapter IV, p. 182) with generating function (Chapter V, p. 249)

$$F(t) = \left(\frac{q}{1-pt}\right)^r, \qquad p + q = 1$$

is a compound Poisson distribution. [Show that $F(t)$ can be put in the form (102′) with $a = r\log(1/q)$, $g(t) = r/a\log(1-pt)^{-1}$, a generating function.]

9.2. Compound Poisson distribution. Independently of the previous considerations, we now consider the compound Poisson distribution as such, defined by its generating function $h(t)$ of (102′), where the α_μ, $\mu = 1, 2, \ldots, m$ define a probability distribution. Mean value and variance of the compound Poisson distribution follow immediately by differentiation of $h(t)$. We have

$$\bar{x} = h'(1) = ag'(1), \qquad s^2 = h''(1) + h'(1) - h'(1)^2 = a[g''(1) + g'(1)].$$

Hence, if for the distribution $\{\alpha_i\}$ we write $\Sigma\, i\alpha_i = \bar{\alpha}$, $\Sigma\, i^2\alpha_i = M_2^{(0)}$, we have (see p. 311)

$$\bar{x} = a\bar{\alpha}, \qquad s^2 = aM_2^{(0)}.$$

Write now in (102′) λ instead of "a," $f(t)$ for $g(t)$, and $h[f(t); \lambda]$ instead of $h(t)$. Then

$$h[f(t); \lambda_1] \cdot h[f(t); \lambda_2] = h[f(t); \lambda_1 + \lambda_2], \tag{116}$$

or, more generally,

$$h[f(t); \lambda_1] \cdot h[g(t); \lambda_2] = h\left[\frac{\lambda_1 f(t) + \lambda_2 g(t)}{\lambda_1 + \lambda_2}; \lambda_1 + \lambda_2\right]. \tag{117}$$

The sum of n random variables each having a compound Poisson distribution with the respective parameters $\lambda_1, \ldots, \lambda_n$ and with $f_1(t)$, $f_2(t), \ldots, f_n(t)$, has a compound Poisson distribution

$$h\left[\frac{\lambda_1 f_1 + \cdots + \lambda_n f_n}{\lambda_1 + \cdots + \lambda_n}; \lambda_1 + \cdots + \lambda_n\right].$$

Consider now for any n the nth root of a given $h(t) = e^{-\lambda + \lambda f(t)}$, viz.,

$$[h(t)]^{1/n} = e^{-(\lambda/n) + (\lambda/n) f(t)}. \tag{118}$$

The nth *root of $h(t)$ is of the same type as $h(t)$ with the same $f(t)$ and parameter λ/n.* We call a distribution *infinitely divisible* if its generating function $g(t)$ is, for each n, the nth power of some generating function $g_n(t)$

$$g(t) = [g_n(t)]^n.$$

(The $g_n(t)$ need not be of the same type as $g(t)$.) In other words: *A distribution is called infinitely divisible if for each positive integer n it can be represented as the n-fold convolution of some other distribution with itself.* Or, again, equivalently: *A distribution $P(x)$ is infinitely divisible if for each positive integer n it is the distribution of a sum of n independent random* *variables* $x_1 + x_2 + \cdots + x_n$ *each* x_i *having the same d.f.* $P_n(x)$ (dependent on n).

We see that:

A compound Poisson distribution is infinitely divisible.[3] We do not follow up the theory of infinitely divisible distributions although it is closely connected with the problems considered in this chapter.[4] Remember that the Gaussian distribution is "infinitely divisible." Denoting its characteristic function by $f(u) = \exp(iau - \frac{1}{2}s^2u^2)$ we see that $f_n(u)$ is the c.f. of a Gaussian with mean value and variance a/n and s^2/n.

Problem 21. Show that the Pascal distribution $p(k) = a^k/(1 + a)^{k+1}$, $k = 0, 1, \ldots, a > 0$, is infinitely divisible. Show that it is a particular negative binomial distribution.

Problem 22. Show that the Cauchy distribution $P(x) = \frac{1}{\pi}\left(\frac{\pi}{2} + \arctan \frac{x - b}{a}\right)$, $a > 0$ is infinitely divisible.

