FOURIER ANALYSIS OF DISTRIBUTION FUNCTIONS.
A MATHEMATICAL STUDY OF THE LAPLACE-GAUSSIAN LAW.

BY

CARL-GUSTAV ESSEEN

in UPPSALA.

Contents.

PART II.

Probability Distributions in More than One Dimension.

Introduction.

For about two hundred years the *normal*, or, as it also is called, the *Laplace-Gaussian distribution function*

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{y^2}{2}} dy$$

has played an important rôle in the theory of probability and its statistical applications. Thus, for instance, the distribution of the random errors in a series of equivalent physical measurements may with good approximation be represented by $\Phi\left(\frac{x}{\sigma}\right)$, σ being the dispersion. To explain this many hypotheses have been proposed. One of the most convincing is the *hypothesis of elementary errors*, introduced by HAGEN and BESSEL. According to this hypothesis the error of a measurement etc. is regarded as the sum of a large number of independent errors, so-called elementary errors. Let $X_1, X_2, \ldots$ be a sequence of one-dimensional random variables (r. v.), each variable representing, for instance, an elementary error, with the same or different distribution functions (d. f.), the mean value zero and the finite dispersions σ_i $(i = 1, 2, \ldots)$. If

$$s_n^2 = \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2$$

and $\overline{F_n(x)}$ is the d. f. of

$$Z_n = \frac{X_1 + X_2 + \cdots + X_n}{s_n},$$

then under certain conditions $\overline{F_n(x)}$ is approximately equal to $\Phi(x)$ for large values of n. This is the *Central Limit Theorem* of the theory of probability.

From this we infer that it is of fundamental importance in the theory of probability and mathematical statistics to determine the range of validity of the Central Limit Theorem. LAPLACE [1] [1] and others having formulated the theorem more or less explicitly as early as about the end of the eighteenth century, it was first proved under fairly general conditions by the Russian mathematicians TCHEBYCHEFF, MARKOFF and LIAPOUNOFF [1, 2]. Liapounoff used what are now called *characteristic functions* (c. f.), the others employed the moments of the

[1] [ ] refers to the bibliography at the end of the work.

distributions. If $F(x)$ is a d. f., the c. f. $f(t)$ is the Fourier-Stieltjes transform of $F(x)$:

$$f(t) = \int_{-\infty}^{\infty} e^{itx}\, dF(x).$$

LIAPOUNOFF also succeeded in estimating the remainder term, showing that

$$|\overline{F_n(x)} - \Phi(x)| \leq K \cdot \frac{\log n}{\sqrt{n}}, \tag{1}$$

K being independent of n under certain conditions.

The inequality (1) has later been studied by CRAMÉR [1, 3, 5]. β_{3i} being the third absolute moment of X_i, and the quantities B_{2n} and B_{3n} being defined by

$$B_{2n} = \frac{1}{n}(\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2),$$

$$B_{3n} = \frac{1}{n}(\beta_{31} + \beta_{32} + \cdots + \beta_{3n}),$$

he shows that

$$|\overline{F_n(x)} - \Phi(x)| \leq 3 \cdot \frac{B_{3n}}{B_{2n}^{3/2}} \cdot \frac{\log n}{\sqrt{n}}. \tag{2}$$

In recent years important works on the Central Limit Theorem have further been performed by LINDEBERG [1], LÉVY [1, 2], KHINTCHINE [1] and others.

Though the normal d. f. may often be used with good approximation to represent the distribution of a statistical material, there are many cases where the agreement is not satisfactory. To obtain a better result it has been proposed to expand the d. f. $F(x)$ in a series of $\Phi(x)$ and its derivatives[1] (we suppose the mean value $= 0$ and the dispersion $= 1$):

$$F(x) = \Phi(x) + \frac{c_3}{3!}\Phi^{(3)}(x) + \cdots + \frac{c_\nu}{\nu!}\Phi^{(\nu)}(x) + \cdots, \tag{3}$$

the coefficient c_ν being determined by

$$c_\nu = (-1)^\nu \int_{-\infty}^{\infty} H_\nu(x)\, dF(x),$$

where $H_\nu(x)$ is the νth Hermite polynomial:

[1] or, to be exact, the frequency function $F'(x)$ in a series of $\Phi'(x)$ and its derivatives.

$$H_\nu(x) = (-1)^\nu e^{\frac{x^2}{2}} \frac{d^\nu}{dx^\nu}\left(e^{-\frac{x^2}{2}}\right).$$

The coefficient c_ν only depends on the first ν moments of $F(x)$.

The expansion (3) was introduced by BRUNS [1], EDGEWORTH [1], CHARLIER [1, 2, 3, 4] and others, being called by Charlier an A series, as distinguished from another expansion by Charlier, the B series.[1] It is possible to deduce the A series in a formal way, as Edgeworth and Charlier did, by using the hypothesis of elementary errors. A more rigorous mathematical proof was needed, however. In the applications there was nevertheless often a good agreement between $F(x)$ and the sum of the first terms in (3).

A question that naturally arose was that of the convergence properties of the A series. Compare CRAMÉR [2]. The essential question, however, is the asymptotic behaviour of the partial sums in (3) and the order of magnitude of the remainder term. Starting from the hypothesis of elementary errors, regarding $F(x)$ as the d. f. of the variable

$$\frac{X_1 + X_2 + \cdots + X_n}{\sigma \sqrt{n}},$$

where the variables X_i are mutually independent and for the sake of simplicity each X_i has the same d. f. with the mean value zero, the dispersion $\sigma \neq 0$ and finite moments of arbitrarily high order, it is easily seen that[2]

$$c_\nu = O\left(\frac{1}{n^{\frac{\nu}{2} - \left[\frac{\nu}{3}\right]}}\right), \qquad (\nu = 3, 4, 5, \ldots).$$

If α_ν is the moment of order ν of X_i it is found that

$$c_3 = -\frac{\alpha_3}{\alpha_2^{3/2}} \cdot \frac{1}{n^{1/2}}, \qquad c_4 = \frac{\alpha_4 - 3\alpha_2^2}{\alpha_2^2} \cdot \frac{1}{n},$$

$$c_5 = -\frac{\alpha_5 - 10\alpha_2\alpha_3}{\alpha_2^{5/2}} \cdot \frac{1}{n^{3/2}}, \qquad c_6 = \frac{10\alpha_3^2}{\alpha_2^3} \cdot \frac{1}{n} + \frac{\alpha_6 - 15\alpha_2\alpha_4 - 10\alpha_3^2 + 30\alpha_2^3}{\alpha_2^3} \cdot \frac{1}{n^2}.$$

Every c_ν $(\nu > 5)$ generally contains different powers of $\frac{1}{\sqrt{n}}$. After a rearrangement of (3) according to powers of $\frac{1}{\sqrt{n}}$ it becomes equal to

[1] The series derived by Edgeworth was not formally identical with (3) but a rearrangement of it. Compare (4).

[2] CRAMÉR [2, 3],

$$F(x) = \Phi(x) + \frac{p_1(x)}{n^{1/2}} e^{-\frac{x^2}{2}} + \frac{p_2(x)}{n} e^{-\frac{x^2}{2}} + \cdots + \frac{p_\nu(x)}{n^{\nu/2}} e^{-\frac{x^2}{2}} + \cdots, \tag{4}$$

where $p_\nu(x)$ is a polynomial in x, the coefficients of which are only dependent on the moments α. This is the development of Edgeworth (in the following we call it the *Edgeworth expansion*) which has later been studied by CRAMÉR [3, 5], especially with regard to the order of magnitude of the remainder term.

The function $f(t)$ being the c. f. of each variable X_i, the *Cramér condition* (C) implies that

$$\overline{\lim_{t \to \pm\infty}} |f(t)| < 1, \tag{C}$$

this for instance being the case when the d. f. of X_i contains an absolutely continuous component. Provided that the condition (C) is satisfied, Cramér shows that

$$F(x) = \Phi(x) + \sum_{\nu=1}^{k-3} \frac{p_\nu(x)}{n^{\nu/2}} e^{-\frac{x^2}{2}} + O\left(\frac{1}{n^{\frac{k-2}{2}}}\right), \qquad (k \text{ an integer} \geq 3). \tag{5}$$

In this case the expansion (4) may be regarded as an asymptotic series. It follows from (5) with $k = 3$, that the Liapounoff remainder term in (1) can be improved to $O\left(\frac{1}{\sqrt{n}}\right)$ if the condition (C) is satisfied.

Contents of Part I.

It has been supposed that $\log n$ is on the whole superfluous in (1), but this was not proved until a few years ago. It follows from a somewhat more general theorem of the present work (Chap. III, Theorem 1) that (2) can be replaced by

$$|\overline{F_n(x)} - \Phi(x)| \leq 7.5 \cdot \frac{B_{3n}}{B_{2n}^{3/2}} \cdot \frac{1}{\sqrt{n}}. \tag{6}$$

The inequality (6) was proved in an earlier work of the present author[1] and at the same time by BERRY[2] [1], independently of each other.

One of the *main problems* of this work may be formulated thus: Given a sequence of independent r. v.'s $X_1, X_2, \ldots$ all having the same[3] d. f. $F(x)$ with

[1] ESSEEN [1].

[2] This work is not yet accessible in Sweden. I have found in a review in Mathematical Reviews, 2 (1941) p. 228, that an inequality like (6) is to be found in BERRY [1]. My own proof of (6) was completed in the autumn of 1940. See ESSEEN [1].

[3] As we are most interested in principles we generally restrict ourselves to this case.

mean value zero, the dispersion $\sigma \neq 0$ and some finite absolute moments of higher order, study the d. f. $F_n(x)$ of the variable

$$\frac{X_1 + X_2 + \cdots + X_n}{\sigma \sqrt{n}}$$

as $n \to \infty$, especially the *remainder term problem*. A complete discussion of this question necessitates the introduction of a certain class of d. f.'s called *lattice distributions*. A d. f. is a lattice distribution if it is purely discontinuous, the jumps belonging to a sequence of equidistant points. This is one of the most usual types of d. f. met with in the applications.

Three different cases may occur, which together cover all possibilities.

1. The condition (C) is satisfied. Then the expansion (5) holds.

2. $F(x)$ is a lattice distribution. Even if all moments are finite an expansion like (5) is impossible with $k > 3$, there being jumps of $F_n(x)$ of order of magnitude $\frac{1}{\sqrt{n}}$. By adding an expression to (5) containing a discontinuous function, it is possible to diminish the order of magnitude of the remainder term.

3. Condition (C) is not satisfied and the distribution is not of lattice type. It is found that

$$F_n(x) = \Phi(x) + \frac{\alpha_3}{6\sigma^3\sqrt{2\pi n}}(1 - x^2)e^{-\frac{x^2}{2}} + o\left(\frac{1}{\sqrt{n}}\right),$$

α_3 being the third moment of X_i.

These questions are investigated in Chapter IV and the results make it possible to determine the asymptotic maximum deviation of $F_n(x)$ from $\Phi(x)$.

In Chapter V we study the dependence of the remainder term on n and on x; the results are applied to the so-called Uniform Law of Great Numbers.

The theorems of Chapters III—V are based on a theorem concerning the connection between the difference of two d. f.'s and the difference between their c. f.'s. The proof is given in Chapter II. In proving the inequalities (1) and (2) respectively, Liapounoff and Cramér used a convolution method. Liapounoff considered the convolution of the difference between the d. f.'s with a convenient normal d. f., while Cramér applied Riemann-Liouville integrals. By these methods, however, it is not possible to obtain the real order of magnitude of the remainder term. In proving the inequality (6) and others, consequences of the main theorem of Chapter II, we consider the convolution with a function having the Fourier-Stieltjes transform equal to zero outside a finite interval. It is just this property of the transform that is essential.

The c. f.'s being the most important analytical implements of this work, an account of their theory is given in Chapter I. Many of the theorems stated here are well known but are included for the sake of continuity. A closer study is devoted to certain questions, for instance the problem whether two c. f.'s equal to each other in an interval about the zero point are identical or not. The c. f.'s form a sub-class of a more general set of functions, the class (T). We begin Chapter I by investigating these functions.

Contents of Part II.

In Part II we study the Central Limit Theorem and the remainder term problem for r. v.'s in k dimensions. Concerning the remainder term problem there have hitherto been only rough estimations. The results are applied to the so-called χ^2 method. It follows from Chapter VIII that the remainder term problem is intimately connected with the lattice point problem of the analytic theory of numbers. For further information the reader is referred to the introduction of Chapter VII.

I take the opportunity of expressing my warmest thanks to Prof. Arne Beurling, Uppsala, for suggesting this investigation and for his kind interest and valuable advice in the course of the work.

PART I.

Distribution Functions of One Variable.

Chapter I.

Functions of Bounded Variation and Their Fourier-Stieltjes Transforms.

The concept of the distribution function plays an important rôle not only in the theory of probability but also in several other branches of mathematics. As an introduction to the study of this class of functions we shall, however, devote the first sections of this chapter to a treatment of a more general class, the functions of bounded variation. For the proofs of several theorems mentioned below reference is made to Beurling [1], Bochner [1] and Cramér [5]. In the following treatment Lebesgue or Lebesgue-Stieltjes integrals are used.

1. **Functions of bounded variation.** Let $F(x)$ be a real or complex-valued function of the real variable x and of bounded variation on the whole real axes:

$$V(F) = \int_{-\infty}^{\infty} |dF(x)| < \infty. \tag{1}$$

Further let $F(-\infty) = 0$. It is well known that $F(x)$ has at the most an enumerable set of discontinuity points. In such a point we define

$$F(x) = \tfrac{1}{2}[F(x+0) + F(x-0)].$$

The class of all such functions is denoted by (V). A sub-class (V_P) consists of those real functions $F(x) < (V)$ which are non-decreasing. If $F(x) < (V_P)$ and $F(+\infty) = 1$, then $F(x)$ is a distribution function (d. f.).

By a well-known theorem of Lebesgue, $F(x)$, belonging to (V), can be represented as the sum of three components:

$$F(x) = F_1(x) + F_2(x) + F_3(x),$$

where $F_1(x)$ is absolutely continuous, $F_2(x)$ singular, i. e. continuous and having the derivative $= 0$ almost everywhere, and where $F_3(x)$ is the step function, i. e. constant in every interval of continuity of $F(x)$ and having the jump $F(x+0) - F(x-0)$ at every point of discontinuity. Hence it is convenient to divide (V) into three sub-classes:

$$(V) = (V_1) + (V_2) + (V_3),$$

(V_1) being the class of absolutely continuous functions etc. In the same way

$$(V_P) = (V_{P_1}) + (V_{P_2}) + (V_{P_3}).$$

By the *point spectrum* Q of a function $F(x) < (V)$ we understand the set of those points x for which $F(x+0) - F(x-0) \neq 0$.[1] The set Q is at the most enumerable and may be empty. By the vectorial sum $Q = Q_1 + Q_2$ of two such sets we understand the set of those points x which may be written $x = x_1 + x_2$ where $x_1 < Q_1$ and $x_2 < Q_2$. By definition Q is empty if either Q_1 or Q_2 is empty.

The concept of convolution (»Faltung») of two functions plays an important rôle in this work.

[1] This definition is in accordance with the terminology of Wintner, see WINTNER [1, 2]. It would perhaps be more correct to call Q the point spectrum of the Fourier-Stieltjes transform of $F(x)$.

For every pair of functions $F_1(x)$ and $F_2(x)$ belonging to (V) with point spectra Q_1 and Q_2 there exists a uniquely determined function $F(x)$ belonging to (V) with point spectrum Q such that the convolution

$$F_1 * F_2 = \int_{-\infty}^{\infty} F_1(x-y)\,d\,F_2(y) = \int_{-\infty}^{\infty} F_2(x-y)\,d\,F_1(y) \tag{2}$$

exists and is equal to $F(x)$ for every x not contained in $Q_1 + Q_2$. If $x < Q_1 + Q_2$ then $F(x)$ is defined by

$$F(x) = \tfrac{1}{2}[F(x+0) + F(x-0)].$$

Further $Q < Q_1 + Q_2$.

If F_1 and F_2 belong to (V_P), then F belongs to (V_P) and $Q = Q_1 + Q_2$. If F_1 and F_2 are d.f.'s, F is also a d.f.

The convolution of three functions F_1, F_2 and F_3 is defined by: $F_1 * F_2 * F_3 = F_1 * (F_2 * F_3)$. Correspondingly for n functions. The convolution operation is easily found to be commutative. If $F_1, F_2, \ldots, F_n$ belong to (V_P) and Q is the point spectrum of their convolution, then $Q = Q_1 + Q_2 + \cdots + Q_n$.

If either of the functions F_1 and F_2 in (2) is continuous, F is also continuous. This explains why Q is defined as empty if Q_1 or Q_2 is empty. It is, however, possible further to specialize the continuity properties of F. It is easily seen that if F_1 or F_2 belongs to (V_1) then $F < (V_1)$, that if F_1 or F_2 belongs to (V_2) then F belongs to (V_2) or $(V_1 + V_2)$, that if both F_1 and F_2 belong to (V_3) then F belongs to (V_3) or is constant.

By (1) and (2) we obtain the following important inequalities:

$$\begin{cases} V(F_1 + F_2) \leq V(F_1) + V(F_2) \\ V(F_1 * F_2) \leq V(F_1)\, V(F_2). \end{cases} \tag{3}$$

Finally we introduce that concept of *convergence* which is especially convenient for the (V)-class. A sequence of functions $\{F_n(x)\}$ belonging to (V) is said to converge to a function $F(x) < (V)$ if $\lim\limits_{n\to\infty} F_n(x) = F(x)$ at every point of continuity of $F(x)$.

2. **Functions of class** (T). If it is possible to represent a function $f(t)$ of the real variable t as the Fourier-Stieltjes transform of a function $F(x) < (V)$,

$$f(t) = \int_{-\infty}^{\infty} e^{itx}\,d\,F(x), \tag{4}$$

$f(t)$ by definition belongs to the class (T).[1] If $F(x) < (V_P)$ we say that $f(t)$ belongs to (T_P). It is immediately clear that $f(t) < (T)$ is a uniformly continuous and bounded function:

$$|f(t)| \leq \int_{-\infty}^{\infty} |dF(x)| = V(F) < \infty.$$

In the following we denote a function $< (V)$ and its transform by the same letters, capital letters for (V)- and small letters for (T)-functions. Between classes (V) and (T) there is a one-to-one correspondence, as the following well-known *inversion theorem* shows:

If $F(x) < (V)$ *and* $f(t) = \int_{-\infty}^{\infty} e^{itx}\, dF(x)$, *then*

$$F(x) - F(\xi) = \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-i\xi t} - e^{-ixt}}{it} f(t)\, dt, \tag{5}$$

$$F(x+0) - F(x-0) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} e^{-ixt} f(t)\, dt. \tag{6}$$

Corresponding to the decomposition $F = F_1 + F_2 + F_3$ we obtain $f = f_1 + f_2 + f_3$, f_1 being the transform of the absolutely continuous part F_1 of F etc. The component f_1 is called the ordinary, f_2 the singular and f_3 the almost periodic part of f. In conformity to this decomposition we put $(T) = (T_1) + (T_2) + (T_3)$, (T_1) being the class of ordinary functions of (T) etc. Correspondingly $(T_P) = (T_{P_1}) + (T_{P_2}) + (T_{P_3})$.

For later purposes it is desirable to investigate the properties of $f(t)$ for large values of t. The regularity of $F(x)$ is here of predominant importance. We only consider the case where $F(x)$ contains but one component.

a. $F(x) < (V_1)$. Then it follows by the Riemann-Lebesgue theorem that

$$\lim_{t\to\pm\infty} |f(t)| = 0.$$

b. $F(x) < (V_3)$. Denote by a_ν the jump of $F(x)$ at $x = x_\nu$, $(\nu = 0, \pm 1, \pm 2, \ldots)$. Then $\sum |a_\nu| = V(F) < \infty$ and

[1] This notation is used in BEURLING [1].

$$f(t) = \sum_{\nu} a_\nu e^{i x_\nu t}.$$

Thus $f(t)$ is almost periodic and $\overline{\lim\limits_{t \to \pm\infty}} |f(t)| > 0$, provided that $f(t) \not\equiv 0$.

c. $F(x) < (V_2)$. Both cases, $\lim\limits_{t \to \pm\infty} |f(t)| = 0$ and $\overline{\lim\limits_{t \to \pm\infty}} |f(t)| > 0$, may occur. On the whole the singular transforms have hitherto been very little known. However, $|f(t)|$ is small in mean, this being a consequence of:

$$\lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} |f(t)|^2 \, dt = 0.$$

Later we shall return to cases a—c and investigate them more closely supposing $F(x)$ to be a d. f.

The study of the convolution of functions in (V) is considerably facilitated by passing over to the transforms. This depends on the following *convolution theorem:*

If F_1 and F_2 belong to (V), f_1 and f_2 being their transforms, then $f_1 f_2$ is the transform of $F_1 * F_2$. Conversely, if f_1 and f_2 belong to (T). $f_1 f_2$ also belongs to (T), being the transform of $F_1 * F_2$.

The extension to n functions is immediate.