Problem 23. Let $\alpha > 0$, $\beta > 0$ and the density

$$p(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \qquad x > 0$$

$$= 0 \qquad\qquad x \leqslant 0$$

(For $\beta = \frac{1}{2}$ and 2α integral we obtain the χ^2-distribution (Chapter VIII, Section 6).) Prove that $f(u) = (1 - (iu/\beta))^{-\alpha}$ and that $p(x)$ is infinitely divisible.

Problem 24. The d.f. of a finite sum $x_1 + \cdots + x_n$ of independent infinitely divisible variables x_ν is infinitely divisible.

10. A Generalization of the Theorem of Section 8

We have, so far, assumed that he n distributions $p_\nu(\mu)$ of Section 8 were arithmetical distributions with a finite number of non-negative labels $0, 1, 2, \ldots, m$. (If the non-negative labels are not integers but rational they can be considered as multiples of a new unit.) Consider the labels $0, 1, \ldots, m$. If $m < x$, not all theoretically possible terms will appear in $q(x)$ but only those which can be formed by means of $\alpha_1, \alpha_2, \ldots, \alpha_m$. But if $m \geqslant x$, the expression for $q(x)$ is always the same, independent of

[3] Some applications and literature on the compound Poisson distribution are in W. FELLER, "On a general class of 'contagious' distributions." *Ann. Math. Statist.* **14** (1943), pp. 389–399.

[4] See the monograph of Gnedenko and Kolmogorov [11].

the value of m; formula (112′) for $q(6)$ is the same for $m = 6$ as for $m = 100$. If there are denumerably many labels 0, 1, 2, 3, ..., we have for any $q(x)$ the same formula as we would have if $m = x$; this holds for (107) as well as for (112) and (115). On the other hand, if negative labels are admitted, the situation changes. Take equal distributions $P_\nu(x) = P(x)$ with jumps at -1, 0, $+1$ only, where $p(0)$ is the "large" probability. The result (109) remains true; with $a_1 = np(1)$, $a_2 = np(-1)$, $a = a_1 + a_2$ we obtain $\lim_{n\to\infty} q_n(n_1, n_2) = \psi(n_1; a_1)\psi(n_2; a_2)$. But the summation rule (108′) by means of which $q(x)$ followed from (109) is now of a changed type, since $q(x)$ is now the sum of all the infinitely many products for which $(+1)n_1 + (-1)n_2 = n_1 - n_2 = x$.[1] Likewise, in a formula corresponding to (112) or to (115), the summation will now extend to infinity so that convolutions of any order occur. Thus, our explicit answers are no longer of practical value.

Khintchine [16], p. 21 ff, has considered *a general situation of rare events, in the case of one and the same basic distribution.* $P(x)$ is the distribution of each of n random variables $x_1, x_2, ..., x_n$; at $x = 0$ there is one large probability $P(+0) - P(-0) = p_0 = 1 - a/n$, and otherwise $P(x)$ is arbitrary. We can write

$$P(x) = p_0\epsilon(x) + (1 - p_0)\Phi(x). \tag{119}$$

Here, $\Phi(x) = [P(x) - p_0\epsilon(x)]/(1 - p_0)$ is a distribution. In fact, since $\epsilon(-\infty) = 0$, $\epsilon(+\infty) = 1$, $P(-\infty) = 0$, $P(+\infty) = 1$, we see that $\Phi(-\infty) = 0$, $\Phi(+\infty) = (1 - p_0)/(1 - p_0) = 1$; and it is easily seen that $\Phi(x)$ is non-decreasing. In the previously considered arithmetical case defined by $\alpha_1, \alpha_2, ..., \alpha_m$, $\Phi(x)$ reduced to the distribution introduced in (113′). The convolutions $\Phi_k(x)$ are defined by formula (114). The decisive relation is similar to (112′):

$$Q_n(x) = \sum_{k=0}^{n} \binom{n}{k}\left(\frac{a}{n}\right)^k\left(1 - \frac{a}{n}\right)^{n-k}\Phi_k(x). \tag{120}$$