When studying the (T)-functions it is often convenient to introduce the metric $T(f)$, defined by

$$T(f) \equiv V(F). \tag{7}$$

This has been done by Bochner [2] and Beurling [1], the latter having obtained important results by the comparison of $T(f)$ with $M(f) = \overline{\mathrm{Bound}}_{-\infty < t < \infty} |f(t)|$. From (3), (7) and the convolution theorem it follows:

$$\begin{cases} T(f_1 + f_2) \leq T(f_1) + T(f_2) \\ \quad T(f_1 f_2) \leq T(f_1)\, T(f_2). \end{cases} \tag{8}$$

A very important class of functions, especially with regard to the applications, is formed by those *entire functions of exponential type* which on the real axes belong to the Lebesgue class $L(-\infty, \infty)$. An entire function $H(z)$ is of exponential type a if

$$H(z) = O(e^{a|z|}), \qquad (a > 0, |z| \to \infty).$$

The following lemma holds[1]:

[1] See for example Plancherel-Pólya [1], p. 229.

Lemma 1. *If $H(z)$, $z = x + iy$, is an entire function of exponential type a, if $H(x)$ belongs to $L(-\infty, \infty)$ and if $h(t) = \int_{-\infty}^{\infty} e^{itx} H(x)\,dx$, then $h(t) = 0$ for $|t| \geq a$.*

Finally we may touch upon the *convergence* in (T). Consider a sequence of functions $\{f_n(t)\}$ $(n = 1, 2, \ldots)$, belonging to (T), such that $T(f_n) \leq K < \infty$ for every n. Such a sequence may converge to a function not belonging to (T), for instance the sequence $\{e^{-nt^2}\}$. We quote the following *convergence theorem*[1]:

Let $\{f_n(t)\}$ be a sequence of functions belonging to (T) such that $T(f_n) \leq K < \infty$ for every n. A sub-sequence can always be chosen, converging to a function $f(t)$ which almost everywhere and at every continuity point is equal to a function $g(t) < (T)$, such that $T(g) \leq K$. — If $\{f_n(t)\}$ converges to a continuous function $f(t)$, then $f(t) < (T)$ and $T(f) \leq K$.

3. **Minimum extrapolation in** (T). Consider a function $f(t)$ defined on a set e which is made up by a sum of intervals, and suppose $f(t)$ to be continuous on e. Further, suppose that there exists a function $g(t) < (T)$ such that $f(t) = g(t)$ on e. Then we say that $f(t)$ *belongs to* (T) *on* e and we put by definition

$$T_e(f) = \underline{\text{Bound}}\, T(g), \tag{9}$$

g running through all the functions in (T) equal to f on e. By the convergence theorem, § 2, it follows that, if $f(t) < T$ on e, there is at least one function $f_e(t) < T$ such that

$$T_e(f) = T(f_e).$$

The function $f_e(t)$ is called the *minimum extrapolation of f with regard to e* (in French »prolongement minimal»). Later we shall show by examples that a minimum extrapolation need not be uniquely determined. The concept of minimum extrapolation has been introduced by Beurling[2] and is of great importance in many questions.

With regard to a later application we shall briefly determine the minimum extrapolation of a certain function.

Theorem 1. *If $f(t) = -\dfrac{1}{it}$ on the set $\omega(|t| \geq T)$, $f(t)$ belongs to (T_1) with regard to ω and*

[1] Beurling [1], p. 4.

[2] Beurling [1], p. 4. This paper is a summary of an earlier work of Beurling, presented to the University of Uppsala in 1936. See Beurling [1], p. 1.

$$T_\omega(f) = \frac{\pi}{2}\cdot\frac{1}{T}.$$

Without loss of generality we may suppose $T = 1$. We put

(10)
$$f_1(t) = \begin{cases} g(t) \text{ for } |t| \leq 1 \\ -\frac{1}{it} \quad » \quad |t| \geq 1, \end{cases}$$

$g(t)$ being chosen so that $f_1(t)$ is continuous and belongs to (T). As $f_1(t) < < L^2(-\infty, \infty)$, it is readily seen that $f_1(t) < (T_1)$, i. e.

$$f_1(t) = \int_{-\infty}^{\infty} e^{itx} F_1'(x)\,dx.$$

Hence from (10) and the Fourier inversion formula:

$$F_1'(x) = \frac{1}{\pi}\int_1^\infty \frac{\sin xt}{t}\,dt + \frac{1}{2\pi}\int_{-1}^{1} e^{-ixt} g(t)\,dt$$

or

$$F_1'(x) = \tfrac{1}{2}\operatorname{sign} x - \frac{1}{\pi}\int_0^1 \frac{\sin xt}{t}\,dt + \frac{1}{2\pi}\int_{-1}^{1} e^{-ixt} g(t)\,dt.$$

Hence

(11)
$$F_1'(x) = \tfrac{1}{2}\operatorname{sign} x - H_1(x),$$

where

(12)
$$H_1(x) = \frac{1}{\pi}\int_0^1 \frac{\sin xt}{t}\,dt - \frac{1}{2\pi}\int_{-1}^{1} e^{-ixt} g(t)\,dt.$$

From (12) it follows that $H_1(x)$ is an entire function of exponential type 1. According to (11) the problem is now to determine an entire function $H_1(x)$ of exponential type 1 such that

(13)
$$T(f_1) = \int_{-\infty}^{\infty} \left|\tfrac{1}{2}\operatorname{sign} x - H_1(x)\right| dx = \text{min.} = T_\omega(f).$$

It is easily found that we may choose $H_1(x)$ to be an *odd* function, and further that $H_1'(x) = H(x)$ belongs to $L(-\infty, \infty)$. Conversely, for every such function $H_1(x)$ the transform of $\frac{1}{2}\operatorname{sign} x - H_1(x)$ satisfies the conditions on $f_1(t)$. Now it evolves that our problem is connected with a theorem of Bohr [1] concerning exponential polynomials:

If $\varphi(x) = \sum_{\nu=1}^{n} a_\nu e^{i\lambda_\nu x}$, λ_ν *being real numbers,* $|\lambda_\nu| \geq 1$ *and* $|\varphi'(x)| \leq 1$, *then*

$$|\varphi(x)| \leq \frac{\pi}{2}. \tag{14}$$

In order to show that such a function $\varphi(x)$ is bounded by an absolute constant we use an argument that has been given in lectures by Prof. Beurling. Let $H(x)$ be an even entire function of exponential type 1, belonging to $L(-\infty, \infty)$ on the real axes. If $H_1(x) = \int_0^x H(y)\,dy$ we further suppose that

$$\int_0^\infty \left|\tfrac{1}{2} - H_1(y)\right| dy < \infty. \tag{15}$$

Now from Lemma 1;

$$\int_{-\infty}^{\infty} e^{i\lambda x} H(x)\,dx = 0 \text{ for } |\lambda| \geq 1.$$

Hence

$$\int_{-\infty}^{\infty} \varphi(x-y) H(y)\,dy = 0$$

and further

$$\varphi(x) = \int_{-\infty}^{\infty} \varphi(x-y)\, d\,\tfrac{1}{2} \operatorname{sign} y.$$

By subtraction we obtain

$$\varphi(x) = \int_{-\infty}^{\infty} \varphi(x-y)\, d\left[\tfrac{1}{2} \operatorname{sign} y - H_1(y)\right].$$

Partial integration gives with regard to (15):

$$\varphi(x) = \int_{-\infty}^{\infty} \varphi'(x-y)\left[\tfrac{1}{2} \operatorname{sign} y - H_1(y)\right] dy, \tag{16}$$

or

$$|\varphi(x)| \leq \int_{-\infty}^{\infty} \left|\tfrac{1}{2} \operatorname{sign} y - H_1(y)\right| dy. \tag{17}$$

According to (17) all functions $\varphi(x)$ are bounded by an absolute constant. It remains to be proved, however, that $\frac{\pi}{2}$ is the best possible constant.

On comparison of (13) and (17) the equivalence of the two problems follows. In both cases the function $H_1(x)$ has to satisfy the same conditions: to be an odd entire function of exponential type 1 and to make the integral (15) convergent and as small as possible.

There are several proofs of Bohr's theorem. I shall, however, give one more proof, the minimum extrapolation being thus determined.[1]

Let us consider the function $\psi(x)$ with the period 2π represented in Fig. 1. By the expansion of $\psi(x)$ in a Fourier series it is easily found that $\psi(x)$ meets all conditions in Bohr's theorem. It is very probable that $\psi(x)$ is the extremal function. Under all circumstances it follows from $\psi\left(\frac{\pi}{2}\right)=\frac{\pi}{2}$ that

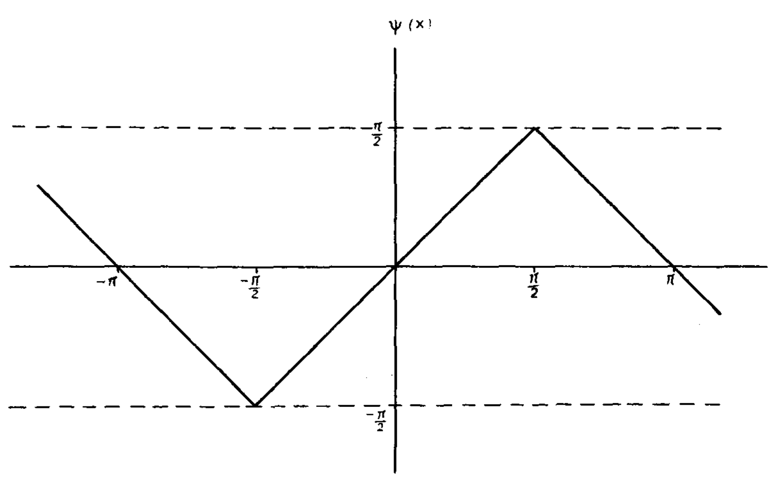

Fig. 1.

$$\frac{\pi}{2} \leq \int_{-\infty}^{\infty} \left|\tfrac{1}{2}\operatorname{sign} y - H_1(y)\right| dy. \tag{18}$$

We shall now show that it is possible to determine H_1 in such a manner that there is equality in (18). Then Theorem 1 and Bohr's theorem are proved.

Let $\varphi \equiv \psi$ in (16) and (17) and $x = \frac{\pi}{2}$. $\psi'(x)$ being ± 1 with the period

[1] A similar method has been used by von Sz. Nagy-Strausz [1] who give a proof of Bohr's theorem using the minimum extrapolation of $-\frac{1}{it}$ with regard to $|t| \geq 1$. However, the rôle of the minimum extrapolation is not explicitly stated; further, since our determination of the minimum extrapolation is not the same as in the cited paper and may have an interest of its own, I have found it convenient to treat the question once more.

2π, in order not to increase the value of the integral by the step $(16) \to (17)$ it is necessary that the variations of sign of $\frac{1}{2}$ sign $y - H_1(y)$ occur in accordance with Fig. 2.

Thus we form an odd entire function $H_1(z)$, $z = x + iy$, satisfying the conditions

$$(19) \qquad \begin{cases} 1^\circ \; H_1(z) = O(e^{|z|}), \\ 2^\circ \; H_1'(x) < L(-\infty, \infty), \\ 3^\circ \; \tfrac{1}{2} \operatorname{sign} x - H_1(x) = 0 \text{ for } x = n\pi \qquad (n = \pm 1, \pm 2, \pm 3, \pm \cdots), \\ 4^\circ \; C = \int\limits_{-\infty}^{\infty} |\tfrac{1}{2} \operatorname{sign} x - H_1(x)|\, dx < \infty. \end{cases}$$

We will show that (19) uniquely determines H_1 and that $C = \frac{\pi}{2}$.

We prefer, however, to consider the function

$$(20) \qquad G(z) = \tfrac{1}{2} - H_1(\pi z).$$

From (19) and (20) we obtain:

$$(21) \qquad \begin{cases} 1^\circ \; G(z) = O(e^{\pi |z|}), \\ 2^\circ \; G(z) = \begin{cases} 0 \text{ for } z = 1, 2, \ldots \\ \tfrac{1}{2} \quad \text{»} \quad z = 0, \\ 1 \quad \text{»} \quad z = -1, -2, \ldots \end{cases} \\ 3^\circ \quad C = 2\pi \int\limits_{0}^{\infty} |G(x)|\, dx < \infty. \end{cases}$$

By an interpolation formula of VALIRON [1] we obtain from (21):

$$(22) \qquad G(z) = \frac{\sin \pi z}{\pi} \left(\frac{1}{2z} - \sum_{n=1}^{\infty} \frac{(-1)^n z}{n(z+n)} + a \right),$$

a being a constant later to be determined. But

$$\sum_{n=1}^{\infty} \frac{(-1)^n z}{n(z+n)} = -\log 2 - \sum_{n=1}^{\infty} \frac{(-1)^n}{z+n}.$$

Here we introduce the function

$$(23) \qquad \beta(z) = \sum_{n=0}^{\infty} \frac{(-1)^n}{z+n},$$

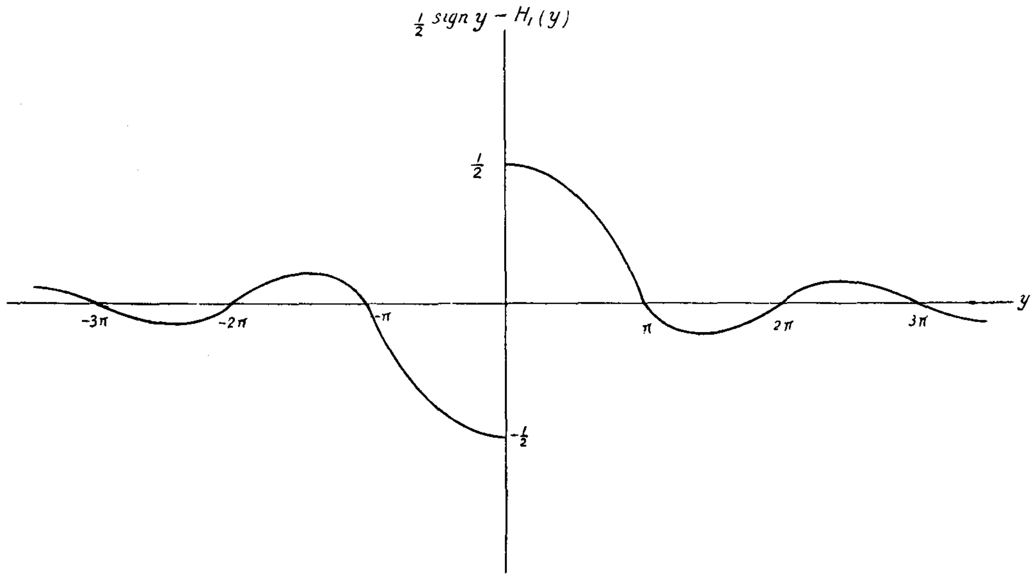

Fig. 2.

known from the theory of the Γ-function. From (22) and (23) we thus obtain:

$$G(z) = \frac{\sin \pi z}{\pi} \left\{ \beta(z) - \frac{1}{2z} + \log 2 + a \right\}.$$

We now observe that

$$\beta(x) - \frac{1}{2x} = \sum_{n=0}^{\infty} \frac{1}{(x+2n)(x+2n+1)(x+2n+2)}. \tag{24}$$

Thus $\beta(x) - \frac{1}{2x} = O\left(\frac{1}{x^2}\right)$ as $x \to \infty$, and hence from (21: 3°) $a + \log 2 = 0$. Thus

$$G(z) = \frac{\sin \pi z}{\pi} \left\{ \beta(z) - \frac{1}{2z} \right\}. \tag{25}$$

From (25) it is easily found that $H_1(z)$ has the required properties. It only remains to evaluate

$$C = 2 \cdot \int_0^{\infty} \left| \left\{ \beta(x) - \frac{1}{2x} \right\} \sin \pi x \right| dx. \tag{26}$$

It is easily seen that the following operations are allowed. According to (24) $\beta(x) - \frac{1}{2x} > 0$ for $x > 0$. Thus

$$C = 2 \cdot \int_0^\infty \left\{\beta(x) - \frac{1}{2x}\right\} |\sin \pi x| \, dx = 2 \cdot \left\{\int_0^1 + \int_1^2 + \int_2^3 + \cdots\right\} =$$

$$= 2 \cdot \int_0^1 \left\{\beta(x) - \frac{1}{2x} + \beta(x+1) - \frac{1}{2(x+1)} + \beta(x+2) - \frac{1}{2(x+2)} + \cdots\right\} \sin \pi x \, dx.$$

Observing that $\beta(x) + \beta(x+1) = \frac{1}{x}$ we obtain:

$$C = 2 \cdot \int_0^1 \left\{\frac{1}{2x} - \frac{1}{2(x+1)} + \frac{1}{2(x+2)} - \cdots\right\} \sin \pi x \, dx = \int_0^\infty \frac{\sin \pi x}{x} dx = \frac{\pi}{2}.$$

Hence the theorem is proved.

4. **A uniqueness theorem in** $(T_2 + T_3)$. We shall later consider the problem of the unique determination of a function in (T_P), knowing its values in an interval about the zero point. Here we shall treat the case where $f(t) < (T)$ is known in an infinite interval $t \leq a$. Without loss of generality we may suppose $a = 0$. If $F(x)$ is real, the solution is immediate, for then $\overline{f(-t)} = f(t)$.

Theorem 2.[1] *A function $f(t) < (T_2 + T_3)$ is uniquely determined by its values in an infinite interval. This need not be true if $f(t) < (T_1)$.*

P r o o f o f T h e o r e m 2. In the proof we may suppose the interval in question to be $t \leq 0$. Now suppose that there is actually another function $f_1(t) < (T)$, equal to $f(t)$ for $t \leq 0$, but not identical with $f(t)$ for $t > 0$ and belonging to the same (T)-class as $f(t)$. Putting $f_2(t) = f(t) - f_1(t)$ we have $f_2(t) = 0$ for $t \leq 0$, $f_2(t) \not\equiv 0$ for $t > 0$. We now use the following theorem, the proof of which will be postponed somewhat.

Theorem 2 a. *If a function belongs to (T) and is equal to zero for $t \leq 0$, it is the Fourier-Stieltjes transform of an absolutely continuous function.*

Hence from Theorem 2 a $f_2(t) = f(t) - f_1(t) < (T_1)$. But if $f(t) < (T_2 + T_3)$ this also holds regarding $f_1(t)$ and $f(t) - f_1(t)$. Thus $f_2(t) \equiv 0$, contrary to hypothesis. On the other hand there are functions belonging to (T_1) which are zero in an infinite interval. See for instance Lemma 1. Thus the theorem is proved.

P r o o f o f T h e o r e m 2 a. Let $f(t)$ satisfy the conditions of Theorem 2 a. By hypothesis

[1] BEURLING [1], p. 4, mentions this theorem incidentally without proof.

(27)
$$f(t) = \int_{-\infty}^{\infty} e^{it\xi}\, dF(\xi),$$

where

(28)
$$\int_{-\infty}^{\infty} |dF(\xi)| = V < \infty.$$

Now we form

(29)
$$G(z) = \frac{1}{2\pi} \int_{0}^{\infty} e^{-izt} f(t)\, dt, \qquad (z = x + iy).$$

Evidently $G(z)$ is analytic and regular for $y < 0$. We suppose from now on that this condition is satisfied. But since $f(t) = 0$ for $t \leq 0$, we may write (29) in the following way:

(30)
$$G(z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ixt} e^{y|t|} f(t)\, dt.$$

From (27) and (30) we obtain:

$$G(z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dF(\xi) \int_{-\infty}^{\infty} e^{y|t|} e^{it(\xi - x)}\, dt$$

or

(31)
$$G(z) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{|y|}{(x-\xi)^2 + y^2}\, dF(\xi).$$

From (31) it follows:

$$\int_{-\infty}^{\infty} |G(x+iy)|\, dx \leq \int_{-\infty}^{\infty} |dF(\xi)| \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{|y|\, dx}{(x-\xi)^2 + y^2}$$

or

(32)
$$\int_{-\infty}^{\infty} |G(x+iy)|\, dx \leq \int_{-\infty}^{\infty} |dF(\xi)| = V < \infty.$$

According to a theorem of Hille-Tamarkin [1] there exists a function $H(x)$ such that:

(33)
$$\begin{cases} 1^\circ \ \lim\limits_{y \to -0} G(x+iy) = H(x) \text{ almost everywhere}, \\ 2^\circ \ \int_{-\infty}^{\infty} |H(x)|\, dx \leq V < \infty, \\ 3^\circ \ \lim\limits_{y \to -0} \int_{-\infty}^{X} G(x+iy)\, dx = \int_{-\infty}^{X} H(x)\, dx. \end{cases}$$

From (31) we further obtain:

$$(34) \qquad \int_{-\infty}^{X} G(x+iy)\,dx = \int_{-\infty}^{\infty} dF(\xi) \int_{-\infty}^{X} \frac{1}{\pi} \frac{|y|\,dx}{(x-\xi)^2+y^2}.$$

On account of the well-known properties of the kernel $\frac{1}{\pi}\frac{|y|}{(x-\xi)^2+y^2}$ we have from (33: 3°) and (34):

$$\tfrac{1}{2}[F(X+0)+F(X-0)] = \int_{-\infty}^{X} H(x)\,dx,$$

or this relation and (33: 2°) show that $F(x)$ is absolutely continuous. Hence the theorem is proved.

There is an analogous theorem in the unit circle.