Consider in fact a term $\binom{n}{k}(1 - p_0)^k p_0^{n-k}\Phi_k(x)$. Here $p_0^{n-k}(1 - p_0)^k$ is the probability that a certain group of k variables out of n are different from zero; $\Phi_k(x)$ is the probability that the sum of these k variables, which we know to be different from zero, is $\leqslant x$. Now all groups of k variables have the same $\Phi_k(x)$ — defined by (113), (113′), and (114), and the same $(1 - a/n)^{n-k}(a/n)^k$; hence, we have only to multiply this product by $\binom{n}{k}$, in order to take care of all possible groups of k out of n, and then by

[1] One can get the "sum 6" by obtaining 100,000 times the result -1 and 100,006 times the result $+1$.

$\Phi_k(x)$. Finally, the sum over k from 0 to n is to be taken, and with $1 - p_0 = a/n$ we obtain (120).

It is then easily shown (see Khintchine [16], p. 21 ff) that, *as* $n \to \infty$, $Q_n(x)$ *tends toward*

$$Q(x) = e^{-a} \sum_{k=0}^{\infty} \frac{a^k}{k!} \Phi_k(x). \tag{121}$$

In (121), in contrast to (115), the summation goes from 0 to ∞, rather than from 0 to $[x]$. The distribution will appear again in Appendix Four.

In the very special case where $P(x) = (1 - a/n)\,\epsilon(x) + (a/n)\,\epsilon(x - 1)$, as in the original Poisson situation, it is seen that $\Phi(x) = \epsilon(x - 1)$ and, from that $\Phi_k(x) = \epsilon(x - k)$, $k = 0, 1, \ldots$. If then by the rule (121) we form

$$Q(x) = e^{-a} \sum_{k=0}^{\infty} \frac{a^k}{k!} \epsilon(x - k),$$

we indeed recover Poisson's $\psi(x; a)$.

APPENDIX FOUR

REMARKS ON ADDITIVE TIME-DEPENDENT STOCHASTIC PROCESSES

(a) *Introduction*. In the last two chapters we studied independent events. We considered successions of trials, numbered $1, 2, \ldots, n, \ldots$, such that the prevailing conditions for the nth trial were given from the start, once and for all, and did not depend on the outcome of previous trials. In Chapter IV we discussed certain forms of dependence, among them Markov chains. There, the probability p_{ij} for a system to assume a state i at time n depended on the stated j of the system at time $n - 1$ (but not on the "path" by which this state had been reached). Instead of the numbers $0, 1, 2, \ldots, n, \ldots$ we could use instants $t_0, t_1, \ldots, t_n, \ldots$. *The instants* t_ν *were considered given.*

This assumption of a sequence of given instants for the trials or observations is often not adequate. There are important processes where the changes happen at times which are themselves subject to chance. Whether we consider statistics of incoming telephone calls at an exchange, the occurrences of an illness, the radioactive disintegration of an atom, the mutation of genes, etc., it seems natural to assume that *such an event can happen in a random manner at any instant of a certain time interval, which may also be infinite.* The term "stochastic" originally had no meaning other than "by chance" or "at random"; it is, however, very often used in relation to time-dependent processes where, as just explained, *the time t at which a change of the system occurs is itself subject to chance.*

Within the class of stochastic processes, important distinctions appear. First, the events can be independent or they can be in chains (or have other forms of dependence beyond the scope of this book). In the first case, in addition to the term "independent" the more specific term *additive* is often used. The change of the system at time t is then independent of the past and even independent of the present instant, t. For the second case the term *Markovian* process is in general use. In this case it is assumed that the change of the system depends on its present state, t.

The changes of the system may be discontinuous or continuous. Suppose that a system S changes from state A to state B by means of jumps. The system has been in a state A from time t_0 to t_1; at time t_1 it changes abruptly to a state A' which it occupies from t_1 to t_2, etc., and the instants $t_0, t_1, t_2, \ldots$ of the jumps are subject to chance. We call such a process *discontinuous.* (The word "discrete" was used in the case of *given* instants $t_0, t_1, t_2, \ldots$.) A stochastic process with at most denumerably many states will naturally appear discontinuous. If, on the other hand, there is a continuum of states, one can visualize a continuous random change between t and $t + dt$ from A to $A + dA$ and we speak then of a *continuous* process.