Theorem 2 b. *Let $\mu(\theta)$ be a function of bounded variation in $(0, 2\pi)$. If the Fourier-Stieltjes coefficients*

$$c_n = \frac{1}{2\pi}\int_0^{2\pi} e^{-in\theta}\,d\mu(\theta), \qquad (n = 0, \pm 1, \pm 2, \pm \ldots),$$

satisfy the condition $c_n = 0$ for $n < 0$, then $\mu(\theta)$ is absolutely continuous.

The proof is similar to that of Theorem 2 a.

5. **Distribution functions and their characteristic functions.** Henceforth we restrict ourselves to that sub-class of (V) which was denoted by (V_P), and we further suppose every function $F(x) < (V_P)$ to be so normalized that $F(-\infty) = 0$, $F(+\infty) = 1$. The class (V_P) then consists of the set of all d. f.'s, i. e. those real non-decreasing functions which are 0 for $x = -\infty$, 1 for $x = +\infty$. The class (T_P) is formed by those functions $f(t)$ which may be represented as the Fourier-Stieltjes transform of a function $F(x) < (V_P)$:

$$(35) \qquad f(t) = \int_{-\infty}^{\infty} e^{itx}\,dF(x).$$

The function $f(t)$ is called the *characteristic function* (c. f.) of $F(x)$ and has the following properties: it is

$$(36) \qquad \begin{cases} 1^\circ \text{ uniformly continuous,} \\ 2^\circ \text{ bounded: } |f(t)| \leq f(0) = \int_{-\infty}^{\infty} dF(x) = 1, \\ 3^\circ \text{ hermitian, i. e. } \overline{f(-t)} = f(t). \end{cases}$$

The c.f.'s naturally have all the properties of the (T)-functions but also show certain special features.

It is very important with regard to applications, to study the convergence of a sequence of c.f.'s.

A necessary and sufficient condition for the convergence of a sequence $\{F_n(x)\}$ *of d.f.'s to a d.f.* $F(x)$ *is, that the sequence of the corresponding c.f.'s* $\{f_n(t)\}$ *converges for all values of* t *to a function* $f(t)$, *continuous at* $t = 0$. *The limit* $f(t)$ *is then identical with the c.f. of* $F(x)$ *and* $\{f_n(t)\}$ *converges to* $f(t)$ *uniformly in every finite t-interval.*

Under somewhat less general conditions a similar theorem was first proved by LÉVY [1], pp. 195—197. In its present form the *convergence theorem* was proved contemporaneously by LÉVY [2], p. 49, and CRAMÉR [5], p. 29. See also CRAMÉR [7], p. 77, where a correction is made, and compare the convergence theorem in § 2 and Theorem 4 this chapter, § 6.

6. **A uniqueness theorem.** From the inversion formula in § 2 it follows that a d.f. $F(x)$ is uniquely determined by its c.f., i.e. if $f(t)$ is known for all t. We shall here consider the question: Do there exist two c.f.'s equal to each other in an interval about the zero point but not identically equal? GNEDENKO [1] has given an example of such an occurrence. Since there has been some obscurity as to this point, we give some further examples and theorems, starting with the following lemma.[1]

Lemma 2. *Let* $f(t)$ *be an even real bounded function which decreases steadily to zero as* $t \to \infty$ *and is convex downwards. Then, if* $f(0) = 1, f(t) < (T_{P_1})$.

Examples.

a. Suppose that $f(t)$ satisfies the conditions of Lemma 2, that $f(\pm 1) > 0$ and that $f'(\pm 1)$ exist. Form the even function (Fig. 3)

$$g(t) = \begin{cases} f(t) & \text{for } 0 \le |t| < 1 \\ f(1) + f'(1)(|t| - 1) & \text{for } 1 \le |t| < 1 - \dfrac{f(1)}{f'(1)}. \\ 0 & \text{for } \quad |t| \ge 1 - \dfrac{f(1)}{f'(1)} \end{cases}$$

From Lemma 2 it follows that both $f(t)$ and $g(t)$ belong to (T_{P_1}). Obviously the d.f.'s of $f(t)$ and $g(t)$ are not identical in spite of the fact that $f(t) = g(t)$ for $|t| \le 1$.

[1] TITCHMARSH [1], p. 170.

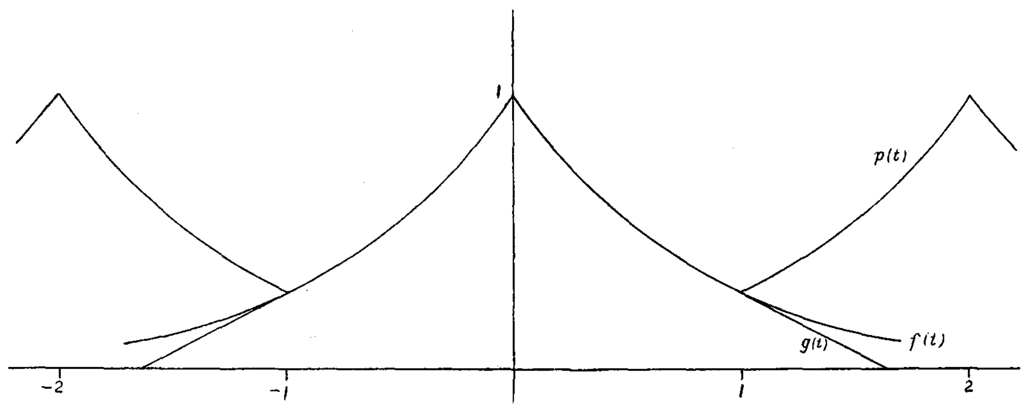

Fig. 3.

b. Let $f(t)$ be defined as in example a. Form $p(t) = f(t)$ for $|t| \leq 1$ and then continue $p(t)$ periodically with the period 2 (Fig. 3). Then $p(t) < (T_{P_3})$, for if we expand $p(t)$ in a Fourier series,

$$p(t) \sim \sum a_n e^{i n \pi t}, \quad a_n = \tfrac{1}{2} \int_{-1}^{1} e^{-i n \pi t} p(t)\, dt = \int_{0}^{1} \cos n \pi t \cdot f(t)\, dt,$$

it is easily seen as in the proof of Lemma 2 that $a_n \geq 0$. Furthermore, by a well-known theorem on Fourier series

$$p(t) = \sum a_n e^{i n \pi t}, \quad \sum a_n = p(0) = 1.$$

Hence $p(t)$ is the c.f. of a purely discontinuous d.f. with the jump $a_n \geq 0$ at $x = n\pi$. By the construction, $p(t) = f(t)$ for $|t| \leq 1$, but the d.f.'s are not identical.

In the examples given above we may for instance choose $f(t) = e^{-|t|}$, the c.f. of the Cauchy distribution $\frac{1}{2} + \frac{1}{\pi} \operatorname{arctg} x$.

Remarks.

1. Examples a and b show that a minimum extrapolation need not be unique. Let $h(t) = f(t)$ for $|t| \leq 1$. Then all the functions $f(t)$, $g(t)$ and $p(t)$ are minimum extrapolations of $h(t)$ with respect to $|t| \leq 1$.

2. In examples a and b the derivative at $t = 0$ does not exist. This is, however, by no means necessary.

Let us for a moment consider $\varphi(t) = e^{-\frac{t^2}{2}}$, the c.f. of the normal d.f. $\Phi(x)$. Does there exist a d.f. $\not\equiv \Phi(x)$ with the c.f. equal to $\varphi(t)$ in an interval about

$t = 0$? This is an important question with regard to the applications to the theory of probability. We will show that in this and many other cases the c. f. is uniquely determined by its values in an interval about $t = 0$. We base our argument upon the following lemma.[1]

Lemma 3. *A necessary and sufficient condition for the c.f. $f(t)$ of the d.f. $F(x)$ to have a finite derivative $f^{(2k)}(0)$, (k a positive integer), at $t = 0$, is that*

$$\alpha_{2k} = \int_{-\infty}^{\infty} x^{2k}\, dF(x) < \infty .$$

The sufficiency of the condition is immediately clear. In order to show the necessity we may without loss of generality suppose $k = 1$. Then by hypothesis $f''(0)$ exists and is finite. Now

$$-\tfrac{1}{2} f''(0) = \lim_{t \to 0} \frac{1 - \frac{f(t) + f(-t)}{2}}{t^2} .$$

From $f(t) = \int_{-\infty}^{\infty} e^{itx}\, dF(x)$ it follows:

$$\frac{f(t) + f(-t)}{2} = \int_{-\infty}^{\infty} \cos tx \, dF(x),$$

or

$$(37) \qquad -\tfrac{1}{2} f''(0) = \lim_{t \to 0} \int_{-\infty}^{\infty} \frac{1 - \cos tx}{t^2}\, dF(x).$$

From (37) we obtain, with regard to $\frac{1 - \cos tx}{t^2}\, dF(x) \geq 0$:

$$-\tfrac{1}{2} f''(0) \geq \int_{-a}^{a} \frac{x^2}{2}\, dF(x)$$

for every a. Hence $\int_{-\infty}^{\infty} x^2\, dF(x) < \infty$ and the lemma is proved.

We now enunciate the following theorem:

Theorem 3. *Let $F(x)$ and $G(x)$ be two d.f.'s and $f(t)$ and $g(t)$ the corresponding c.f.'s, such that*

1° $g(t) = f(t)$ *in an interval about* $t = 0$,

2° $\alpha_k = \int_{-\infty}^{\infty} x^k\, dF(x) < \infty$ *for* $k = 0, 1, 2, 3, \ldots$

[1] LÉVY [1], p. 174.

If the Stieltjes-Hamburger problem of moments with regard to $\{\alpha_k\}$ is determined, i. e. if the series $\sum_{k=1}^{\infty} \frac{1}{\alpha_{2k}^{\frac{1}{2k}}}$ *diverges, then* $F(x) \equiv G(x)$.

Proof. By the Stieltjes-Hamburger problem of moments we mean the determination of a non-decreasing function $\psi(x)$ belonging to a given sequence of numbers $\{c_k\}$ so that

$$\int_{-\infty}^{\infty} x^k \, d\psi(x) = c_k, \qquad (k = 0, 1, 2, 3, \ldots). \tag{38}$$

The problem is said to be *determined* if $\psi(x)$ is uniquely defined by (38), this, being the case, according to CARLEMAN [1], if and only if $\sum_{k=1}^{\infty} \frac{1}{c_{2k}^{1/2k}}$ diverges. (Here we naturally suppose that there exists a solution of the problem.)

By 2°, Theorem 3, $f^{(k)}(0)$ exists for every k and by 1° $g^{(k)}(0) = f^{(k)}(0)$ for every k. Thus by Lemma 3 every moment of $G(x)$ exists and from $f^{(k)}(0) = i_k \alpha_k$ it results that

$$\alpha_k = \int_{-\infty}^{\infty} x^k \, dF(x) = \int_{-\infty}^{\infty} x^k \, dG(x), \qquad (k = 0, 1, 2, 3, \ldots). \tag{39}$$

By hypothesis the problem of moments with regard to $\{\alpha_k\}$ is determined. Hence according to (39) $G(x) \equiv F(x)$ and the theorem is proved.

Theorem 3 especially holds if $f(t)$, $(t = \sigma + i\tau)$, is analytic and regular at $t = 0$, for if the Taylor series of $f(t)$ about $t = 0$ has a positive radius of convergence ϱ, then from Lemma 3 it is easily seen that $f(t)$ is analytic in $-\varrho < \tau < \varrho$, i. e. analytic and regular on the whole real axes. Hence, if $g(t) = f(t)$ in an interval about $t = 0$, $g(t)$ is also analytic and regular for all real t, and thus $g(t) \equiv f(t)$. Let us observe the important example $f(t) = e^{-\frac{t^2}{2}}$, the c. f. of the normal d. f. $\Phi(x)$. Here $f(t)$ is analytic and regular for $t = 0$, and thus the c. f. $g(t)$ of a d. f. $G(x)$ cannot be equal to $e^{-\frac{t^2}{2}}$ in an interval about $t = 0$ without $G(x) \equiv \Phi(x)$.

Let us finally consider the convergence theorem in § 5 from the point of view of this section. It is generally necessary and sufficient for the convergence of a sequence of d. f.'s $\{F_n(x)\}$ to a d. f. $F(x)$ that the corresponding c. f.'s $\{f_n(t)\}$ converge for *all* t to a function $f(t)$ continuous at $t = 0$, or if we only consider the convergence in an interval about $t = 0$ there may be several d. f.'s with the

c. f.'s equal to $f(t)$ in the interval in question. With regard to Theorem 3, an analysis of the proof of the convergence theorem mentioned above shows that it may be replaced, for instance, by the following:

Theorem 4. *A sufficient condition for the convergence of a sequence of d.f.'s* $\{F_n(x)\}$ *with the c.f.'s* $\{f_n(t)\}$ *to a d.f.* $F(x)$ *with the c.f.* $f(t)$ *is that* $\lim_{n\to\infty} f_n(t) = f(t)$ *for all* t *in the general case, or that* $\lim_{n\to\infty} f_n(t) = f(t)$ *in an interval about* $t=0$, *provided that the Stieltjes-Hamburger problem of moments with regard to* $F(x)$ *is determined.*

7. **On the approach towards 1 of the modulus of a characteristic function.** For later purposes it is of importance to consider $f(t)$ for large values of t, and further to investigate whether and when $|f(t_0)| = 1$ for a finite $t_0 \neq 0$.

We call a d.f. a *lattice distribution* if the following condition is satisfied: $F(x)$ is a purely discontinuous d. f. with the jumps situated only in a sequence of equidistant points. For instance, a purely discontinuous function $F(x)$ with $F(-0)=\frac{1}{4}$, $F(+0)=\frac{3}{4}$ and the jumps $\frac{1}{2^{n+2}}$ at $x=\pm n$ $(n=1, 2, 3, \ldots)$ is a lattice distribution. The most common example is the Bernoulli distribution having two jumps p and q $(p+q=1, p>0, q>0)$ at two points x_1 and x_2. The reason of the term »lattice» will become more clear in the multi-dimensional case. The lattice distributions are most frequently met with, besides the absolutely continuous distributions, in statistical applications.

Theorem 5.[1] *If and only if* $F(x)$ *is a lattice distribution, there exists a finite* $t_0 \neq 0$ *such that* $|f(t_0)| = 1$.

This condition is necessary, for if we suppose that $t_0 \neq 0$ and $|f(t_0)| = f(0) = 1$, then $f(t_0) e^{i\theta_0} = f(0)$ for some real θ_0, or

$$\int_{-\infty}^{\infty} (1 - e^{i(\theta_0 + t_0 x)})\, dF(x) = 0.$$

On taking real parts we obtain $\int_{-\infty}^{\infty} g(x)\, dF(x) = 0$ where $g(x) = 1 - \cos(\theta_0 + t_0 x)$. As $g(x) \geq 0$ and continuous, $g(x)$ must be 0 at every point where $dF(x) > 0$. But $g(x) = 0$ only for

$$x = x_0 + \nu \cdot \frac{2\pi}{t_0}, \quad \left(x_0 = -\frac{\theta_0}{t_0};\ \nu = 0, \pm 1, \pm 2, \pm \ldots\right), \tag{40}$$

[1] Wintner [2], p. 48.

and thus $F(x)$ must be a purely discontinuous function with the jumps

$$a_\nu \geq 0 \text{ for } x = x_0 + \nu \cdot \frac{2\pi}{t_0}, \qquad (\nu = 0, \pm 1, \pm 2, \pm \ldots), \tag{41}$$

and no other discontinuities. Thus $F(x)$ is a lattice distribution.

The condition is sufficient, for if $F(x)$ is a lattice distribution let it be defined by (41). Then

$$f(t) = \sum_\nu a_\nu e^{it\left(x_0 + \nu \cdot \frac{2\pi}{t_0}\right)};$$

hence $|f(t)|$ is periodic with the period t_0. Thus $|f(t_0)| = f(0) = 1$.[1]

The proof of the following theorem will be delayed until Chapter VII, § 1, where it is proved in the multi-dimensional case. By I we denote an arbitrary interval of the real axes and by $m_I(E)$ the measure of those t-points, belonging to I, for which a certain property E is satisfied.

Theorem 6. *Let $F(x)$ be a d.f. with the mean value zero, the c.f. $f(t)$ and the finite moments*

$$\alpha_2 = \int_{-\infty}^{\infty} x^2 \, dF(x); \quad \beta_3 = \int_{-\infty}^{\infty} |x|^3 \, dF(x).$$

For every ε, $(0 < \varepsilon \leq 1)$, and for every interval I of length $c_1 \cdot \frac{\alpha_2}{\beta_3}$, the inequality

$$m_I(|f(t)|^2 \geq 1 - \varepsilon) \leq c_2 \frac{\sqrt{\varepsilon}}{\sqrt{\alpha_2}}$$

holds, c_1 and c_2 being absolute constants.

We now proceed to the study of $|f(t)|$ for large values of t. Let us first recapitulate the results of § 2.

a. If $f(t) < (T_{P_1})$, then $\lim\limits_{t \to \pm\infty} |f(t)| = 0$.

b. If $f(t) < (T_{P_3})$, then $f(t) = \sum\limits_\nu a_\nu e^{i x_\nu t}$, a_ν being the jump of $F(x)$ at $x = x_\nu$, and $\sum\limits_\nu a_\nu = 1$. Then $f(t)$ is almost periodic, and since $f(0) = 1$, it follows that $\overline{\lim\limits_{t \to \pm\infty}} |f(t)| = 1$.

---

[1] It may happen that $|f(t)| = 1$ for every t. Then it is easily seen that $F(x) = E(x - a) = \begin{cases} 0 \text{ for } x < a \\ 1 \;\;\text{»}\;\; x > a \end{cases}$, a being a constant. We always exclude this case.

c. $f(t) < (T_{P_2})$. It is known that $|f(t)|$ is small in mean. There are[1], however, singular transforms $f(t)$ such that

$$\overline{\lim_{t \to \pm\infty}} |f(t)| > 0.$$

On the other hand there exist[2] singular transforms such that

$$\lim_{t \to \pm\infty} |f(t)| = 0.$$

We are especially interested in the question whether there exists a singular transform $f(t)$ with the property that

$$\overline{\lim_{t \to \pm\infty}} |f(t)| = 1.$$

I have not found any example of such a function in the literature until lately, when a paper by L. SCHWARTZ [1] became available in Sweden. Two years ago I found another example, using the following lemmata.

Lemma 4.[3] *Let $F_1(x), F_2(x), \ldots, F_n(x), \ldots$ be a sequence of purely discontinuous d.f.'s and suppose that the convolutions $\Psi_n(x) = F_1 * F_2 * F_3 * \cdots * F_n$, $(n = 1, 2, 3, \ldots)$, converge to a d.f. $\Psi(x)$ as $n \to \infty$. Then $\Psi(x)$ is purely discontinuous or purely singular or absolutely continuous.*

Lemma 5.[4] *Let $F_1(x), F_2(x), \ldots, F_n(x), \ldots$ be a sequence of purely discontinuous d.f.'s and suppose that the convolutions $\Psi_n(x) = F_1 * F_2 * F_3 * \cdots * F_n$, $(n = 1, 2, 3, \ldots)$, converge to a d.f. $\Psi(x)$. If d_n denotes the maximum jump of $F_n(x)$, the necessary and sufficient condition for $\Psi(x)$ to be continuous is that*

$$\lim_{n \to \infty} \prod_{\nu=1}^{n} d_\nu = 0.$$

Example. Let $\{\lambda_n\}$, $(n = 1, 2, 3, \ldots, \lambda_1 > 2)$, be a non-decreasing sequence of numbers such that $\lim_{n \to \infty} \lambda_n = \infty$, and let $F_n(x)$ be a purely discontinuous d. f. with the jump $1 - \frac{1}{\lambda_n}$ at $x = 0$ and the jump $\frac{1}{\lambda_n}$ at $x = 2^{-n}$, $(n = 1, 2, 3, \ldots)$. The corresponding c. f. $f_n(t)$ is obtained from

---

[1] CARLEMAN [2], p. 225, RIESZ [1], p. 312, JESSEN-WINTNER [1], p. 61.

[2] MENCHOFF [1], LITTLEWOOD [1]; many examples in the American Journal of Mathematics from 1935 and onwards.

[3] JESSEN-WINTNER [1], p. 86.