A most important class of stochastic processes is the *stationary* class, also designated *homogeneous in time.* The process is such that the origin of the time-axis plays no role; only *changes* of time matter. In other words, the position of a time interval on the time axis is of no interest; only the length of the time interval is considered. A stationary process may be additive or Markovian or of a more general type, where the influence of the past cannot be neglected. The stationary processes are of great importance. The diffusion of gases per unit mass is, in general, considered stationary.

In the following, we shall illustrate some of these concepts and derive some basic results, which complement the results of Chapter VI.

(b) *Elementary discontinuous stochastic process.* We consider repetitive events like incoming telephone calls, radioactive disintegration, etc. The time of occurrence of such an event is not known beforehand. Denote by $p_k(t)$ the probability that in a time interval of length t, k events of the type under consideration occur.[1] We designate by $\nu(t)$ ($\geqslant 0$) an integral-valued random variable such that $\nu(t_0 + t) - \nu(t_0)$ *is the number of telephone calls, say, during time t.* The process is considered *stationary* or homogeneous in time, i.e., t_0 plays no role.[2] We also assume that the conditions at time $t \leqslant t_0$ play no role; the probability for the number of calls in the time interval $(t_0\,, t_0 + t)$ does not depend on the incoming calls in other non-overlapping time intervals (a, b); the process is *additive*, and homogeneous.

In terms of this random variable, $p_k(t)$ is the probability that $\nu(t_0 + t) - \nu(t_0)$ equals k.

$$p_k(t) = \Pr\{\nu(t_0 + t) - \nu(t_0) = k\}. \tag{1}$$

Then $p_0(t) = f(t)$ is the probability of no event and $1 - p_0(t) = \phi(t)$ the probability of at least one event in the time t.

For a small time interval t, we now make the following assumption for $\phi(t)$. There is a constant $\lambda > 0$ such that $\phi(t)$ is of the form λt to within a magnitude of the order $o(t)$.

$$\phi(t) = \lambda t + o(t), \qquad \lambda > 0. \tag{2}$$

This implies that the probability for a call tends to zero as $t \to 0$; $\phi(0) = 1 - p_0(0) = 0$, $p_0(0) = 1$. Now consider $p_0(t) = 1 - \phi(t)$, the probability of no event during the time interval t. It follows from our assumptions that $p_0(t + h) = p_0(t)\, p_0(h)$, i.e., the probability of no event during time $t + h$ is the combined probability of no event during t and no event during h. Then

$$p_0(t + h) = p_0(t)p_0(h) = p_0(t)[1 - \phi(h)] = p_0(t)[1 - \lambda h + o(h)]. \tag{3}$$

Now divide by h and let $h \to 0$; then $o(h)/h \to 0$ and we obtain

$$\begin{aligned} p_0'(t) &= -\lambda p_0(t), \\ p_0(t) &= Ke^{-\lambda t}, \end{aligned} \tag{4}$$

and since $p_0(0) = 1$,

$$p_0(t) = e^{-\lambda t}, \qquad \phi(t) = 1 - e^{-\lambda t}. \tag{4'}$$

[1] See the remarkable presentation in Khintchine [16], p. 19 ff.

[2] Of course there are rush hours and quiet hours. We may assume that we are investigating rush-hour traffic; the process may then be considered stationary.

From (2) we see that λ is a physical constant which characterizes the frequency of the event; and from (4′) we see that the larger λ, the smaller the probability of no call during a short time t.