[4] LÉVY [3].

$$f_n(t) = 1 - \frac{1}{\lambda_n} + \frac{1}{\lambda_n} e^{it \cdot 2^{-n}}.$$

Now we suppose that

$$\prod_1^\infty \left(1 - \frac{1}{\lambda_n}\right) = 0. \tag{42}$$

By the convolution theorem in § 2 the d. f. $\Psi_n(x) = F_1 * F_2 * \cdots * F_n$ has the c. f.

$$\psi_n(t) = \prod_{\nu=1}^{n} \left(1 - \frac{1}{\lambda_\nu} + \frac{1}{\lambda_\nu} e^{it \cdot 2^{-\nu}}\right).$$

It is easily seen that $\psi_n(t)$ converges for all t to a continuous function

$$\psi(t) = \prod_{\nu=1}^{\infty} \left(1 - \frac{1}{\lambda_\nu} + \frac{1}{\lambda_\nu} e^{it \cdot 2^{-\nu}}\right) \text{ as } n \to \infty. \tag{43}$$

By the convergence theorem in § 5 $\lim\limits_{n \to \infty} \Psi_n(x) = \Psi(x)$ exists and is a d. f. By Lemma 5 and (42) $\Psi(x)$ is continuous, and by Lemma 4 it is either purely singular or absolutely continuous. Putting $t = 2\pi \cdot 2^m$ in (43), (m a positive integer), it is easily found that

$$\lim_{m \to \infty} |\psi(2\pi \cdot 2^m)| = 1. \tag{44}$$

But (44) shows that $\Psi(x)$ cannot be absolutely continuous. Hence it is singular and its c. f. has the required property.

d. Let us finally consider the general case: $f(t) < (T_P)$. By § 1 we may write

$$F(x) = b_1 F_1(x) + b_2 F_2(x) + b_3 F_3(x), \quad \left(b_i \geq 0, \sum_{i=1}^{3} b_i = 1, i = 1, 2, 3\right); \tag{45}$$

here the functions $F_i(x)$ are d. f.'s, $F_1(x)$ being absolutely continuous etc. Correspondingly

$$f(t) = b_1 f_1(t) + b_2 f_2(t) + b_3 f_3(t). \tag{46}$$

If $b_1 > 0$, i. e. if $F(x)$ has an absolutely continuous component, we obtain from (46) and case a: $\overline{\lim\limits_{t \to \pm\infty}} |f(t)| < 1$.

If $b_1 = 0$, i. e. if $f(t) < (T_{P_2} + T_{P_3})$, cases b and c show that sometimes $\overline{\lim\limits_{t \to \pm\infty}} |f(t)| = 1$, sometimes $\overline{\lim\limits_{t \to \pm\infty}} |f(t)| < 1$.

If we sum up the results of cases a—d, combined with Theorem 5, we obtain:

Let $F(x)$ *be a d.f. with the c.f.* $f(t)$.

If $F(x)$ *has an absolutely continuous component, then*

$$\overline{\lim_{t\to\pm\infty}}\,|f(t)| < 1.$$

If $F(x)$ *is purely discontinuous, then*

$$\overline{\lim_{t\to\pm\infty}}\,|f(t)| = 1.$$

If $F(x)$ *is purely singular or if* $F(x) < (V_{P_2} + V_{P_3})$ *either*

$$\overline{\lim_{t\to\pm\infty}}\,|f(t)| = 1$$

or

$$\overline{\lim_{t\to\pm\infty}}\,|f(t)| < 1$$

may occur.

If $\overline{\lim\limits_{t\to\pm\infty}}\,|f(t)| < 1$, *then there exists a constant* $c > 0$ *such that* $|f(t)| < e^{-c}$ *for* $|t| \geq 1$.

Chapter II.

Estimation of the Difference Between Two Distribution Functions by the Behaviour of Their Characteristic Functions in an Interval About the Zero Point.

1. **On** $\int\limits_{-\infty}^{\infty} |F(x) - G(x)|\,dx$. We have earlier found that two c.f.'s may be equal to each other in an interval about the zero point without the corresponding d.f.'s being necessarily identical. It will, however, be shown that they are approximately equal in the mean. As a measure of the difference we may consider

$$\varrho(F, G) = \int\limits_{-\infty}^{\infty} |F(x) - G(x)|\,dx.$$

Theorem 1. *If* $F(x)$ *and* $G(x)$ *are two d.f.'s,* $f(t)$ *and* $g(t)$ *being the corresponding c.f.'s such that*

$$f(t) = g(t) \text{ on } |t| \leq L,$$

then

$$\varrho(F, G) = \int\limits_{-\infty}^{\infty} |F(x) - G(x)|\,dx \leq \frac{\pi}{L}. \tag{1}$$

Proof. From $f(t) - g(t) = \int\limits_{-\infty}^{\infty} e^{itx}\, d(F(x) - G(x))$ we obtain by partial integration:

$$\frac{f(t) - g(t)}{-it} = \int\limits_{-\infty}^{\infty} e^{itx} (F(x) - G(x))\, dx. \tag{2}$$

From (2) and the definition of the metric $T(f)$, (Chapter I, (7)), we obtain:

$$\varrho(F, G) = T\left\{\frac{f(t) - g(t)}{-it}\right\}. \tag{3}$$

But since $f(t) = g(t)$ in $|t| \leq L$ we have $\frac{f(t) - g(t)}{-it} = 0$ for $|t| \leq L$. Hence we may write:

$$\frac{f(t) - g(t)}{-it} = (f(t) - g(t)) \cdot a(t), \tag{4}$$

where $a(t) = -\frac{1}{it}$ for $|t| \geq L$ and for the rest arbitrary. We now choose

$$a(t) \equiv \text{the minimum extrapolation of } -\frac{1}{it} \tag{5}$$

with respect to $|t| \geq L$. From Chapter I, Theorem 1, it follows:

$$T(a(t)) \leq \frac{\pi}{2} \cdot \frac{1}{L}. \tag{6}$$

Further

$$T(f(t) - g(t)) = \int\limits_{-\infty}^{\infty} |d(F(x) - G(x))| \leq 2. \tag{7}$$

From (3), (4), (6) and (7) we now obtain:

$$\varrho(F, G) = T\{(f(t) - g(t))\, a(t)\} \leq T(f(t) - g(t)) \cdot T(a(t)) \leq 2 \cdot \frac{\pi}{2} \cdot \frac{1}{L},$$

and the theorem is proved. This method has been used by Beurling (loc. cit. p. 13) in similar cases.

Remark.

Even if $f(t) \not\equiv g(t)$ in an interval about $t = 0$, it is possible to make an estimation of $\varrho(F, G)$, provided that further conditions are imposed. I confine myself to this indication.

2. **On** $|F(x) - G(x)|$. We now proceed to the proof of a theorem that is fundamental with regard to its applications.[1]

[1] Esseen [1], p. 3.

Theorem 2 a. *Let A, T and ε be arbitrary positive constants, $F(x)$ a non-decreasing function, $G(x)$ a real function of bounded variation on the whole real axes, $f(t)$ and $g(t)$ the corresponding Fourier-Stieltjes transforms such that*

$1°$ $F(-\infty) = G(-\infty) = 0, \quad F(+\infty) = G(+\infty),$

$2°$ *$G'(x)$ exists everywhere and $|G'(x)| \leq A$, .*

$$(8) \qquad 3° \quad \int_{-T}^{T} \left| \frac{f(t) - g(t)}{t} \right| dt = \varepsilon.$$

To every number $k > 1$ there corresponds a finite positive number $c(k)$, only depending on k, such that

$$(9) \qquad |F(x) - G(x)| \leq k \cdot \frac{\varepsilon}{2\pi} + c(k) \cdot \frac{A}{T}.$$

We also need a theorem[1], analogous to Theorem 2 a, where, however, $G(x)$ is not supposed to be continuous.

Theorem 2 b. *Let A, T and ε be arbitrary positive constants, let $F(x)$ be a non-decreasing, purely discontinuous function and $G(x)$ a real function of bounded variation on the whole real axes, $f(t)$ and $g(t)$ the corresponding Fourier-Stieltjes transforms such that*

$1°$ $F(-\infty) = G(-\infty) = 0, \quad F(+\infty) = G(+\infty),$

$2°$ *if $G(x)$ is discontinuous at $x = x_\nu$, $(x_\nu < x_{\nu+1},\ \nu = 0, \pm 1, \pm 2, \pm \cdots)$, there exists a constant $L > 0$ such that* Min. $(x_{\nu+1} - x_\nu) \geq L$,

$3°$ *$|G'(x)| \leq A$ everywhere except when $x = x_\nu$, $(\nu = 0, \pm 1, \pm 2, \pm \cdots)$,*

$4°$ *$F(x)$ may be discontinuous only at $x = x_\nu$, $(\nu = 0, \pm 1, \pm 2, \pm \cdots)$,*

$$(10) \qquad 5° \quad \int_{-T}^{T} \left| \frac{f(t) - g(t)}{t} \right| dt = \varepsilon.$$

Then to every number $k > 1$ there correspond two finite positive constants $c_1(k)$ and $c_2(k)$, only depending on k, such that

$$(11) \qquad |F(x) - G(x)| \leq k \cdot \frac{\varepsilon}{2\pi} + c_1(k) \cdot \frac{A}{T},$$

provided that $T \cdot L \geq c_2(k)$.

---

[1] Esseen [2], p. 7, without proof.

Proof of Theorem 2 a. From $f(t) - g(t) = \int\limits_{-\infty}^{\infty} e^{itx}\, d\,(F(x) - G(x))$ and integrating by parts we obtain

$$\frac{f(t) - g(t)}{-it} = \int\limits_{-\infty}^{\infty} e^{itx}\,(F(x) - G(x))\, dx. \tag{12}$$

Thus $\frac{f(t) - g(t)}{-it}$ is the Fourier transform of $F(x) - G(x)$.

In order to understand the theorem better let us first suppose

$$\int\limits_{-\infty}^{\infty} \left| \frac{f(t) - g(t)}{t} \right| dt = \varepsilon < \infty$$

and let $T = \infty$ in (8). By means of the Fourier inversion formula we obtain from (12):

$$F(x) - G(x) = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} e^{-ixt}\, \frac{f(t) - g(t)}{-it}\, dt.$$

Hence

$$|F(x) - G(x)| \leq \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} \left| \frac{f(t) - g(t)}{t} \right| dt = \frac{\varepsilon}{2\pi}.$$

Thus $c(k) \cdot \frac{A}{T}$ in (9) can be interpreted as a remainder term, corresponding to a finite T.

It is sufficient to prove the theorem for $A = T = 1$, for if we put

$$F_1(x) = \frac{T}{A}\, F\left(\frac{x}{T}\right), \quad G_1(x) = \frac{T}{A}\, G\left(\frac{x}{T}\right),$$

$f_1(t)$ and $g_1(t)$ being the corresponding transforms, then $|G_1'(x)| \leq 1$ and

$$\int\limits_{-1}^{1} \left| \frac{f_1(t) - g_1(t)}{t} \right| dt = \frac{T \cdot \varepsilon}{A}.$$

Now suppose that the theorem holds for $A = T = 1$. Then

$$|F_1(x) - G_1(x)| \leq k \cdot \frac{\varepsilon}{2\pi} \cdot \frac{T}{A} + c(k),$$

or

$$|F(x) - G(x)| \leq k \cdot \frac{\varepsilon}{2\pi} + c(k) \cdot \frac{A}{T},$$

the desired inequality. Thus in the following we take $A = T = 1$.

In the proof, which is based on a convolution method, we use two auxiliary functions $H(x)$ and $h(t)$ with the following properties:

$$(13) \quad \begin{cases} 1^\circ \quad H(x) \text{ and } h(t) \text{ are real even non-negative functions;} \\ 2^\circ \quad \int_{-\infty}^{\infty} H(x)\,dx = 1; \quad b = \int_{-\infty}^{\infty} |x|\, H(x)\,dx < \infty; \\ 3^\circ \quad h(t) = \int_{-\infty}^{\infty} e^{itx} H(x)\,dx; \quad h(0) = 1; \quad h(t) = 0 \text{ for } |t| \geq 1; \end{cases}$$

$$0 \leq h(t) \leq 1 \text{ for } |t| \leq 1.$$

To obtain an example of such a function we may proceed as follows. Let

$$k(t) = \begin{cases} 1 - |t| & \text{for } |t| \leq 1 \\ 0 & \text{for } |t| \geq 1 \end{cases}.$$

Hence the Fourier transform

$$K(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ixt} k(t)\,dt = \frac{1}{2\pi} \left(\frac{\sin \frac{x}{2}}{\frac{x}{2}} \right)^2.$$

By means of the convolution theorem of Fourier integrals

$$H(x) = \frac{3}{8\pi} \left(\frac{\sin \frac{x}{4}}{\frac{x}{4}} \right)^4$$

and

$$h(t) = \frac{\int_{-\infty}^{\infty} k(2t - s)\, k(s)\,ds}{\int_{-\infty}^{\infty} (k(s))^2\,ds}$$

are Fourier transforms; it is easily seen that they have the required properties.

In this connection we quote a theorem by Ingham [1].

Lemma 1. *If $\varepsilon(x)$ is an assigned positive function tending steadily to zero when $x \to \infty$, there exists a non-null function $h(t)$, equal to zero outside an assigned interval $(-l, l)$, and having a Fourier transform $H(x)$ satisfying*

$$H(x) = O(e^{-|x|\varepsilon(|x|)}),\ (x \to \pm\infty),$$

if and only if

$$\int_1^\infty \frac{\varepsilon(x)}{x}\,dx$$

converges.

The sufficiency is easily proved by considering a function $H(x)$ of the type:

$$H(x) = \prod_1^\infty \frac{\sin \varrho_n x}{\varrho_n x},$$

the quantities ϱ_n being suitably chosen.

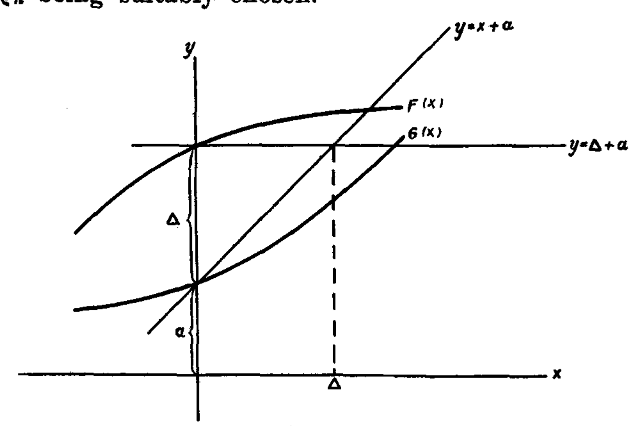

Fig. 4.

After these preliminaries we proceed to the proof of Theorem 2 a. Put

(14) $$\varDelta = \operatorname*{Max}_{-\infty<x<\infty} |F(x) - G(x)|.$$

Without loss of generality we may suppose that $\varDelta = |F(x) - G(x)|$ for $x = 0$, since a translation x_0 of x is equivalent to a multiplication of the transform by e^{itx_0} of modulus 1. Further we may suppose $F(0) > G(0)$.

Since $F(x)$ is non-decreasing and $|G'(x)| \leq 1$ it is easily seen from Fig. 4 that

(15) $$F(x) - G(x) \geq \varDelta - x \text{ for } 0 \leq x \leq \varDelta.$$

Now consider the integral

$$V(x) = \int_{-\infty}^{\infty} H(x-y)\,[F(y) - G(y)]\,dy,$$

$H(x)$ and $h(t)$ being defined by (13). Formally we obtain from (12), (13: 3°) and the Parseval formula:

$$(16)\qquad \int_{-\infty}^{\infty} H(x-y)\,[F(y)-G(y)]\,dy = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ixt}\,\frac{f(t)-g(t)}{-it}\,h(t)\,dt.$$

The validity of (16) may be proved in the following way. It is immediately clear that $(f(t)-g(t))\cdot h(t)$ is the Fourier-Stieltjes transform of $V(x)$. By the inversion formula (5), Chap. I, we have

$$V(x)-V(x_1)=\frac{1}{2\pi}\int_{-1}^{1}\frac{f(t)-g(t)}{-it}\,h(t)\,(e^{-itx}-e^{-itx_1})\,dt.$$

In view of (8) and the Riemann-Lebesgue theorem it is readily seen, that $V(x_1)\to 0$ and that $\frac{1}{2\pi}\int_{-1}^{1}\frac{f(t)-g(t)}{-it}\,h(t)\,e^{-itx_1}\,dt\to 0$ when $x_1\to\infty$. Hence (16), the central formula of the proof.

From (15) and (16) we have

$$\int_{0}^{\Delta}(\Delta-y)\,H(x-y)\,dy-\int_{\Delta}^{\infty}H(x-y)\,|F(y)-G(y)|\,dy-$$
$$-\int_{-\infty}^{0}H(x-y)\,|F(y)-G(y)|\,dy\le\frac{1}{2\pi}\int_{-1}^{1}\left|\frac{f(t)-g(t)}{t}\,h(t)\right|dt\le\frac{\varepsilon}{2\pi}\cdot$$

But $|F(y)-G(y)|\le\Delta$. Hence

$$\int_{0}^{\Delta}(\Delta-y)\,H(x-y)\,dy-\Delta\int_{\Delta}^{\infty}H(x-y)\,dy-\Delta\int_{-\infty}^{0}H(x-y)\,dy\le\frac{\varepsilon}{2\pi}\cdot$$

By means of (13: 2°) this is easily transformed to

$$\int_{-x}^{\Delta-x}(2\Delta-x-y)\,H(y)\,dy-\Delta\le\frac{\varepsilon}{2\pi}$$

or

$$(17)\qquad \int_{-x}^{\Delta-x}(2\Delta-x)\,H(y)\,dy-\Delta\le\frac{\varepsilon}{2\pi}+b.$$

In (17) we put $x=m\cdot\Delta$, $(0<m<1)$, and obtain

$$(18) \qquad \varDelta\left\{(2-m)\int\limits_{-m\varDelta}^{(1-m)\varDelta} H(y)\,dy - 1\right\} \leq \frac{\varepsilon}{2\pi} + b.$$

Given an arbitrary number $k > 1$ we can always choose $m(k)$ sufficiently small and $\alpha(k)$ sufficiently large so that

$$(19) \qquad (2-m(k))\int\limits_{-m(k)\alpha(k)}^{(1-m(k))\alpha(k)} H(y)\,dy - 1 = 1/k.$$

Now two cases may occur:

1. $\varDelta \leq \alpha(k)$,
2. $\varDelta > \alpha(k)$.

Hence from (18) and (19):

$$\varDelta\cdot\frac{1}{k} \leq \frac{\varepsilon}{2\pi} + b \text{ or } \varDelta \leq k\cdot\frac{\varepsilon}{2\pi} + k\cdot b.$$

Thus

$$\varDelta \leq \mathrm{Max}\left(k\frac{\varepsilon}{2\pi} + kb;\ \alpha(k)\right) \leq k\cdot\frac{\varepsilon}{2\pi} + kb + \alpha(k) = k\cdot\frac{\varepsilon}{2\pi} + c(k),$$

$c(k)$ being a number only depending on k. Hence the theorem is proved. By the construction $c(k) \to \infty$ as $k \to 1$.

Proof of Theorem 2 b. The method of proof is similar to that of Theorem 2 a. As before we may take $A = T = 1$ and, putting

$$(20) \qquad \varDelta = \mathop{\mathrm{Max}}_{-\infty<x<\infty} |F(x) - G(x)|,$$

we may suppose that $|F(0) - G(0)| = \varDelta$. Several cases may occur in the behaviour of $F(x)$ and $G(x)$ about $x = 0$. We restrict ourselves to that represented in Fig. 5; the others are treated in the same way.

From Fig. 5 it is seen that

$$(21) \qquad F(x) - G(x) \geq \varDelta - x \text{ for } 0 \leq x \leq \delta$$

where $\delta = \mathrm{Min}(\varDelta, L/2)$. As before

$$(22) \qquad \int\limits_{-\infty}^{\infty} H(x-y)[F(y) - G(y)]\,dy = \frac{1}{2\pi}\int\limits_{-1}^{1} e^{-ixt}\frac{f(t)-g(t)}{-it}h(t)\,dt,$$

and hence

$$(23) \qquad \int_{-x}^{\delta-x} (2\varDelta - x) H(y)\, dy - \varDelta \leq \frac{\varepsilon}{2\pi} + b.$$

Putting $x = m \cdot \delta$, $(0 < m < 1)$, in (23) we obtain

$$(24) \qquad \varDelta \left\{ (2-m) \int_{-m\delta}^{(1-m)\delta} H(y)\, dy - 1 \right\} \leq \frac{\varepsilon}{2\pi} + b.$$

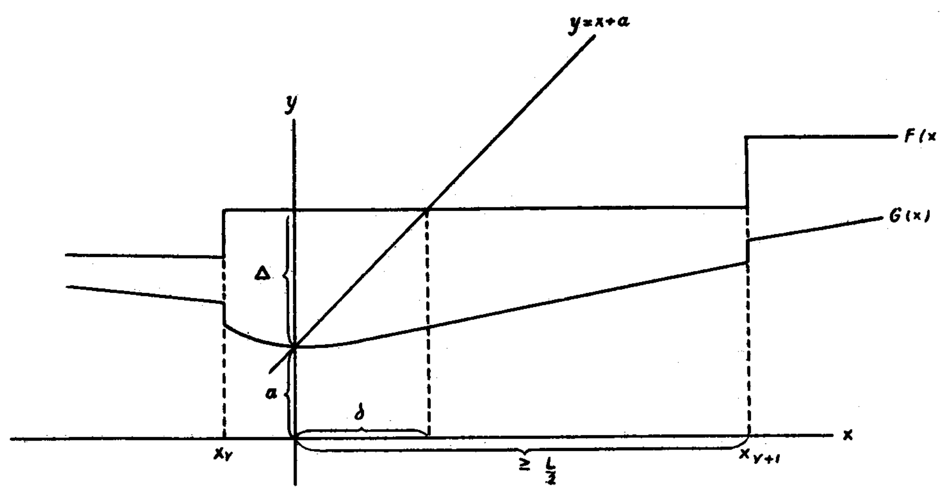

Fig. 5.

Given an arbitrary number $k > 1$ we choose $m(k)$ and $\alpha(k)$ in such a manner that (19) holds. Then two cases may occur.

1. $\varDelta \leq \alpha(k)$.
2. $\varDelta > \alpha(k)$.

If now $\frac{L}{2} > \alpha(k)$, then $\delta = \text{Min.}\, (\varDelta, L/2) > \alpha(k)$, and hence from (19) and (24):

$$\varDelta \leq \text{Max} \left\{ k \frac{\varepsilon}{2\pi} + kb;\ \alpha(k) \right\} \text{ or } \varDelta \leq k \cdot \frac{\varepsilon}{2\pi} + c_1(k),$$

provided that $L > 2\alpha(k) = c_2(k)$, $c_1(k)$ and $c_2(k)$ only depending on k. This proves the theorem.

Chapter III.

Random Variables. Improvement of the Liapounoff Remainder Term.

1. **Random variables.** By a *random variable* X we understand, popularly speaking, a quantity, which may assume certain real values with certain probabilities. The foundations and definitions of the theory of probability have always been subject to discussion and different opinions. In recent years it has been attempted to give the theory of probability a more rigorous structure by connection with the general theory of sets and axiomatically stated definitions. (KOLMOGOROFF [1].) I shall give a brief account of some of the conceptions and definitions that have been used; for further information the reader is referred to the works of KOLMOGOROFF and CRAMÉR [5].

Consider the k-dimensional euclidean space R_k with the variable point $x=(x_1, x_2, \ldots, x_k)$. By S we denote an arbitrary Borel set in R_k. (Only such sets are considered here.) A set function $P(S)$ is called a *probability function* if the following conditions are satisfied.

1. $P(S)$ is defined for every Borel set S; $P(S) \geq 0$.
2. $P(R_k) = 1$.
3. $P(S)$ is completely additive, i. e.

$$P(S_1 + S_2 + S_3 + \cdots) = P(S_1) + P(S_2) + P(S_3) + \cdots,$$

where $S_1, S_2, S_3, \ldots$ are Borel sets, no two of which have a common point.

The probability function $P(S)$ defines the *probability distribution* of a *random variable* X (r. v.) in R_k, $P(S)$ denoting the probability that $X < S$.

By $S_{\xi_1, \xi_2, \ldots, \xi_k}$ we denote the set $x_i \leq \xi_i$, $(i = 1, 2, \ldots, k)$. Then the *distribution function* $F(\xi_1, \xi_2, \ldots, \xi_k)$, corresponding to the r. v. X, is defined by

$$F(\xi_1, \xi_2, \ldots, \xi_k) = P(S_{\xi_1, \xi_2, \ldots, \xi_k}).$$

The d. f. $F(\xi_1, \xi_2, \ldots, \xi_k)$ is a point function, uniquely defined by $P(S)$. Conversely, by a well-known theorem of Lebesgue, $F(\xi_1, \xi_2, \ldots, \xi_k)$ determines $P(S)$ uniquely. Every d. f. $F(\xi_1, \xi_2, \ldots, \xi_k)$ has the following properties:

1. In each variable ξ_i, F is a non-decreasing function, continuous to the right, and $\lim\limits_{\xi_i \to -\infty} F = 0$.
2. As all variables $\xi_i \to +\infty$, F tends to the limit 1.

If $X=(X_1, X_2, \ldots, X_k)$ is a r. v. in R_k and $Y=(Y_1, Y_2, \ldots, Y_m)=f(X)$ is a B-measurable vector function, finite and uniquely defined for all points X of R_k, then $f(X)$ is a r. v. in R_m.

Let X_1 and X_2 be two r. v.'s in R_{k_1} and R_{k_2} with the probability functions $P_1(S_1)$ and $P_2(S_2)$ respectively and consider the combined variable $X=(X_1, X_2)$ in the product space $R_{k_1}R_{k_2}$ with the probability function $P(S)$. If S denotes the set formed by X as $X_1 < S_1$ and $X_2 < S_2$, then X_1 and X_2 are *independent* if

$$P(S)=P_1(S_1)\,P_2(S_2).$$

In the same way the mutual independence of n r. v.'s is defined.

In this and the following two chapters we only consider probability distributions in one dimension.

2. **Probability distributions in one dimension.** Consider a one-dimensional r. v. X with the d. f. $F(x)$. In a discontinuity point we put

$$F(x)=\tfrac{1}{2}(F(x+0)+F(x-0)),$$

thus slightly modifying the definition in § 1. The difference is unimportant, but now $F(x)$ belongs to the class (V_P) in Chapter I.

By α_k and β_k we always denote the moment and the absolute moment respectively of order k:

$$\alpha_k=\int_{-\infty}^{\infty} x^k\,dF(x); \quad \beta_k=\int_{-\infty}^{\infty} |x|^k\,dF(x). \tag{1}$$

The number k is generally a positive integer but sometimes we also consider absolute moments where this need not be the case. The following important inequalities are well known[1]:

$$\beta_1 \le \beta_2^{\frac{1}{2}} \le \beta_3^{\frac{1}{3}} \le \beta_4^{\frac{1}{4}} \le \cdots. \tag{2}$$

Two moments play an especially important rôle in the statistical applications, *the mean value* or the mathematical expection $m(X)$ of X and the *dispersion* $\sigma(X)$:

$$m(X)=\int_{-\infty}^{\infty} x\,dF(x); \quad \sigma^2(X)=\int_{-\infty}^{\infty}(x-m)^2\,dF(x)=\alpha_2-\alpha_1^2. \tag{3}$$

[1] E. g. HARDY-LITTLEWOOD-PÓLYA [1], p. 157.

Let k be a positive integer and β_k finite and consider the c. f.

$$f(t) = \int_{-\infty}^{\infty} e^{itx}\, dF(x). \tag{4}$$

The derivatives $f^{(\nu)}(t)$ obviously exist and are finite for $\nu = 1, 2, \ldots, k$. For small values of t we obtain the expansion

$$f(t) = 1 + \sum_{\nu=1}^{k} \frac{\alpha_\nu}{\nu!}(i\,t)^\nu + o(|t|^k). \tag{5}$$

If $\beta_{k+\delta}$, $(0 < \delta < 1)$, is finite, (5) may be replaced by

$$f(t) = 1 + \sum_{\nu=1}^{k} \frac{\alpha_\nu}{\nu!}(i\,t)^\nu + \theta \cdot \beta_{k+\delta} |t|^{k+\delta}, \tag{5'}$$

where the modulus of θ is bounded by a finite quantity, only depending on k and δ. For small values of t the following expansion holds:

$$\log f(t) = \sum_{\nu=1}^{k} \frac{\gamma_\nu}{\nu!}(i\,t)^\nu + o(|t|^k); \tag{6}$$

the coefficients γ_ν are called the *semi-invariants* of the distribution.

Let us now consider two independent r. v.'s X_1 and X_2. The sum $X = X_1 + X_2$ is also a r. v. The connection between the d. f.'s of X_1, X_2 and X is expressed by the *addition theorem*[1]:

Let X_1 and X_2 be two independent r. v.'s with the d. f.'s $F_1(x)$ and $F_2(x)$ respectively and the corresponding c. f.'s $f_1(t)$ and $f_2(t)$. The d. f. $F(x)$ of the r. v. $X = X_1 + X_2$ is obtained from

$$F(x) = F_1 * F_2 = \int_{-\infty}^{\infty} F_1(x-y)\, dF_2(y) = \int_{-\infty}^{\infty} F_2(x-y)\, dF_1(x).$$

The c. f. $f(t)$ of $F(t)$ is expressed by

$$f(t) = f_1(t) \cdot f_2(t).$$

Concerning the properties of $F(x)$, see Chapter I, § 1. The generalization of the addition theorem to a sum of any number of independent variables is immediate. We also observe the following relations: If X_1, X_2, $\ldots$, X_n are a

[1] Lévy [1], p. 186.

sequence of independent r. v.'s and $X = a_1 X_1 + a_2 X_2 + \cdots + a_n X_n$, where the coefficients a_i are constants, then

$$(7) \qquad \begin{cases} m(X) = a_1 m(X_1) + a_2 m(X_2) + \cdots + a_n m(X_n); \\ \sigma^2(x) = a_1^2 \sigma^2(X_1) + a_2^2 \sigma^2(X_2) + \cdots + a_n^2 \sigma^2(X_n). \end{cases}$$

3. **Improvement of the Liapounoff remainder term.** This section is devoted to an investigation of the Liapounoff remainder term (see Introduction, (1) and (2)). Our aim is the proof of the inequality (6) in the Introduction, which relation we shall obtain from a somewhat more general theorem.

Consider a sequence of independent r. v.'s $X_1, X_2, \ldots, X_n$ such that each variable X_ν has the d. f. $F_\nu(x)$, the c. f. $f_\nu(t)$, the mean value zero, the dispersion σ_ν, the moments $\alpha_{k\nu}$, the absolute moments $\beta_{k\nu}$ and the semi-invariants $\gamma_{k\nu}$, ($\nu = 1, 2, 3, \ldots, n$; $k = 3, 4, \ldots$ or sometimes, and then we only consider absolute moments, $2 < k \leq 3$.) By $\overline{F_n(x)}$ we denote the d. f. of the variable

$$Z_n = \frac{X_1 + X_2 + \cdots + X_n}{s_n},$$

where $s_n^2 = \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2$, and by $\overline{f_n(t)}$ the corresponding c. f. These notations are used throughout the chapter. It follows from (7) that Z_n has the mean value zero and the dispersion 1.

Our problem is to study the difference

$$\overline{F_n(x)} - \Phi(x),$$

$\Phi(x)$ being the normal d. f.:

$$(8) \qquad \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{y^2}{2}} dy$$

with the c. f. $e^{-\frac{t^2}{2}}$.

Let us first observe the following relations, consequences of the addition theorem:

$$(9) \qquad \begin{cases} \overline{F_n(x)} = F_1(s_n x) * F_2(s_n x) * \cdots * F_n(s_n x); \\ \overline{f_n(t)} = \prod_{\nu=1}^{n} f_\nu\left(\frac{t}{s_n}\right). \end{cases}$$

According to Cramér [5], p. 70, we introduce the quantities:

$$(10)\qquad \begin{cases} B_{kn} = \frac{1}{n}(\beta_{k1} + \beta_{k2} + \cdots + \beta_{kn}), & \Gamma_{kn} = \frac{1}{n}(\gamma_{k1} + \gamma_{k2} + \cdots + \gamma_{kn}), \\ \varrho_{kn} = \frac{B_{kn}}{B_{2n}^{k/2}}, \quad \lambda_{kn} = \frac{\Gamma_{kn}}{\Gamma_{2n}^{k/2}}, \end{cases}$$

$$(11)\qquad T_{kn} = \frac{\sqrt{n}}{4\,\varrho_{kn}^{3/k}}.$$

It is readily observed that

$$(12)\qquad 1 \leq \varrho_{2n}^{1/2} \leq \varrho_{3n}^{1/3} \leq \ldots$$

If all the d. f.'s $F_\nu(x)$ are equal with the dispersion σ, the absolute moments β_k and the semi-invariants γ_k, we observe that ϱ_{kn} and λ_{kn} are independent of n:

$$(13)\qquad \varrho_{kn} = \frac{\beta_k}{\sigma^k}, \quad \lambda_{kn} = \frac{\gamma_k}{\sigma^k}.$$

After these preliminaries we may state the following theorem, containing the desired improvement of the Liapounoff remainder term.

Theorem 1. *Let $X_1, X_2, \ldots, X_n$ be a sequence of independent r. v.'s such that each variable X_ν has the mean value zero and the finite absolute moment $\beta_{k\nu}$, $(\nu = 1, 2, 3, \ldots, n)$, of given order k, $(2 < k \leq 3)$. Then*

$$(14)\qquad |\overline{F_n(x)} - \Phi(x)| \leq c(k)\left(\frac{\varrho_{kn}}{n^{\frac{k-2}{2}}} + \frac{\varrho_{kn}^{\frac{1}{k-2}}}{n^{\frac{1}{2}}}\right),$$

where $c(k)$ is a finite positive constant only depending on k, and ϱ_{kn} is defined by (10).

Remarks.

1. If $k = 3$ we obtain by (14):

$$(15)\qquad |\overline{F_n(x)} - \Phi(x)| \leq C \cdot \frac{\varrho_{3n}}{\sqrt{n}},$$

C being an absolute constant. It is possible to show that C may be chosen $= 7.5$. The calculations are simple but rather laborious, and I omit them here. If all the d. f.'s are equal, (15) and (13) give:

$$(15')\qquad |\overline{F_n(x)} - \Phi(x)| \leq C \cdot \frac{\beta_3}{\sigma^3 \sqrt{n}}.$$

2. It is interesting to observe that the remainder term $O\left(\frac{1}{\sqrt{n}}\right)$ in (15) generally cannot be improved even if moments of all orders are finite. Consider the case, where every d. f. $F_\nu(x)$ is identical with a d. f. having the jumps $\frac{1}{2}$ at $x = \pm 1$. It is readily observed that $\overline{F_n(x)}$ has jumps in the vicinity of $x = 0$, asymptotically equal to $\sqrt{\frac{2}{\pi}} \cdot \frac{1}{\sqrt{n}}$, $(n \to \infty)$. We shall return to this question in the next chapter.

The proof of Theorem 1 is based on Theorem 2 a, Chap. II. Before we proceed to the proof, however, we shall record some lemmata concerning the expansion of $\overline{f_n(t)}$. These lemmata are of an elementary nature, being easily obtained by expanding each $f_\nu(t)$ in a Taylor series about $t = 0$. Concerning Lemma 2 a the reader is referred to Cramér [5], pp. 71 and 74; the others are proved in a similar way.

Throughout this chapter we denote by c_k an unspecified finite positive constant only depending on k.

Lemma 1. *If k is an assigned number,* $(2 < k \leq 3)$, $\beta_{k\nu} < \infty$ *for* $\nu = 1, 2, 3, \ldots, n$, ϱ_{kn} *defined by* (10) *and*

$$|t| \leq \frac{\sqrt{n}}{(24\, \varrho_{kn})^{\frac{1}{k-2}}},$$

then

$$\left|\overline{f_n(t)} - e^{-\frac{t^2}{2}}\right| \leq c_k \frac{\varrho_{kn}}{n^{\frac{k-2}{2}}} |t|^k e^{-\frac{t^2}{4}}.$$

Lemma 2. *If the d.f.'s are equal, if k is an integer* ≥ 3, $\beta_{k\nu} < \infty$ *for* $\nu = 1, 2, 3, \ldots, n$, ϱ_{kn} *defined by* (10) *and*

$$|t| \leq T_{kn} = \frac{\sqrt{n}}{4\, \varrho_{kn}^{3/k}},$$

then

a. $$\left|\overline{f_n(t)} - e^{-\frac{t^2}{2}} - \sum_{\nu=1}^{k-3} \frac{P_\nu(it)}{n^{\nu/2}} e^{-\frac{t^2}{2}}\right| \leq \frac{c_k}{T_{kn}^{k-2}} (|t|^k + |t|^{3(k-2)})\, e^{-\frac{t^2}{4}},$$

and

b. $$\left|\overline{f_n(t)} - e^{-\frac{t^2}{2}} - \sum_{\nu=1}^{k-2} \frac{P_\nu(it)}{n^{\nu/2}} e^{-\frac{t^2}{2}}\right| \leq \frac{\delta(n)}{n^{\frac{k-2}{2}}} (|t|^k + |t|^{3(k-1)})\, e^{-\frac{t^2}{4}},$$

where $\delta(n)$ *only depends on n and* $\lim\limits_{n\to\infty} \delta(n) = 0$.

Here $P_\nu(i\,t) = \sum_{j=1}^{\nu} c_{j\nu}(i\,t)^{\nu+2j}$ *is a polynomial of degree* $3\,\nu$ *in* (it), *the coefficient* $c_{j\nu}$ *being a polynomial in* $\lambda_{3n}, \lambda_{4n}, \ldots, \lambda_{\nu-j+3,n}$ *and hence according to* (13) *independent of* n.

For example, if the d. f.'s $F_\nu(x)$ are equal, with the moments $\alpha_3, \alpha_4, \ldots$ and the dispersion σ, we obtain

$$(16)\qquad \begin{cases} P_1(i\,t) = \dfrac{1}{3!}\dfrac{\alpha_3}{\sigma^3}(i\,t)^3, \\[2ex] P_2(i\,t) = \dfrac{1}{4!}\dfrac{\alpha_4 - 3\,\alpha_2^2}{\sigma^4}(i\,t)^4 + \dfrac{10}{6!}\dfrac{\alpha_3^2}{\sigma^6}(i\,t)^6. \end{cases}$$

Proof of Theorem 1. The theorem is an immediate consequence of Theorem 2 a, Chap. II, and Lemma 1, this chap. In Theorem 2 a, Chap. II, we put

$$(17)\qquad \begin{cases} F(x) = \overline{F_n(x)}, & G(x) = \Phi(x), \\[1ex] f(t) = \overline{f_n(t)}, & g(t) = e^{-\frac{t^2}{2}}, \\[1ex] A = \operatorname{Max}|\Phi'(x)| = \dfrac{1}{\sqrt{2\pi}}, & T = \dfrac{\sqrt{n}}{(24\,\varrho_{kn})^{\frac{1}{k-2}}}. \end{cases}$$

It only remains to estimate

$$(18)\qquad \varepsilon = \int_{-T}^{T} \left| \frac{\overline{f_n(t)} - e^{-\frac{t^2}{2}}}{t} \right| d\,t.$$

From Lemma 1 we have

$$\varepsilon \leq c_k \frac{\varrho_{kn}}{n^{\frac{k-2}{2}}} \int_{-\infty}^{\infty} |\,t\,|^{k-1} e^{-\frac{t^2}{4}} d\,t = c_k \cdot \frac{\varrho_{kn}}{n^{\frac{k-2}{2}}},$$

and hence from the main theorem

$$|\overline{F_n(x)} - \Phi(x)| \leq \frac{a}{2\,\pi}\,\frac{c_k\,\varrho_{kn}}{n^{\frac{k-2}{2}}} + \frac{c(a)}{\sqrt{2\,\pi}}\,\frac{(24\,\varrho_{kn})^{\frac{1}{k-2}}}{n^{\frac{1}{2}}}$$

for every $a > 1$. This is the desired inequality.

Chapter IV.

Asymptotic Expansions in the Case of Equal Distribution Functions.

Consider a sequence of independent r. v.'s $X_1, X_2, \ldots, X_n$, all having the same d. f. $F(x)$, the c. f. $f(t)$, the mean value zero, the finite dispersion $\sigma \neq 0$, the moments α_k and the absolute moments β_k, $(k = 3, 4, \ldots)$. By $F_n(x)$ we denote the d. f. of the variable

$$Z_n = \frac{X_1 + X_2 + \cdots + X_n}{\sigma \sqrt{n}}$$

with the c. f. $f_n(t)$. These notations are used throughout the chapter. From (7), Chapter III, it follows that Z_n has the mean value zero and the dispersion 1. As before, the following relations hold:

$$\begin{cases} F_n(x) = (F(\sigma \sqrt{n}\, x))^{n*},^{1} \\ f_n(t) = \left\{ f\left(\dfrac{t}{\sigma \sqrt{n}}\right)\right\}^n. \end{cases} \tag{1}$$

In Theorem 1, Chap. III, we found that

$$|F_n(x) - \Phi(x)| \leq C \cdot \frac{\beta_3}{\sigma^3 \sqrt{n}},$$

C being an absolute constant, provided that $\beta_3 < \infty$. As was mentioned in the Introduction (5) in connection with the Charlier A series, it is sometimes possible to obtain an asymptotic expansion of $F_n(x)$ in $\Phi(x)$ and its derivatives, thus lowering the order of magnitude of the remainder term. At the same time a theoretical explanation of the usefulness of the A series is obtained.

The possibility of such an expansion is conditioned by the behaviour of $|f(t)|$ for large values of t. Three cases may occur which together cover all possibilities.

a. $\overline{\lim\limits_{t \to \pm\infty}} |f(t)| < 1$. This is the Cramér condition (C). The condition (C) being satisfied, Cramér has given an estimation of the remainder term of the Edgeworth expansion (5) in the introduction. (See also Theorem 1, this chap.)

The estimation of the remainder term becomes more delicate if (C) is no longer satisfied. We devote this chapter to the study of this question.

[1] This means the convolution of n functions $F(\sigma \sqrt{n}\, x)$.

b. $\overline{\lim_{t \to \pm\infty}} |f(t)| = 1$, but $|f(t)| < 1$ for every finite $t \neq 0$. We will show (Theorem 2) that

$$F_n(x) = \Phi(x) + \frac{\alpha_3}{6\sigma^3\sqrt{2\pi n}}(1 - x^2)e^{-\frac{x^2}{2}} + o\left(\frac{1}{\sqrt{n}}\right), \qquad (n \to \infty).$$

c. $|f(t_0)| = 1$ for a finite $t_0 \neq 0$. Then by Theorem 5, Chap. I, $F(x)$ is a *lattice distribution* and $|f(t)|$ is periodic, with the period t_0. Especially $|f(\nu t_0)| = 1$ for every $\nu = 0, \pm 1, \pm 2, \pm \cdots$. It will be shown, that $F_n(x)$ has discontinuities of order of magnitude $\frac{1}{\sqrt{n}}$. Cramér's estimation of the remainder term, valid in case a., breaks down for $k > 3$. By adding an expression containing a discontinuous term, we shall, however, obtain an expansion which makes it possible to lower the order of magnitude of the remainder term (Theorems 3 and 4).