We make now the *additional assumption that the probability of more than one event during the small interval t is* $o(t)$ and denote this probability by $\psi(t)$

$$\psi(t) = 1 - p_0(t) - p_1(t) = \phi(t) - p_1(t) = o(t). \tag{5}$$

Then, using again the assumptions of additivity and stationarity, as well as (2) and (5), we obtain for $k > 0$, the (immediately understood) relation

$$\begin{aligned} p_k(t+h) &= p_k(t)p_0(h) + p_{k-1}(t)p_1(h) + o(h) \\ &= p_k(t)[1 - \phi(h)] + p_{k-1}(t)[\phi(h) - \psi(h)] + o(h) \\ &= p_k(t)[1 - \lambda h + o(h)] + p_{k-1}(t)[\lambda h + o(h)] + o(h). \end{aligned} \tag{6}$$

Thence

$$\frac{p_k(t+h) - p_k(t)}{h} = -\lambda p_k(t) + \lambda p_{k-1}(t) + \frac{o(h)}{h}$$

and, as $h \to 0$

$$\frac{1}{\lambda} p_k'(t) = p_{k-1}(t) - p_k(t), \qquad k > 0. \tag{7}$$

This is a difference-differential equation. To solve it we introduce a new function $u_k(t)$ by

$$p_k(t) = e^{-\lambda t} u_k(t), \qquad k > 0.$$

Then (7) becomes

$$u_k'(t) = \lambda u_{k-1}(t),$$

and $u_k(0) = p_k(0) = 0$ for $k > 0$, since $p_0(0) = 1$. Therefore

$$u_k(t) = \lambda \int_0^t u_{k-1}(s)\, ds, \qquad u_0(t) = e^{\lambda t} p_0(t) = 1, \tag{8}$$

where (4′) has been used. Thus, from (8),

$$u_k(t) = \frac{(\lambda t)^k}{k!}, \tag{8′}$$

and

$$p_k(t) = \frac{(\lambda t)^k e^{-\lambda t}}{k!} \qquad k = 0, 1, \dots . \tag{9}$$

We have proved under the assumptions (2), (3), and (5):

The probability distribution of k events (e.g., telephone calls) *during the time interval t* is a *Poisson distribution with mean value* λt, [λ defined by (2), i.e., $\phi(t)/t \to \lambda$]. *The Poisson distribution appears here not as a limit formula* (as in Chapters IV and VI) *but as the direct solution of a physical problem, and mathematically as the solution of a functional equation.*

(c) *A general discontinuous stochastic process.* We consider again a random variable which at instants distributed at random undergoes sudden changes. We again denote by $\phi(t)$ the probability of at least one sudden change during t; and we assume that the process is stationary and additive. Again, as in (b), (2) and (5) are to hold. In (b), however, we counted only the number of jumps, or, in other words it was tacitly assumed that each jump was of magnitude $+1$. Now we consider also the amplitude of each jump; the jump which happens at time t_i has an amplitude x_i which is itself a random variable of additive character and independent of t_i. We assume that all amplitudes have the same distribution function $\Phi(x) = \Phi_1(x) =$ *the probability of an amplitude* $\leqslant x$. If there are k jumps in the time interval t we denote by $x(t)$ the sum of the k amplitudes, each with distribution $\Phi_1(x)$. Then $\Phi_k(x)$ is the probability that the sum of the k amplitudes has a value $\leqslant x$; here $\Phi_k(x)$, as in Chapter VI, Eqs. (113′) and (114) is the k-fold convolution of $\Phi_1(x)$ by itself. *Let $F(x, t)$ be the distribution of $x(t)$*, i.e., *the probability that the sum of all amplitudes which correspond to jumps during time t be $\leqslant x$.*

In a less abstract way, consider again the telephone call problem and assume that the probability of k calls during time t is given by (9). Now, we associate the random variable x (the "amplitude") with the duration of each call, x being the length of the respective conversation. In another example, taken from insurance, one observes the cumulative damage $x(t)$ due to fire accidents during time t. $F(x, t)$ is the probability that the total damage through those accidents is $\leqslant x$ or, in the previous example, $F(x, t)$ is the total time the telephones in the exchange have been in use due to incoming calls during time t.