For the different cases of the behaviour of $|f(t)|$ for large values of t reference is made to the end of Chap. I. We observe that (C) is satisfied if $F(x)$ has an absolutely continuous component. We should also remark that cases a. and c. are most frequently met with in statistical applications, the case b. having mainly a theoretical interest.

The reason for the dominating importance of the behaviour of $|f(t)|$ for large values of t may be explained by the following discussion. The proofs are based on Theorem 2, Chap. II. We have to estimate an integral of the type

$$I = \int_1^{\lambda} \frac{|f(t)|^n}{t} dt, \tag{2}$$

λ being suitably chosen. In case a. we obtain $|f(t)|^n < e^{-cn}$, $(c > 0, |t| \geq 1)$. Hence if λ is a power of n we have $I = O(e^{-c_1 n})$, $(c_1 > 0)$. In case b. we may choose $\lambda = \lambda(n) \to \infty$ when $n \to \infty$, so that $I = o\left(\frac{1}{\sqrt{n}}\right)$. In case c. even

$$\int_1^{\nu t_0} \frac{|f(t)|^n}{t} dt \sim \frac{\text{const.}}{\sqrt{n}}, \qquad (\nu \text{ an integer so that } \nu t_0 > 1),$$

as is easily confirmed. Now, the larger λ may be chosen and the smaller I, the smaller is the remainder term. Thus case a. is very favourable, case b. not so good and case c. the most unfavourable of all.

We devote § 1 of this chap. to Cramér's estimation of the remainder term in the Edgeworth expansion. In § 2 we treat case b., and in §§ 3 and 4 case c., using two different methods. Finally, in § 5, we investigate the asymptotic maximum deviation from the normal d. f.

1. **The Edgeworth expansion and Cramér's estimation of the remainder term.** We start from the relation

$$(3)\qquad (-it)^{\nu} e^{-\frac{t^2}{2}} = \int_{-\infty}^{\infty} e^{itx}\, d\,\Phi^{(\nu)}(x),$$

where $\Phi^{(\nu)}$ denotes the νth derivative of Φ. Consider the polynomial $P_\nu(it)$ in Lemma 2, Chap. III, and replace each power $(it)^{\nu+2j}$ by $(-1)^{\nu+2j}\,\Phi^{(\nu+2j)}(x)$. We then obtain a linear aggregate of the derivatives of $\Phi(x)$, symbolically denoted by $P_\nu(-\Phi)$:

$$(4)\qquad P_\nu(-\Phi) = \sum_{j=1}^{\nu} (-1)^{\nu+2j} c_{j\nu}\, \Phi^{(\nu+2j)}(x).$$

From (3) it is easily found that

$$(5)\qquad P_\nu(it) = \int_{-\infty}^{\infty} e^{itx}\, d\, P_\nu(-\Phi(x)).$$

For example, from (16), Chap. III:

$$(6)\qquad \begin{cases} P_1(-\Phi) = -\dfrac{1}{3!}\dfrac{\alpha_3}{\sigma^3}\Phi^{(3)}(x) = \dfrac{\alpha_3}{6\sigma^3\sqrt{2\pi}}(1-x^2)\, e^{-\frac{x^2}{2}}; \\ P_2(-\Phi) = \dfrac{1}{4!}\dfrac{\alpha_4 - 3\alpha_2^2}{\sigma^4}\Phi^{(4)}(x) + \dfrac{10}{6!}\dfrac{\alpha_3^2}{\sigma^6}\Phi^{(6)}(x). \end{cases}$$

Now we can formulate the following theorem:

Theorem 1. *Let $X_1, X_2, \ldots, X_n$ be a sequence of independent r. v.'s all having the same d. f. with the mean value zero, the dispersion $\sigma \neq 0$ and the finite absolute moment β_k, (k being an integer ≥ 3). If the condition*

$$(C)\qquad \overline{\lim_{t \to \pm\infty}}\, |f(t)| < 1$$

holds, then

$$(7)\qquad F_n(x) = \Phi(x) + \sum_{\nu=1}^{k-2} \frac{P_\nu(-\Phi)}{n^{\nu/2}} + o\left(\frac{1}{n^{\frac{k-2}{2}}}\right), \qquad (n \to \infty).$$

The proof is to be found in Cramér [3], p. 57 and [5], p. 81. There is, however, a slight difference between Cramér's expansion and (7), (cf. the Introduction (5)). By adding one more term to the expansion and applying Lemma 2 b, Chap. III, instead of Lemma 2 a, we have replaced the remainder term $O\left(\frac{1}{n^{\frac{k-2}{2}}}\right)$ by $o\left(\frac{1}{n^{\frac{k-2}{2}}}\right)\cdot$ The proof is analogous to that of Theorem 1, Chap. III, and is an immediate consequence of Lemma 2 b, Chap. III, and Theorem 2 a, Chap. II.

If $k = 3$, we have from Theorem 1 and (6):

$$F_n(x) = \Phi(x) + \frac{\alpha_3}{6\sigma^3\sqrt{2\pi n}}(1 - x^2)e^{-\frac{x^2}{2}} + o\left(\frac{1}{\sqrt{n}}\right), \qquad (n \to \infty), \tag{8}$$

provided that the condition (C) is satisfied. In the next section, however, we shall obtain (8) under less restrictive hypotheses.

2. **A further improvement of the Liapounoff remainder term.** We devote this section to the proof of the expansion (8) under more general conditions.

Theorem 2.[1] *Let $X_1, X_2, \ldots, X_n$ be a sequence of independent r. v.'s with the same d. f. $F(x)$, the c. f. $f(t)$, the mean value zero, the dispersion $\sigma \neq 0$, the third moment α_3 and the finite absolute third moment β_3. If $F(x)$ is not a lattice distribution, then*

$$F_n(x) = \Phi(x) + \frac{\alpha_3}{6\sigma^3\sqrt{2\pi n}}(1 - x^2)e^{-\frac{x^2}{2}} + o\left(\frac{1}{\sqrt{n}}\right), \qquad (n \to \infty). \tag{9}$$

Before we proceed to the proof, which is based on Theorem 2 a, Chap. II, we have to investigate an integral of the type (2).

Lemma 1. *If $F(x)$ is a d.f. which is not a lattice distribution, if $f(t)$ is the corresponding c. f. and if ω an assigned positive number, there exists a positive function $\lambda(n)$, so that*

$$\lim_{n \to \infty} \lambda(n) = \infty$$

and

$$I = \int_{\omega}^{\lambda(n)} \frac{|f(t)|^n}{t}\, dt = o\left(\frac{1}{\sqrt{n}}\right). \tag{10}$$

The proof is immediately clear if $\overline{\lim\limits_{t \to \pm\infty}} |f(t)| < 1$, for there then exists a constant $c > 0$ such that $|f(t)| < e^{-c}$ for $|t| \geq \omega$, (cf. the end of Chap. I). Putting $\lambda(n) = n$ we obtain

[1] Esseen [1], p. 14.

$$I \leq \int_{\omega}^{n} \frac{e^{-c \cdot n}}{t}\, dt = o\left(\frac{1}{\sqrt{n}}\right).$$

Now suppose that $\overline{\lim\limits_{t\to\pm\infty}} |f(t)| = 1$. Since $F(x)$ is not a lattice distribution, there does not exist any $t_0 \neq 0$ such that $|f(t_0)| = 1$. Now we define the function $\eta(t)$ by:

$$1 - \frac{1}{\eta(t)} = \operatorname*{Max}_{\omega \leq \tau \leq t} |f(\tau)|. \tag{11}$$

Obviously $\eta(t)$ is a continuous, non-decreasing function for all finite values of t. Since $\overline{\lim\limits_{t\to\pm\infty}} |f(t)| = 1$ we have

$$\lim_{t\to+\infty} \eta(t) = +\infty. \tag{12}$$

Now from (11):

$$I = \int_{\omega}^{\lambda(n)} \frac{|f(t)|^n}{t}\, dt \leq \int_{\omega}^{\lambda(n)} \frac{\left(1 - \frac{1}{\eta(t)}\right)^n}{t}\, dt. \tag{13}$$

For a given value of n we distinguish between two cases.

1. $\eta(n) \leq \sqrt{n}$. Putting $\lambda(n) = n$ we obtain from (13):

$$I \leq \int_{\omega}^{n} \frac{\left(1 - \frac{1}{\sqrt{n}}\right)^n}{t}\, dt \leq e^{-\frac{\sqrt{n}}{2}} \log\left(\frac{n}{\omega}\right) = o\left(\frac{1}{\sqrt{n}}\right).$$

2. $\eta(n) \geq \sqrt{n}$. Then from (13)

$$I \leq \int_{\omega}^{\lambda(n)} \frac{\left(1 - \frac{1}{\eta(\lambda(n))}\right)^n}{t}\, dt.$$

We now choose $\lambda(n) = \eta^{-1}(\sqrt{n})$, $\eta^{-1}(t)$ being the inverse function of $\eta(t)$. Obviously $\lim\limits_{t\to\infty} \eta^{-1}(t) = \infty$ and $\lambda(n) \leq n$. Thus

$$I \leq \int_{\omega}^{n} \frac{\left(1 - \frac{1}{\sqrt{n}}\right)^n}{t}\, dt = o\left(\frac{1}{\sqrt{n}}\right).$$

In either case $\lim\limits_{n\to\infty} \lambda(n) = \infty$ and $I = o\left(\frac{1}{\sqrt{n}}\right)$. This proves the lemma.

Proof of Theorem 2. By Lemma 2 b, Chap. III, we have

$$(14)\qquad \left| f_n(t) - e^{-\frac{t^2}{2}} - \frac{P_1(it)}{\sqrt{n}} e^{-\frac{t^2}{2}} \right| \leq C \cdot \frac{\delta(n)}{\sqrt{n}} (|t|^3 + |t|^6) e^{-\frac{t^2}{4}}$$

for $|t| \leq T_{3n}$, where $T_{3n} = \frac{\sqrt{n}}{4 \varrho_{3n}}$, $\varrho_{3n} = \frac{\beta_3}{\sigma^3}$ independent of n, C a constant and $\lim_{n \to \infty} \delta(n) = 0$. Further by (5) and (6) $P_1(it)$ is the Fourier-Stieltjes transform of

$$P_1(-\Phi) = \frac{\alpha_3}{6 \sigma^3 \sqrt{2 \pi}} (1 - x^2) e^{-\frac{x^2}{2}}.$$

Now apply Theorem 2 a, Chap. II, with

$$(15)\qquad \begin{cases} F(x) = F_n(x), & G(x) = \Phi(x) + \dfrac{P_1(-\Phi)}{\sqrt{n}}, \\ f(t) = f_n(t), & g(t) = e^{-\frac{t^2}{2}} + \dfrac{P_1(it)}{\sqrt{n}} e^{-\frac{t^2}{2}}, \\ A = \operatorname{Max} |G'(x)| < \infty, & \\ T = \lambda(n) \sqrt{n}\, \sigma, \text{ where } \lambda(n) \text{ is defined by Lemma 1 with } \omega = \dfrac{1}{4 \varrho_{3n} \cdot \sigma}. \end{cases}$$

Without loss of generality we may suppose $T \geq T_{3n}$.

It only remains to estimate

$$(16)\qquad \varepsilon = \int_{-T}^{T} \left| \frac{f_n(t) - g(t)}{t} \right| dt = \int_{-T}^{-T_{3n}} + \int_{-T_{3n}}^{T_{3n}} + \int_{T_{3n}}^{T} = \varepsilon_1 + \varepsilon_2 + \varepsilon_3.$$

From (14) it immediately follows that

$$(17)\qquad \varepsilon_2 = o\left(\frac{1}{\sqrt{n}}\right).$$

Furthermore it is easily seen that

$$\varepsilon_1 + \varepsilon_3 \leq 2 \int_{T_{3n}}^{T} \left| f\left(\frac{t}{\sigma \sqrt{n}}\right) \right|^n \frac{dt}{t} + o\left(\frac{1}{\sqrt{n}}\right) =$$

$$(18)\qquad = 2 \int_{\omega}^{\lambda(n)} |f(t)|^n \frac{dt}{t} + o\left(\frac{1}{\sqrt{n}}\right) = o\left(\frac{1}{\sqrt{n}}\right),$$

according to Lemma 1. Hence

$$\varepsilon = o\left(\frac{1}{\sqrt{n}}\right). \tag{19}$$

Now the fundamental theorem gives:

$$\left| F_n(x) - \Phi(x) - \frac{P_1(-\Phi)}{\sqrt{n}} \right| \leq \frac{k}{2\pi} o\left(\frac{1}{\sqrt{n}}\right) + c(k) \cdot \frac{A}{\lambda(n)\sqrt{n}\,\sigma}$$

for every $k > 1$. Hence the theorem is proved.

Discussion of Theorems 1 and 2. Let us first remark that Theorem 2 does not hold if $F(x)$ is a lattice distribution. As we shall see in the next section, $F_n(x)$ then has discontinuities in the vicinity of $x = 0$ of order of magnitude $\frac{1}{\sqrt{n}}$. Now suppose that $F(x)$ is not a lattice distribution. If $\overline{\lim\limits_{t \to \pm\infty}} |f(t)| < 1$, there are no difficulties and Theorem 1 gives a satisfactory solution, but if $\overline{\lim\limits_{t \to \pm\infty}} |f(t)| = 1$ it is difficult to improve upon Theorem 2, at least regarding the general case. The order of magnitude of the remainder term seems to depend on arithmetical properties of the point spectrum of the d. f.

If $\overline{\lim\limits_{t \to \pm\infty}} |f(t)| = 1$ and $F(x)$ is not a lattice distribution, it follows from § 7, Chap. I, that $F(x)$ may be a purely singular function. It is not to be expected that this case will occur in practice. A less theoretical case is that where $F(x)$ is a purely discontinuous function, the discontinuities of which occur in a sequence of points with incommensurable mutual distances. We may consider the following example. Let $F(x)$ be a d. f. with the jumps $\frac{1}{4}$ at the points $x = \pm 1$, $\pm \sqrt{2}$. Then $F(x)$ is not a lattice distribution, $f(t) = \frac{1}{2} \cos t + \frac{1}{2} \cos \sqrt{2}\, t$ and $\overline{\lim\limits_{t \to \pm\infty}} |f(t)| = 1$. If n is even, it is found, after some calculation, that $F_n(x)$ has a jump at $x = 0$, asymptotically equal to $\frac{2}{\pi} \cdot \frac{1}{n}$. Though all moments of $F(x)$ are finite, it is not possible to obtain the expansion

$$F_n(x) = \Phi(x) + \frac{P_1(-\Phi)}{\sqrt{n}} + \frac{P_2(-\Phi)}{n} + o\left(\frac{1}{n}\right). \tag{20}$$

In order to obtain an expansion like (20) in this and similar cases, where $f(t)$ is almost periodic, $\left(f(t) = \sum\limits_{\nu} a_\nu e^{i x_\nu t}\right)$, it is necessary to add a discontinuous function (cf. the next section), but this function is no doubt very complicated and dependent on the nature of the irrationalities in $\{x_\nu\}$.

3. **Lattice distributions. First method.** In order that the contents of this section may be more easily understood, let us first consider the case of the symmetrical Bernoulli distribution, $F(x)$ having the jumps $\frac{1}{2}$ at $x = \pm 1$. Here $F(x)$ is a lattice distribution and $f(t) = \cos t$. Now suppose that n is an *even* number. Then $F_n(x)$ is purely discontinuous with the discontinuity points $x = \frac{\nu}{\sqrt{n}}$, $(\nu = 0, \pm 2, \pm 4, \pm \cdots \pm n)$. As is seen in all works concerning the theory of probability, $F_n(x)$ has for a bounded discontinuity point x a jump, asymptotically equal to

$$\frac{2}{\sqrt{2\pi n}} e^{-\frac{x^2}{2}}. \tag{21}$$

Thus in the vicinity of $x = 0$ the jump of $F_n(x)$ and the growth of $\Phi(x)$ over an interval of length $\frac{2}{\sqrt{n}}$ are both equal to $\frac{2}{\sqrt{2\pi n}} + o\left(\frac{1}{\sqrt{n}}\right)$. Further $\frac{1}{2}(F_n(+0) + F_n(-0)) = \Phi(0) = \frac{1}{2}$. Hence the behaviour of $F_n(x)$ and $\Phi(x)$ about $x = 0$ may be represented by Figs. 6 and 7, where, however, the term $o\left(\frac{1}{\sqrt{n}}\right)$ has been neglected.

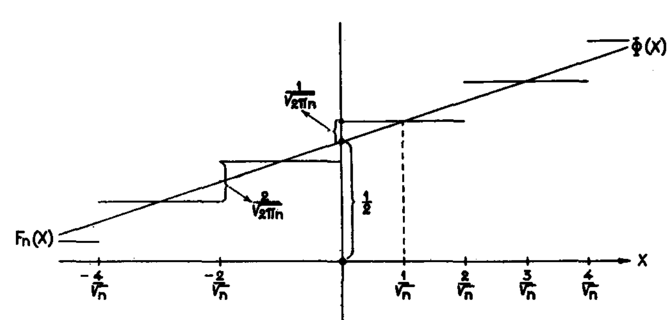

Fig. 6.

In Figs. 6 and 7 it is seen that

$$F_n(x) - \Phi(x) \sim \frac{2}{\sqrt{2\pi n}} Q_1\left(\frac{x\sqrt{n}}{2}\right) \tag{22}$$

for small values of x, where

$$Q_1(x) = [x] - x + \tfrac{1}{2}, \tag{23}$$

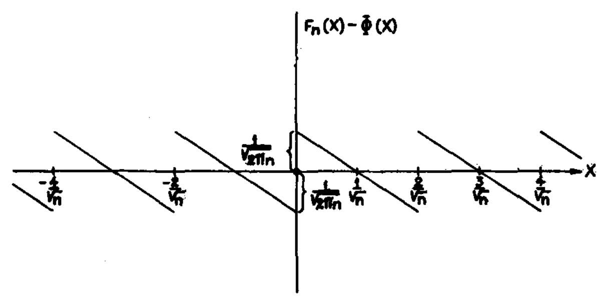

Fig. 7.

$[x]$ being the integral part of x. Now (21) and (22) suggest that we write

$$D_n(x) = \frac{2}{\sqrt{2\pi n}} Q_1\left(\frac{x\sqrt{n}}{2}\right) e^{-\frac{x^2}{2}} \tag{24}$$

and study the expression

$$F_n(x) - \Phi(x) - D_n(x).$$

By the expansion of $Q_1(x)$ in a Fourier series we easily evaluate

$$\Delta_n(t) = \int_{-\infty}^{\infty} e^{itx}\, d\, D_n(x)$$

and find

$$\Delta_n(t) = -\frac{it}{\pi\sqrt{n}} \sum_{\nu=-\infty}^{\infty}{}' \frac{1}{i\nu} e^{-\frac{1}{2}(t+\pi\sqrt{n}\cdot\nu)^2}, \tag{25}$$

the summation being performed for every integer $\nu \neq 0$. As we shall later prove in the general case

$$F_n(x) = \Phi(x) + D_n(x) + o\left(\frac{1}{\sqrt{n}}\right),$$

an expansion similar to that of Theorem 2.

After these preliminaries we proceed to the general case. Let $F(x)$ be a lattice distribution, i. e. a purely discontinuous d. f. with the jumps $a_\nu \geq 0$ situated in

$$x = x_0 + \nu \cdot d, \qquad (\nu = 0, \pm 1, \pm 2, \pm \cdots). \tag{26}$$

By definition $d\,(> 0)$ is the largest number for which (26) holds. For the sake of brevity we say that such a function $F(x)$ belongs to the class (L_d). According to § 7, Chap. I, $|f(t)|$ is periodic with the period t_0, where

(27) $$t_0 = \frac{2\pi}{d} \text{ and } f(t + t_0) = e^{i t_0 x_0} f(t).$$

By § 1, Chap. I, it follows that $F_n(x)$ is a lattice distribution with the jumps situated in

(28) $$x = \frac{1}{\sigma \sqrt{n}} (n x_0 + \nu_1 d + \nu_2 d + \cdots + \nu_n d),$$
$$(\nu_1, \nu_2, \ldots, \nu_n = 0, \pm 1, \pm 2, \pm \cdots).$$

Hence it is easily seen that the least non-negative discontinuity point ξ_n of $F_n(x)$ may be written as

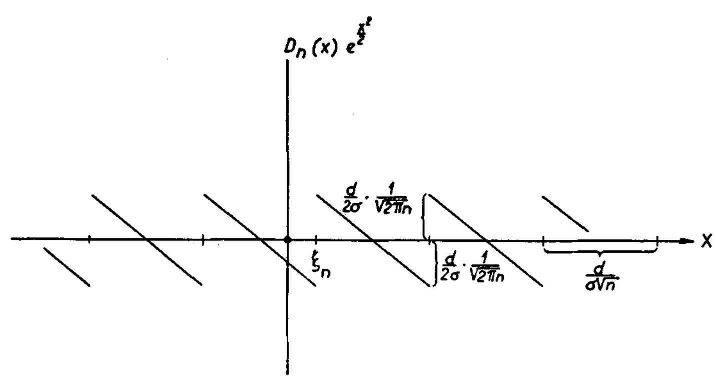

Fig. 8.