There can be $k = 0, 1, 2, \ldots$ events ("jumps") during t; the respective probabilities are given by (9) and $\Phi_k(x)$ is the probability that the sum of the corresponding k amplitudes is $\leqslant x$. Hence (in terms of the insurance example) $p_k(t)\, \Phi_k(x)$ is the probability that there are k accidents during t and that they cause a total damage $\leqslant x$. Therefore

$$F(x, t) = e^{-\lambda t} \sum_{k=0}^{\infty} \frac{(\lambda t)^k}{k!} \Phi_k(x). \tag{10}$$

Thus *we have obtained directly*, by answering a physical question, *the general Poisson distribution* (121) of Chapter VI, Section 10 with "a" replaced by λt. We note that if the characteristic function of $\Phi(x)$ is denoted by $f(u) = \int e^{iux}\, d\Phi(x)$, then the characteristic function of $F(x, t)$ is

$$G(u, t) = e^{-\lambda t} \sum_{k=0}^{\infty} \frac{(\lambda t)^k}{k!} f(u)^k = e^{\lambda t (f(u)-1)}. \tag{10'}$$

(d) *An additive continuous stochastic process.* We remember the main problem of Chapter VI which, in the language of random walk, is as follows: A point starts from the origin and moves along a straight line in successive steps of unequal length, one each second, say. The change of its abscissa at the kth step is x_k and the probability for the next step is independent of x_k and of everything that happened before time k. We are interested in the distribution of $\sum_{k=1}^{n} x_k$ after n seconds, and in particular as the time n tends toward infinity. Equivalently, one may assume that the single steps happen in very small time intervals so that the total time remains finite as the number of steps tends to infinity. It is then not unnatural to consider the process as a continuous one; that means, to consider a continuously varying magnitude x—for example the abscissa of a gas molecule—and to think of $x = x(t)$.[1] We do not plan to rewrite and reinterpret our limit theorems in terms of a continuous parameter[2] t instead of the discontinuous parameter n; now we wish to obtain the normal law directly, as the exact solution of a functional equation which corresponds to a certain continuous stochastic process.

Let us consider the change of a quantity x, for example the abscissa of a moving particle. We assume *independent changes*, which means here the following: *the difference $\Delta x = x(t + \Delta t) - x(t)$ $(\Delta t > 0)$ is a random variable which is independent of the values of x* before time t. Consequently, if we consider a number of non-overlapping time intervals $i_1, i_2, \ldots$, then the corresponding changes of x are mutually independent random variables[3]; $x(t)$ can then be considered as the sum of $x(0)$ plus the changes corresponding to any number of successive non-overlapping intervals which cover the time from 0 to t. We thus assume *additivity in space*, also denoted as *independent changes*; but we *do not assume here homogeneity in time. We denote by $P(z, t, \tau)$ the distribution function of the*

[1] This problem was first considered by P. Bachelier, *Calcul des probabilités.* Paris, 1912. A fundamental paper in this direction is A. Kolmogoroff, "Über die analytischen Methoden der Wahrscheinlichkeitsrechnung." *Math. Ann.* **104** (1931), pp. 415–458. We follow here Khintchine's previously quoted book [16].

[2] See v. Mises [21], p. 495 ff, where this has been done.

[3] See footnote 4, p. 325.

change z within the time interval from t to τ, that is, the probability that the change which the abscissa of the moving point undergoes in the time between t and τ is less than or equal to z. For the sake of simplicity we assume that a probability density exists. The value of the abscissa of the particle at time t plays no role. (The assumption that the process depends on the *positions* of the particle at times t and τ, and not only on the difference between these positions leads to a Markov process.) The probability density is

$$\frac{\partial P}{\partial z} = p(z, t, \tau), \tag{11}$$

so that $p(z, t, \tau)\,dz$ is the probability that the change of x during (t, τ) lies between z and $z + dz$. Then

$$\int p(z, t, \tau)\,dz = 1 \qquad \text{for all} \quad t, \tau. \tag{11$'$}$$

We make now the following further assumptions:

(i) The density is continuous with respect to t and τ. The mean value $a(t, \tau)$ vanishes for all t, τ:

$$a(t, \tau) = \int zp(z, t, \tau)\,dz = 0. \tag{12}$$

(ii) We denote the variance by $r(t, \tau)$:

$$\int z^2 p(z, t, \tau)\,dz = r(t, \tau); \tag{13}$$

from our assumption that the changes of x in non-overlapping time intervals are independent it follows that

$$r(t, t_1) + r(t_1, \tau) = r(t, \tau), \qquad t < t_1 < \tau. \tag{14}$$

We also assume that $r(t, \tau)$ has a derivative with respect to τ.[4]

(iii) We set $t = 0$, and write then t instead of τ:

$$r(0, t) = r(t), \qquad p(z, 0, t) = \phi(z, t) \tag{15}$$

[4] The assumption of independent changes in successive small time intervals is not too realistic. An analogous remark applies to the assumed differentiability of $r(t, \tau)$, as pointed out by Khintchine.

and assume that $\partial\phi/\partial t$, $\partial\phi/\partial z$, and $\partial^2\phi/\partial z^2$ exist for all $t > 0$ and all z and that $\partial^2\phi/\partial z^2$ is bounded and uniformly continuous for $t > 0$.

(iv) Finally, we introduce a condition analogous to the Lindeberg condition (L) of Chapter VI (p. 294), namely, that for any arbitrarily small $\delta > 0$

$$\lim_{h\to 0} \frac{1}{h} \int_{|\zeta|>\delta} \zeta^2 p(\zeta, t, t+h)\, d\zeta = 0. \tag{L}$$

A consequence is that

$$\lim_{h\to 0} \frac{1}{h} \int_{|\zeta|>\delta} p(\zeta, t, t+h)\, d\zeta = 0, \tag{16}$$

which states that as $h \to 0$ the probability of an appreciable change during a time interval of length h tends strongly toward zero.

With the notation (15) the following basic relation holds which is an analog to the operation of convolution and to the Chapman-Kolmogorov equation (see Chapter IV, p. 214):

$$\phi(z, t+h) = \int \phi(z-\zeta, t) p(\zeta, t, t+h)\, d\zeta, \tag{17}$$

the probability to obtain a change z in the time from 0 to $t + h$ is equal to the sum of all probabilities of changes $z - \zeta$ in the time from 0 to t, multiplied by the probability of the remaining change ζ in the time from t to $t + h$. This equation is called *Smoluchowski's equation.*

(e) *A remark on the stationary case.*[1] Before starting the rather lengthy study of the probability $p(z, t, \tau)$ for the above-defined process, we derive an important mathematical result, valid in the particular case of stationarity (homogeneity in time), whose derivation is almost immediate. We denote by $p(z, t)$ the probability density of an increase z during a time interval t. Equation (17) takes then the simpler form

$$p(z, t_1 + t_2) = \int p(z-\zeta, t_1) p(\zeta, t_2)\, d\zeta, \tag{18}$$

which has the form of a convolution. We introduce the characteristic function of $p(z, t)$, viz.,

$$f(u, t) = \int e^{ixu} p(x, t)\, dx. \tag{19}$$

[1] See examples in Gnedenko [10], p. 279.

In terms of the characteristic functions, Eq. (18) reads

$$f(u, t_1 + t_2) = f(u, t_1)f(u, t_2). \tag{20}$$

Now consider n equal non-overlapping time intervals, each of length $1/n$; we obtain from (20)

$$f(u, t) = \left[f\left(u, \frac{t}{n}\right)\right]^n. \tag{21}$$

Using the terminology introduced in Chapter VI, Section 9.2 we state:

The distribution of an additive and time-homogeneous process is infinitely divisible.

(f) *Explicit expression for $p(z, t, \tau)$.* We return to our not necessarily time-homogeneous problem and apply Taylor's formula to $\phi(z, t + h)$:

$$\phi(z, t + h) = \phi(z, t) + \left(\frac{\partial \phi}{\partial t}\right)_{z,t} \cdot h + o(h). \tag{22}$$

Likewise, by Taylor's formula (the argument if not noted otherwise is z, t),

$$\phi(z - \zeta, t) = \phi(z, t) - \zeta \frac{\partial \phi}{\partial z} + \tfrac{1}{2}\zeta^2 \frac{\partial^2 \phi}{\partial z^2} + \tfrac{1}{2}\zeta^2 \left[\left(\frac{\partial^2 \phi}{\partial z^2}\right)_{z-\theta\zeta,t} - \frac{\partial^2 \phi}{\partial z^2}\right]. \tag{23}$$