(29) $$\xi_n = \frac{d}{\sigma \sqrt{n}} \left\{ \frac{n x_0}{d} - \left[\frac{n x_0}{d} \right] \right\}.$$

In analogy with (24) let us write

(30) $$D_n(x) \equiv \frac{d}{\sigma \sqrt{2 \pi n}} Q_1 \left(\frac{(x - \xi_n) \sigma \sqrt{n}}{d} \right) e^{-\frac{x^2}{2}},$$

where ξ_n and $Q_1(x)$ are defined by (29) and (23).

Thus $D_n(x)$ is a discontinuous function with the period $\frac{d}{\sigma \sqrt{n}}$ and the jump $\frac{d}{\sigma \sqrt{2 \pi n}} e^{-\frac{x^2}{2}}$ at a discontinuity point x (Fig. 8).

We also put

(31) $$\psi_n(x) = Q_1 \left(\frac{(x - \xi_n) \sigma \sqrt{n}}{d} \right),$$

this function being periodic and discontinuous with the period $\frac{d}{\sigma \sqrt{n}}$ and the jump 1 at a discontinuity point. Thus

$$D_n(x) = \frac{d}{\sigma \sqrt{2 \pi n}} \psi_n(x) e^{-\frac{x^2}{2}}. \tag{32}$$

The Fourier-Stieltjes transform

$$\Delta_n(t) = \int_{-\infty}^{\infty} e^{itx} \, d D_n(x) \tag{33}$$

is easily evaluated by the expansion of $\psi_n(x)$ in a Fourier series:

$$\Delta_n(t) = -\frac{i t}{t_0 \sigma \sqrt{n}} \sum_{\nu=-\infty}^{\infty}{}' \frac{e^{-i t_0 x_0 n \nu}}{i \nu} e^{-\frac{1}{2}(t + t_0 \sigma \sqrt{n} \nu)^2}, \tag{34}$$

where t_0 is defined by (27) and the summation is performed for every integer $\nu \neq 0$.

We may now state the following theorem.

Theorem 3.[1] *Let $X_1, X_2, \ldots, X_n$ be a sequence of independent r. v.'s with the same d. f. $F(x)$, the mean value zero, the dispersion $\sigma \neq 0$, the third moment α_3 and the finite absolute third moment β_3. Suppose further that $F(x) < (L_d)$. Then,*

$$F_n(x) = \Phi(x) + \frac{\alpha_3}{6 \sigma^3 \sqrt{2 \pi n}} (1 - x^2) e^{-\frac{x^2}{2}} + \frac{d}{\sigma \sqrt{2 \pi n}} \psi_n(x) e^{-\frac{x^2}{2}} + o\left(\frac{1}{\sqrt{n}}\right) \tag{35}$$

as $n \to \infty$, where $\psi_n(x)$ is defined by (31).

Remarks.

1. We observe that $Q_1(x)$ is the same function that occurs in the Euler summation formula. Furthermore we may notice that $D_n(x)$, apart from n, only depends on the two parameters σ and d and on x_0, regarding the determination of an initial position.

2. We found in Theorem 2 that the expansion (9) holds for every d. f. $F(x)$ which is not a lattice distribution. From (35) it is now obvious why Theorem 2 breaks down if $F(x)$ is a lattice distribution. The expansion (35) contains the same terms as (9) and in addition a discontinuous function with jumps of order of magnitude $\frac{1}{\sqrt{n}}$ in the vicinity of $x = 0$.

[1] Esseen [2], p. 7, without proof.

Proof of Theorem 3. By $P_1(-\Phi)$, $P_1(it)$, T_{3n} and ϱ_{3n} we denote the same functions and quantities as in § 1. The quantity ϑ is an unspecified finite positive constant. We apply Theorem 2 b, Chap. II, putting:

$$(36)\qquad \begin{cases} F(x) = F_n(x), \quad G(x) = \Phi(x) + \dfrac{P_1(-\Phi)}{\sqrt{n}} + \dfrac{d}{\sigma\sqrt{2\pi n}}\psi_n(x)\,e^{-\frac{x^2}{2}}, \\ f(t) = f_n(t), \quad g(t) = e^{-\frac{t^2}{2}} + \dfrac{P_1(it)}{\sqrt{n}}e^{-\frac{t^2}{2}} + \Delta_n(t), \\ L = \dfrac{d}{\sigma\sqrt{n}}, \quad A = \dfrac{\vartheta}{\sqrt{n}}. \end{cases}$$

n is further supposed to be so large that

$$(37)\qquad T = n > T_{3n} = \frac{\sqrt{n}}{4\,\varrho_{3n}}, \text{ and } TL = \frac{d\sqrt{n}}{\sigma} > c_2(k),$$

$c_2(k)$ being the constant in Theorem 2 b, Chap. II.

It only remains to estimate

$$(38)\qquad \varepsilon = \int_{-T}^{T}\left|\frac{f_n(t)-g(t)}{t}\right|dt = \int_{-T}^{-\frac{t_0}{2}\sigma\sqrt{n}} + \int_{-\frac{t_0}{2}\sigma\sqrt{n}}^{\frac{t_0}{2}\sigma\sqrt{n}} + \int_{\frac{t_0}{2}\sigma\sqrt{n}}^{T} = \varepsilon_1 + \varepsilon_2 + \varepsilon_3.$$

1°. Without loss of generality we may suppose $T_{3n} < \frac{1}{2}t_0\sigma\sqrt{n}$. We further observe that $|f_n(t)| = \left|f\left(\frac{t}{\sigma\sqrt{n}}\right)\right|^n < e^{-cn}$ for a constant $c > 0$ in the intervals $T_{3n} \leq |t| \leq \frac{1}{2}t_0\sigma\sqrt{n}$. Thus

$$\varepsilon_2 \leq \int_{-T_{3n}}^{T_{3n}}\left|\frac{f_n(t)-g(t)}{t}\right|dt + O\left(\frac{1}{n}\right) + 2\cdot\int_{T_{3n}}^{\frac{1}{2}t_0\sigma\sqrt{n}}\frac{e^{-cn}}{t}\,dt,$$

or by Lemma 2 b, Chap. III, it is easily found in the usual way that

$$(39)\qquad \varepsilon_2 = o\left(\frac{1}{\sqrt{n}}\right).$$

2°. The estimation of ε_1 and ε_3 is somewhat more laborious. It is immediately clear that

$$\varepsilon_3 = \int\limits_{\frac{1}{2} t_0 \sigma \sqrt{n}}^{n} \left| \frac{f_n(t) - \Delta_n(t)}{t} \right| dt + O\left(\frac{1}{n}\right) = \varepsilon_4 + O\left(\frac{1}{n}\right). \tag{40}$$

Here

$$\varepsilon_4 = \int\limits_{\frac{t_0}{2}}^{\frac{\sqrt{n}}{\sigma}} \left| \frac{(f(t))^n - \Delta_n(\sigma \sqrt{n}\, t)}{t} \right| dt.$$

Now we write

$$\varepsilon_4 = \int\limits_{\frac{t_0}{2}}^{\frac{\sqrt{n}}{\sigma}} = \int\limits_{\frac{t_0}{2}}^{\frac{3 t_0}{2}} + \int\limits_{\frac{3 t_0}{2}}^{\frac{5 t_0}{2}} + \cdots + \int\limits_{k t_0 - \frac{t_0}{2}}^{k t_0 + \frac{t_0}{2}} + \cdots + \int\limits_{\left[\frac{\sqrt{n}}{\sigma t_0} - \frac{1}{2}\right] t_0 + \frac{t_0}{2}}^{\frac{\sqrt{n}}{\sigma}} = I_1 + I_2 + \cdots + I_k + \cdots. \tag{41}$$

Let us especially consider I_k and make the substitution $t = \tau + k t_0$. Hence from (27) and (34):

$$I_k = \int\limits_{-\frac{t_0}{2}}^{\frac{t_0}{2}} \left| \frac{e^{i t_0 x_0 n k} (f(t))^n + \dfrac{(t + k t_0)}{t_0} \sum\limits_{\nu = -\infty}^{\infty}{}' \dfrac{1}{\nu} e^{-i t_0 x_0 n \nu} e^{-\frac{1}{2}(t + k t_0 + \nu t_0)^2 \sigma^2 n}}{t + k t_0} \right| dt.$$

In Σ' only the term with $\nu = -k$ gives any considerable contribution to the integral. We obtain

$$I_k = \int\limits_{-\frac{t_0}{2}}^{\frac{t_0}{2}} \left| \frac{e^{i t_0 x_0 n k} (f(t))^n - e^{i t_0 x_0 n k} e^{-\frac{1}{2}\sigma^2 n t^2} - \dfrac{t}{k t_0} e^{i t_0 x_0 n k} e^{-\frac{1}{2}\sigma^2 n t^2}}{t + k t_0} \right| dt + O(e^{-\vartheta n}).$$

Hence

$$I_k \leq I_k' + I_k'' + O(e^{-\vartheta n}), \tag{42}$$

where

$$I_k' = \int\limits_{-\frac{t_0}{2}}^{\frac{t_0}{2}} \left| \frac{(f(t))^n - e^{-\frac{1}{2}\sigma^2 n t^2}}{t + k t_0} \right| dt,$$

$$I_k'' = \frac{1}{k\,t_0}\int\limits_{-\frac{t_0}{2}}^{\frac{t_0}{2}} \frac{|t|\,e^{-\frac{1}{2}\sigma^2 n t^2}}{t+k\,t_0}\,dt.$$

Now

$$I_k' \leq \frac{2}{k\,t_0}\int\limits_{-\frac{t_0}{2}}^{\frac{t_0}{2}} \left|(f(t))^n - e^{-\frac{1}{2}\sigma^2 n t^2}\right| dt = \frac{2}{k\,t_0\,\sigma\sqrt{n}}\int\limits_{-\frac{1}{2}t_0\sigma\sqrt{n}}^{\frac{1}{2}t_0\sigma\sqrt{n}} \left|f_n(t) - e^{-\frac{1}{2}t^2}\right| dt,$$

or as in 1°:

$$I_k' = \frac{\vartheta}{k\cdot n}.$$

Immediately we obtain

$$I_k'' = \frac{\vartheta}{k^2\cdot n}.$$

Hence from (42):

(43) $$I_k = \frac{\vartheta}{k\cdot n}.$$

Substituting (43) in (41) and performing the summation from $k=1$ to $k=O(\sqrt{n})$ we obtain

$$\varepsilon_4 = \frac{\vartheta}{n}\sum_{k=1}^{O(\sqrt{n})}\frac{1}{k} = O\left(\frac{\log n}{n}\right),$$

and thus from (40)

(44) $$\varepsilon_3 = O\left(\frac{\log n}{n}\right).$$

3°. In the same way

(45) $$\varepsilon_1 = O\left(\frac{\log n}{n}\right).$$

Summing up the results of 1°—3° we obtain

(46) $$\varepsilon = o\left(\frac{1}{\sqrt{n}}\right).$$

By Theorem 2 b, Chap. II, (36), (37), (38) and (46) we have

(47) $$\left|F_n(x) - \Phi(x) - \frac{P_1(-\Phi)}{\sqrt{n}} - \frac{d}{\sigma\sqrt{2\pi n}}\,\psi_n(x)\,e^{-\frac{x^2}{2}}\right| \leq \frac{k}{2\pi}\,o\left(\frac{1}{\sqrt{n}}\right) + c_1(k)\cdot\frac{\vartheta}{\sqrt{n}}\cdot\frac{1}{n}$$

for every $k > 1$. Hence the theorem is proved.

The expansion (35) is similar to the expansion (7) with $k=3$, only a discontinuous term being added. By $P_\nu(-\Phi)$ and $P_\nu(it)$ we understand the same functions as in § 1. If the Cramér condition (C) is satisfied and $F(x)$ has a finite absolute moment β_k of order k, (k an integer ≥ 3), then by Theorem 1

$$F_n(x) = \Pi_{n,k}(x) + o\left(\frac{1}{n^{\frac{k-2}{2}}}\right), \tag{48}$$

where

$$\Pi_{n,k}(x) = \Phi(x) + \sum_{\nu=1}^{k-2} \frac{P_\nu(-\Phi)}{n^{\nu/2}}. \tag{49}$$

Is it possible in the lattice distribution case to obtain an expansion analogous to (48) with the same order of magnitude of the remainder term?

We introduce the following functions, occurring in the Euler summation formula:

$$\left\{\begin{aligned} Q_1(x) &= \sum_{\nu=1}^{\infty} \frac{\sin 2\nu\pi x}{\nu\pi}; \\ Q_2(x) &= \sum_{\nu=1}^{\infty} \frac{\cos 2\nu\pi x}{2(\nu\pi)^2}; \\ &\cdots\cdots\cdots\cdots \\ Q_{2\lambda}(x) &= \sum_{\nu=1}^{\infty} \frac{\cos 2\nu\pi x}{2^{2\lambda-1}(\nu\pi)^{2\lambda}}; \\ Q_{2\lambda+1}(x) &= \sum_{\nu=1}^{\infty} \frac{\sin 2\nu\pi x}{2^{2\lambda}(\nu\pi)^{2\lambda+1}}. \end{aligned}\right. \tag{50}$$

These functions are all periodic with the period 1. For $0 \leq x < 1$ the following relations hold:

$$Q_1(x) = -x + \tfrac{1}{2}; \quad Q_2(x) = \frac{x^2}{2} - \frac{x}{2} + \frac{1}{12}; \quad Q_3(x) = \frac{x^3}{6} - \frac{x^2}{4} + \frac{x}{12}; \cdots \tag{51}$$

The functions $Q_2, Q_3, \ldots$ are continuous while Q_1 has the jump 1 at every integer x. Further

$$Q'_{2\lambda}(x) = -Q_{2\lambda-1}(x); \quad Q'_{2\lambda+1}(x) = Q_{2\lambda}(x); \quad Q_{2\lambda}(0) = \frac{B_\lambda}{\underline{|2\lambda}},$$

where B_λ are the Bernoullian numbers; $Q_{2\lambda+1}(0) = 0$ for $\lambda \geq 1$.

We can now formulate the following theorem.

Theorem 4. *Let $X_1, X_2, \ldots, X_n$ be a sequence of independent r.v.'s with the same d.f. $F(x)$, the mean value zero, the dispersion $\sigma \neq 0$ and the finite absolute moment β_k of order k, (k an integer ≥ 3). If $F(x) < (L_d)$, then*

$$F_n(x) = \Pi_{n,k}(x) + \sum_{\nu=1}^{k-2} h_\nu \cdot \left(\frac{d}{\sigma\sqrt{n}}\right)^\nu Q_\nu\left\{\frac{(x-\xi_n)\,\sigma\sqrt{n}}{d}\right\}\frac{d^\nu}{dx^\nu}(\Pi_{n,k}(x)) + o\left(\frac{1}{n^{\frac{k-2}{2}}}\right), \qquad (n \to \infty), \tag{52}$$

where $\Pi_{n,k}$ and Q_ν are defined by (49) and (50), ξ_n by (29) and where

$$h_\nu = \begin{cases} +1 \text{ for } \nu \text{ of the form } 4m+1,\ 4m+2, \\ -1 \text{ for } \nu \text{ of the form } 4m-1,\ 4m. \end{cases}$$

The proof follows from Theorem 2b, Chap. II, and offers no difficulties. It is analogous to the proof of Theorem 3, this theorem being a special case of Theorem 4 with $k = 3$. I confine myself to state the Fourier-Stieltjes transform $g(t)$ of the right-hand side of (52), (apart from the remainder term).

$$g(t) = e^{-\frac{t^2}{2}} + \sum_{\nu=1}^{k-2}\frac{P_\nu(it)}{n^{\nu/2}}e^{-\frac{t^2}{2}} - t\sum_{\mu=1}^{k-2}\ \sum_{\lambda=-\infty}^{\infty}{}' \left\{\frac{1}{(t_0\sigma\sqrt{n}\,\lambda)^\mu}e^{-it_0\sigma\xi_n\sqrt{n}\,\lambda}(t + t_0\sigma\sqrt{n}\,\lambda)^{\mu-1}\cdot e^{-\frac{1}{2}(t+t_0\sigma\sqrt{n}\,\lambda)^2}\left(1 + \sum_{\nu=1}^{k-2}\frac{P_\nu(it + it_0\sigma\sqrt{n}\,\lambda)}{n^{\nu/2}}\right)\right\}.$$

Remarks.

1. In the expansion (52) there are terms of order of magnitude less than $o\left(\frac{1}{n^{\frac{k-2}{2}}}\right)$. This is due to the fact that both the expression and the proof are more easy to handle in the present form.

2. By comparison between (48) and (52) it is seen how the discontinuities enter into the expansion in the lattice distribution case. Obviously the jump $a_n(\xi)$ of $F_n(x)$ at a discontinuity point ξ is expressed by

$$a_n(\xi) = \frac{d}{\sigma\sqrt{n}}\left\{\varphi(\xi) + \sum_{\nu=1}^{k-3}\frac{P_\nu(-\varphi(\xi))}{n^{\nu/2}}\right\} + o\left(\frac{1}{n^{\frac{k-2}{2}}}\right), \tag{53}$$

where

$$\varphi(\xi) = \frac{1}{\sqrt{2\pi}}e^{-\frac{\xi^2}{2}}.$$

In the next section we shall obtain (53) more directly with a better remainder term.

4. **Lattice distributions. Second method.** Let $F(x)$ be a lattice distribution with the c. f. $f(t)$. We begin this section by stating an expression for the jump of $F_n(x)$ at a discontinuity point ξ, using a method that goes back to LAPLACE [1] and has been applied with success by CHARLIER [2, 3] and others. Let us for the sake of simplicity suppose that $F(x)$ has the jumps $a_\nu \geq 0$ at $x = \nu$, $(\nu = 0, \pm 1, \pm 2, \pm 3, \pm \ldots)$. Then

$$f(t) = \sum_{\nu=-\infty}^{\infty} a_\nu e^{i\nu t}.$$

Thus

$$(f(t))^n = \sum_{\nu=-\infty}^{\infty} A_\nu e^{i\nu t},$$

where A_ν denotes the jump of $(F(x))^{n*}$ at $x = \nu$, $(F(x))^{n*}$ being a lattice distribution with the only possible discontinuities at $x = \nu$, $(\nu = 0, \pm 1, \pm 2, \pm 3, \pm \ldots)$. Hence

$$A_\nu = \frac{1}{2\pi} \int_{-\pi}^{\pi} (f(t))^n e^{-i\nu t}\, dt.$$

Owing to $F_n(x) = (F(\sigma\sqrt{n}\, x))^{n*}$, $A_\nu = a_n(\xi)$ is the jump of $F_n(x)$ at $\xi = \frac{\nu}{\sigma\sqrt{n}}$. Hence

$$(54) \quad a_n(\xi) = A_\nu = \frac{1}{2\pi} \int_{-\pi}^{\pi} (f(t))^n e^{-i\nu t}\, dt = \frac{1}{2\pi\sigma\sqrt{n}} \int_{-\pi\sigma\sqrt{n}}^{\pi\sigma\sqrt{n}} \left(f\left(\frac{t}{\sigma\sqrt{n}}\right)\right)^n e^{-i\frac{\nu}{\sigma\sqrt{n}}t}\, dt =$$

$$= \frac{1}{t_0\sigma\sqrt{n}} \int_{-\frac{1}{2}t_0\sigma\sqrt{n}}^{\frac{1}{2}t_0\sigma\sqrt{n}} f_n(t)\, e^{-i\xi t}\, dt,$$

since in this case $t_0 = 2\pi$. There is no difficulty in showing that (54) generally holds, $F(x)$ being a lattice distribution with the jumps $a_\nu \geq 0$ at $x = x_0 + \nu \cdot d$, $(\nu = 0, \pm 1, \pm 2, \pm \ldots)$.

Lemma 2. *If $F(x) < (L_d)$ with the finite dispersion $\sigma \neq 0$, then the jump $a_n(\xi)$ of $F_n(x)$ at a discontinuity point ξ is expressed by*

$$(55) \qquad a_n(\xi) = \frac{1}{t_0\sigma\sqrt{n}} \int_{-\frac{1}{2}t_0\sigma\sqrt{n}}^{\frac{1}{2}t_0\sigma\sqrt{n}} f_n(t)\, e^{-i\xi t}\, dt,$$

where $t_0 = \frac{2\pi}{d}$.

Applying Lemma 2 we can now prove the following theorem, where $P_\nu(-\varphi)$ has its usual meaning, only Φ being replaced by φ.

Theorem 5. *If $F(x) < (L_d)$ with the mean value zero, the dispersion $\sigma \neq 0$ and the finite absolute moment β_k of order k, (k an integer ≥ 3), then the jump $a_n(\xi)$ of $F_n(x)$ at a discontinuity point ξ is obtained by*

$$a_n(\xi) = \frac{d}{\sigma\sqrt{n}}\left(\varphi(\xi) + \sum_{\nu=1}^{k-2} \frac{P_\nu(-\varphi(\xi))}{n^{\nu/2}}\right) + o\left(\frac{1}{n^{\frac{k-1}{2}}}\right) \tag{56}$$

as $n \to \infty$, where $\varphi(\xi) = \Phi'(\xi) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\xi^2}{2}}$.