We substitute (23) into the right-hand side of (17) and use the abbreviation $p(\zeta, t, t + h) = p$. Then

$$\int \phi(z - \zeta, t)p(\zeta, t, t + h)\, d\zeta = \phi(z, t) \int p\, d\zeta - \frac{\partial \phi}{\partial z} \int \zeta p\, d\zeta$$
$$+ \frac{1}{2}\frac{\partial^2 \phi}{\partial z^2} \int \zeta^2 p\, d\zeta + \frac{1}{2} \int \zeta^2 [\;] p\, d\zeta,$$

where [ ] stands for the last bracket in (23). Since $\int p\, d\zeta = 1$, $\int p\zeta\, d\zeta = 0$, we obtain using (13)

$$\int \phi(z - \zeta, t)p\, d\zeta = \phi(z, t) + \frac{1}{2}\frac{\partial^2 \phi}{\partial z^2} r(t, t + h) + J$$
$$J = \frac{1}{2} \int \zeta^2 \left[\left(\frac{\partial^2 \phi}{\partial z^2}\right)_{z-\theta\zeta,t} - \frac{\partial^2 \phi}{\partial z^2}\right] p\, d\zeta. \tag{24}$$

In order to estimate J we divide the interval of integration into those parts for which $|\zeta| \leqslant \delta$ and those for which $|\zeta| > \delta$. If $|\zeta| \leqslant \delta$ the difference in the bracket in J is as small as we please if δ is sufficiently small, on account of the uniform continuity of $\partial^2\phi/\partial z^2$; this part of the integral is then small, that is, $o(h)$. In the second part, where $|\zeta| > \delta$, we apply (L); since $\partial^2\phi/\partial z^2$ is uniformly bounded, we see that this part of J is also $o(h)$; therefore J is $o(h)$.

We replace now the left- and the right-hand members of (17) by means of (22) and (24), respectively, and obtain

$$\frac{\partial\phi}{\partial t}\cdot h = \frac{1}{2}\frac{\partial^2\phi}{\partial z^2}r(t, t+h) + o(h).$$

Now (14) for $0, t, t+h$ reads

$$r(0, t) + r(t, t+h) = r(0, t+h),$$

$$\frac{\partial\phi}{\partial t}\cdot h = \frac{1}{2}\frac{\partial^2\phi}{\partial z^2}[r(0, t+h) - r(0, t)] + o(h).$$

Now using the notation (15) and setting $r'(t) = \rho(t)$, we have as $h \to 0$

$$\frac{\partial\phi}{\partial t} = \frac{1}{2}\frac{\partial^2\phi}{\partial z^2}\rho(t). \tag{25}$$

The new variable

$$t' = r(t) = \int_0^t \rho(\tau)\,d\tau, \qquad \frac{dt'}{dt} = \rho(t)$$

transforms (25) into the heat equation

$$\frac{\partial\phi}{\partial t'} = \frac{1}{2}\frac{\partial^2\phi}{\partial z^2}, \tag{26}$$

which is satisfied by

$$\phi(z, t') = \frac{1}{\sqrt{2\pi t'}}\,e^{-z^2/2t'}, \qquad \text{where} \qquad \int\phi(z, t')\,dz = 1. \tag{26'}$$

This is the only solution of (26) which satisfies the last condition in (26′). Hence *we have found*

$$p(z, 0, t) = \frac{1}{\sqrt{2\pi r(t)}}\,e^{-z^2/2r(t)} \tag{27}$$

as the only solution of the above-defined additive stochastic process, under the assumed conditions. As in the case of the Poisson law, we see here again that *a distribution* which in the theory of summation of random variables was obtained as a limit distribution, *appears, in the theory of stochastic processes, as the exact solution of a functional equation.* In the expression for $p(z, t, \tau)$ which is analogous to (27), $r(t, \tau)$ takes the place of $r(t)$.