Compare (53) and (56)!

Proof of Theorem 5. We apply Lemma 2 and write (55) as:

$$a_n(\xi) = \frac{1}{t_0\sigma\sqrt{n}} \int_{-\frac{1}{2}t_0\sigma\sqrt{n}}^{\frac{1}{2}t_0\sigma\sqrt{n}} \left[f_n(t) - e^{-\frac{t^2}{2}} - \sum_{\nu=1}^{k-2} \frac{P_\nu(it)}{n^{\nu/2}} e^{-\frac{t^2}{2}}\right] e^{-i\xi t}\,dt + \tag{57}$$

$$+ \frac{1}{t_0\sigma\sqrt{n}} \int_{-\frac{1}{2}t_0\sigma\sqrt{n}}^{\frac{1}{2}t_0\sigma\sqrt{n}} \left[e^{-\frac{t^2}{2}} + \sum_{\nu=1}^{k-2} \frac{P_\nu(it)}{n^{\nu/2}} e^{-\frac{t^2}{2}}\right] e^{-i\xi t}\,dt = I_1 + I_2.$$

Observing that $|f_n(t)| < e^{-cn}$, (c a positive constant), for $a\sigma\sqrt{n} \leq |t| \leq \frac{t_0}{2}\sigma\sqrt{n}$, where a is an assigned positive number, we obtain in the usual way by Lemma 2 b, Chap. III:

$$I_1 = o\left(\frac{1}{n^{\frac{k-1}{2}}}\right). \tag{58}$$

Further by (5):

$$I_2 = \frac{1}{t_0\sigma\sqrt{n}} \int_{-\infty}^{\infty} \left(e^{-\frac{t^2}{2}} + \sum_{\nu=1}^{k-2} \frac{P_\nu(it)}{n^{\nu/2}} e^{-\frac{t^2}{2}}\right) e^{-i\xi t}\,dt + o\left(\frac{1}{n^{\frac{k-1}{2}}}\right) = \tag{59}$$

$$= \frac{d}{\sigma\sqrt{n}}\left\{\varphi(\xi) + \sum_{\nu=1}^{k-2} \frac{P_\nu(-\varphi(\xi))}{n^{\nu/2}}\right\} + o\left(\frac{1}{n^{\frac{k-1}{2}}}\right).$$

The proof follows from (57), (58) and (59).

If $F(x)$ is a lattice distribution it is possible to obtain a very simple expression of $F_n(x)$.

Lemma 3. *Let $F(x) < (L_d)$ with the finite dispersion $\sigma \neq 0$. Then the discontinuities of $F_n(x)$ are situated in the sequence*

$$\xi_\nu = \xi_0 + \nu \cdot \frac{d}{\sigma \sqrt{n}}, \qquad (\nu = 0, \pm 1, \pm 2, \pm \ldots). \tag{60}$$

Let each of the points x_1 and x_2 be situated midway between two consecutive points of the sequence $\{\xi_\nu\}$:

$$x_1 = \xi' - \tfrac{1}{2}\frac{d}{\sigma\sqrt{n}}, \quad x_2 = \xi'' + \tfrac{1}{2}\frac{d}{\sigma\sqrt{n}},$$

where ξ' and ξ'' belong to $\{\xi_\nu\}$. Then

$$F_n(x_2) - F_n(x_1) = \frac{1}{t_0 \sigma \sqrt{n}} \int_{-\frac{1}{2} t_0 \sigma \sqrt{n}}^{\frac{1}{2} t_0 \sigma \sqrt{n}} f_n(t) \frac{e^{-i x_1 t} - e^{-i x_2 t}}{2 i \sin \frac{d \cdot t}{2 \sigma \sqrt{n}}} \, dt, \tag{61}$$

where $t_0 = \frac{2\pi}{d}$.

Proof. As before $a_n(\xi)$ denotes the jump of $F_n(x)$ at $x = \xi$. Then by Lemma 2

$$F_n(x_2) - F_n(x_1) = \sum_{\xi' \leq \xi \leq \xi''} a_n(\xi) = \frac{1}{t_0 \sigma \sqrt{n}} \int_{-\frac{1}{2} t_0 \sigma \sqrt{n}}^{\frac{1}{2} t_0 \sigma \sqrt{n}} f_n(t) \left(\sum_{\xi' \leq \xi \leq \xi''} e^{-i\xi t} \right) dt. \tag{62}$$

If $\xi' = \xi_0 + \nu_1 \cdot \frac{d}{\sigma\sqrt{n}}$ and $\xi'' = \xi_0 + \nu_2 \cdot \frac{d}{\sigma\sqrt{n}}$ we obtain by (60):

$$\sum_{\xi' \leq \xi \leq \xi''} e^{-i\xi t} = \sum_{\nu_1 \leq \nu \leq \nu_2} e^{-it\left(\xi_0 + \nu \cdot \frac{d}{\sigma\sqrt{n}}\right)} = \frac{e^{-i x_1 t} - e^{-i x_2 t}}{2 i \sin \frac{d \cdot t}{2 \sigma \sqrt{n}}}.$$

Inserting this into (62) we obtain the desired result.

Using Lemma 3 we can prove all the theorems of § 3. Let us for instance consider Theorem 3, the notations being unaltered. By (61) it follows:

$$F_n(x_2) - F\ (x_1) = \frac{1}{t_0 \sigma \sqrt{n}} \int_{-\frac{1}{2} t_0 \sigma \sqrt{n}}^{\frac{1}{2} t_0 \sigma \sqrt{n}} f_n(t) \frac{e^{-i x_1 t} - e^{-i x_2 t}}{2 i \sin \frac{d \cdot t}{2 \sigma \sqrt{n}}} \, dt = \tag{63}$$

$$= \frac{1}{t_0 \sigma \sqrt{n}} \left(\int_{-\frac{1}{2} t_0 \sigma \sqrt{n}}^{-T_{3n}} + \int_{-T_{3n}}^{T_{3n}} + \int_{T_{3n}}^{\frac{1}{2} t_0 \sigma \sqrt{n}} \right) = I_1 + I_2 + I_3.$$

Without loss of generality we may suppose $\frac{1}{2} t_0 \sigma \sqrt{n} > T_{3n}$. As in the proof of Theorem 3, (1°),

$$I_1 + I_3 = O\left(\frac{1}{n}\right). \tag{64}$$

Now

$$I_2 = \frac{1}{t_0 \sigma \sqrt{n}} \int_{-T_{3n}}^{T_{3n}} (f_n(t) - g(t)) \frac{e^{-i x_1 t} - e^{-i x_2 t}}{2 i \sin \frac{d \cdot t}{2 \sigma \sqrt{n}}} d t + \tag{65}$$

$$+ \frac{1}{t_0 \sigma \sqrt{n}} \int_{-T_{3n}}^{T_{3n}} g(t) \frac{e^{-i x_1 t} - e^{-i x_2 t}}{2 i \sin \frac{d \cdot t}{2 \sigma \sqrt{n}}} d t = I_2' + I_2'',$$

where

$$g(t) = e^{-\frac{t^2}{2}} + \frac{P_1(i t)}{\sqrt{n}} e^{-\frac{t^2}{2}}.$$

But from the proof of Theorem 3, (1°), it follows:

$$I_2' = o\left(\frac{1}{\sqrt{n}}\right). \tag{66}$$

Using the expansion $\frac{1}{\sin v} = \frac{1}{v} + O(v)$ for $|v| \leq \frac{\pi}{2}$ we obtain:

$$I_2'' = \frac{1}{2 t_0 \sigma \sqrt{n}} \cdot \frac{2 \sigma \sqrt{n}}{d} \int_{-T_{3n}}^{T_{3n}} g(t) \frac{e^{-i x_1 t} - e^{-i x_2 t}}{i t} d t + O\left(\frac{1}{n}\right) = \tag{67}$$

$$= \frac{1}{2 \pi} \int_{-\infty}^{\infty} g(t) \frac{e^{-i x_1 t} - e^{-i x_2 t}}{i t} d t + O\left(\frac{1}{n}\right).$$

If now

$$G(x) = \Phi(x) + \frac{P_1(-\Phi)}{\sqrt{n}} \tag{68}$$

it follows from (5), (67) and Chap. I, (5):

$$I_2'' = G(x_2) - G(x_1) + O\left(\frac{1}{n}\right). \tag{69}$$

Summing up the results of (63)—(69) we have:

$$F_n(x_2) - F_n(x_1) = G(x_2) - G(x_1) + o\left(\frac{1}{\sqrt{n}}\right),$$

or if $x_1 \to -\infty$,

$$F_n(x_2) = G(x_2) + o\left(\frac{1}{\sqrt{n}}\right). \tag{70}$$

The expansion (70) is valid for every x_2 situated midway between two consecutive points in $\{\xi_\nu\}$. Since $F_n(x)$ is constant between two discontinuity points, Theorem 3 now follows from (70).

5. **On the asymptotic maximum deviation from $\Phi(x)$.** Throughout this section we assume that $F(x)$ is a distribution function with the mean value zero, the dispersion $\sigma \neq 0$, the third moment α_3 and the finite absolute third moment β_3. By (L) we denote the class of lattice distribution functions and by (L_d) the class of lattice distribution functions with the distance d between the equidistant points.

By combination of Theorems 2 and 3 the following theorem is obtained.

Theorem 6. *If $\alpha_3 = 0$, then*

$$\lim_{n\to\infty} \operatorname*{Max}_{-\infty<x<\infty} \sqrt{n}\,|F_n(x) - \Phi(x)| = \begin{cases} 0, & \textit{if } F(x) \not< (L) \\ \dfrac{d}{2\sigma\sqrt{2\pi}} & \textit{if } F(x) < (L_d) \end{cases}.$$

$F(x)$ being subject to some further conditions we can state:

Theorem 7. *If $F(x)$ is symmetrical, i. e. $F(-x) = 1 - F(x)$, and continuous at $x = 0$, then*

$$\lim_{n\to\infty} \operatorname*{Max}_{-\infty<x<\infty} \sqrt{n}\,|F_n(x) - \Phi(x)| \leq \frac{1}{\sqrt{2\pi}}.$$

There is equality if and only if $F(x)$ is the symmetrical Bernoulli distribution function, having the jump $\frac{1}{2}$ at $x = \pm a$, where a is a positive constant.

Proof of Theorem 7. Here $\alpha_3 = 0$. By Theorem 6 it is sufficient to treat the case $F(x) < (L_d)$. We thus have to find an upper bound of

$$\frac{d}{2\sigma\sqrt{2\pi}}. \tag{71}$$

We may suppose on grounds of homogeneity that $d = 1$. Thus under the given conditions $F(x)$ has the jump $a_\nu \geq 0$ for $x = \pm(\nu + \frac{1}{2})$, $(\nu = 0, 1, 2, 3, \ldots)$. Hence

$$\sum_{\nu=0}^{\infty} a_\nu = \tfrac{1}{2} \text{ and } \sigma^2 = 2 \cdot \sum_{\nu=0}^{\infty} (\nu + \tfrac{1}{2})^2 a_\nu.$$

It is easily seen that the least possible value of σ is equal to $\frac{1}{2}$, $F(x)$ then being the symmetrical Bernoulli distribution function with the jump $\frac{1}{2}$ at $x = \pm\frac{1}{2}$. Hence $\dfrac{d}{2\sigma\sqrt{2\pi}} \leq \dfrac{1}{\sqrt{2\pi}}$ and the theorem is proved.

Remarks.

1. If $F(x)$ does not satisfy the conditions in Theorem 7, symmetry and continuity at $x = 0$, it is readily seen from examples that (71) is no longer bounded.

2. Suppose that two persons, Peter and Paul, play a game of chance so that the players may win or lose certain sums of money at every round, the chances of Paul being represented by a purely discontinuous d. f. $F(x)$. Suppose further that $F(x)$ has the mean value and the third moment α_3 zero, the dispersion $\sigma \neq 0$ and the finite absolute third moment β_3. Let us first assume that all the winnings and losses are measured in the same monetary unit. Then $F(x) < (L)$ and the d. f. $(F(x))^{n*}$ of Paul's gain or loss after a large number n of games is approximately normal with a possible error term of order of magnitude $\frac{1}{\sqrt{n}}$. Let us now assume (if possible) that the game is such that some of the winnings and losses are measured in one sort of monetary unit, some in another and that the two units are incommensurable. Then $F(x) \nless (L)$ and by Theorem 6, $(F(x))^{n*}$ differs from the normal d. f. with an error of order of magnitude *less* than in the previous case. Let us return to the first case, assuming $F(x)$ to be symmetrical and continuous at $x = 0$. From Theorem 7 it follows that among the possible games the old game »pitch and toss» gives a $(F(x))^{n*}$ which in the long run most differs from the normal distribution.

Chapter V.

Dependence of the Remainder Term on n and x.

In the two preceding chapters we have investigated the difference between the d. f. $F_n(x)$ of the normalized sum of a large number of independent r. v.'s and the normal d. f. $\Phi(x)$ or a sum of $\Phi(x)$ and its derivatives. We thus obtained an error term containing, besides certain constants, only the parameter n. It is, however, often important not only to estimate the remainder term as a function of n but also as a function of x. This question has earlier been treated in an interesting paper by CRAMÉR [6], who assumes that all the variables have the same d. f. $F(x)$ and that there exists a constant $\sigma > 0$ such that

$$\int_{-\infty}^{\infty} e^{\sigma |x|} \, dF(x) < \infty .$$

In this chapter we shall consider the case where

$$\int_{-\infty}^{\infty} |x|^k \, dF(x) < \infty \tag{1}$$

for an integer $k \geq 2$. Then also

$$\int_{-\infty}^{\infty} |x|^k \, dF_n(x) < \infty . \tag{2}$$

Owing to the well-known generalization of the Bienaymé-Tchebycheff inequality it follows from (2):

$$1 - F_n(x) = O\left(\frac{1}{|x|^k}\right) \text{ for } x \to +\infty;$$

$$F_n(x) = O\left(\frac{1}{|x|^k}\right) \text{ for } x \to -\infty.$$

Hence it is to be expected that

$$|F_n(x) - \Phi(x)| \leq \frac{\vartheta(n)}{1 + |x|^k}, \qquad (x \to \pm \infty),$$

where $\vartheta(n)$ is a quantity tending to zero as $n \to \infty$, provided that certain conditions are satisfied. In this chapter we shall prove the correctness of this supposition. We begin, however, with some remarks concerning the Central Limit Theorem of the theory of probability.

1. **On the Central Limit Theorem.** Let us consider a sequence of independent r.v.'s $X_1, X_2, \ldots, X_n, \ldots$, the variable X_ν, $(\nu = 1, 2, 3, \ldots)$, having the mean value zero and the finite dispersion σ_ν. By $\overline{F_n(x)}$ we denote the d.f. of the variable

$$Z_n = \frac{X_1 + X_2 + \cdots + X_n}{s_n}$$

where $s_n^2 = \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2$. As before the mean value of Z_n is zero and

$$\int_{-\infty}^{\infty} x^2 \, d\overline{F_n(x)} = 1. \tag{3}$$

Under certain very general conditions the sequence $\{\overline{F_n(x)}\}$ converges to the normal d.f. $\Phi(x)$ (the Central Limit Theorem, see Lindeberg [1]). This may be expressed by

$$\Delta(n) = \operatorname*{Max}_{-\infty < x < \infty} |\overline{F_n(x)} - \Phi(x)|; \quad \lim_{n \to \infty} \Delta(n) = 0. \tag{4}$$

According to (3) and

$$\int_{-\infty}^{\infty} x^2 \, d\,\Phi(x) = 1 \tag{5}$$

we also have by the Bienaymé-Tchebycheff inequality

$$|\overline{F_n(x)} - \Phi(x)| = O\left(\frac{1}{x^2}\right) \text{ as } x \to \infty. \tag{6}$$

The relations (4) and (6) give two different estimations of $|F_n(x) - \Phi(x)|$; in (4) the remainder term depends only on n, in (6) only on x. Is it possible to take both (4) and (6) into consideration, using one inequality only? This is, as we shall see, very easy, but has, as far as I know, not hitherto been explicitly stated.

Let $a \geq 1$ be a number later to be determined. We may without loss of generality suppose that $F_n(x)$ is continuous at $x = \pm a$. Then

$$\int_{-a}^{a} x^2 \, d\,\overline{F_n(x)} = \int_{-a}^{a} x^2 \, d\,(\overline{F_n(x)} - \Phi(x)) + \int_{-a}^{a} x^2 \, d\,\Phi(x) =$$

$$= a^2(\overline{F_n(a)} - \Phi(a)) - a^2(\overline{F_n(-a)} - \Phi(-a)) - 2\int_{-a}^{a} x\,(\overline{F_n(x)} - \Phi(x))\,dx + \int_{-a}^{a} x^2 \, d\,\Phi(x).$$

From (4) it follows:

$$\int_{-a}^{a} x^2 \, d\,\overline{F_n(x)} \geq -4a^2 \Delta(n) + \int_{-a}^{a} x^2 \, d\,\Phi(x). \tag{7}$$

From (3), (5) and (7):

$$\int_{|x| \geq a} x^2 \, d\,\overline{F_n(x)} = 1 - \int_{-a}^{a} x^2 \, d\,\overline{F_n(x)} \leq \int_{|x| \geq a} x^2 \, d\,\Phi(x) + 4a^2 \Delta(n). \tag{8}$$

Now the following relations hold:

$$\int_{|x| \geq a} x^2 \, d\,\overline{F_n(x)} \geq \begin{cases} x^2(1 - \overline{F_n(x)}) & \text{for } x \geq a \\ x^2\,\overline{F_n(x)} & \text{for } x \leq -a \end{cases}, \tag{9}$$

$$\int_{|x| \geq a} x^2 \, d\,\Phi(x) \geq \begin{cases} x^2(1 - \Phi(x)) & \text{for } x \geq a \\ x^2\,\Phi(x) & \text{for } x \leq -a \end{cases}, \tag{10}$$

$$\int_{|x| \geq a} x^2 \, d\,\Phi(x) \leq \sqrt{\frac{2}{\pi}}\,\frac{a^2 + 1}{a}\,e^{-\frac{a^2}{2}}. \tag{11}$$

From (8), (9) and (10) we obtain:

$$x^2\,|\overline{F_n(x)} - \Phi(x)| \leq 4\,a^2\,\Delta(n) + \int_{|x|\geq a} x^2\,d\,\Phi(x) \text{ for } |x| \geq a,$$

and hence from (4) and (11):

$$(12) \qquad (1 + x^2)\,|\overline{F_n(x)} - \Phi(x)| \leq \sqrt{\frac{2}{\pi}}\,\frac{a^2+1}{a}\,e^{-\frac{a^2}{2}} + 5\,a^2\,\Delta(n).$$

This inequality obviously holds not only for $|x| \geq a$ but also for all values of x. In (12) we choose a such that the two terms of the right-hand side have the same order of magnitude in n. We suppose n_0 to be so large that $\Delta(n) \leq \frac{1}{2}$ for $n > n_0$. Then, putting

$$a = \sqrt{2 \log \frac{1}{\Delta(n)}},$$

we obtain:

$$(13) \qquad |\overline{F_n(x)} - \Phi(x)| \leq C \cdot \frac{\Delta(n) \log \frac{1}{\Delta(n)}}{1 + x^2} \text{ for } n > n_0,$$

where C is an absolute constant.

The inequality (13) gives, together with (4), the following theorem.

Theorem 1. *Let $X_1, X_2, \ldots, X_n, \ldots$ be a sequence of independent r. v.'s such that the variable X_ν has the mean value zero and the finite dispersion σ_ν, ($\nu = 1, 2, 3, \ldots, n, \ldots$). Further, let $\overline{F_n(x)}$ denote the d. f. of the variable*

$$\frac{X_1 + X_2 + \cdots + X_n}{s_n}$$

where $s_n^2 = \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2$. If

$$\Delta(n) = \operatorname*{Max}_{-\infty<x<\infty} |\overline{F_n(x)} - \Phi(x)|$$

and $\Delta(n) \leq \frac{1}{2}$ for $n > n_0$, there exists an absolute constant C, such that

$$|\overline{F_n(x)} - \Phi(x)| \leq \operatorname{Min}\left\{\Delta(n);\ C \cdot \frac{\Delta(n) \log \frac{1}{\Delta(n)}}{1 + x^2}\right\}$$

for $n > n_0$ and all values of x.

2. **On the remainder term of the asymptotic expansion.** In the remainder of this chapter we consider a sequence of independent r. v.'s $X_1, X_2, \ldots, X_n, \ldots$ with the same d. f. $F(x)$, the c. f. $f(t)$, the mean value zero, the dispersion $\sigma \neq 0$,

the moments α_k and the absolute moments β_k. By $F_n(x)$ we denote as usual the d. f. of the variable

$$\frac{X_1 + X_2 + \cdots + X_n}{\sigma\sqrt{n}}$$

with the c. f. $f_n(t)$. As before

$$(14)\qquad \begin{aligned} F_n(x) &= (F(\sigma\sqrt{n}\,x))^{n*}, \\ f_n(t) &= \left(f\left(\frac{t}{\sigma\sqrt{n}}\right)\right)^n. \end{aligned}$$

The functions $P_\nu(-\Phi)$ and $P_\nu(it)$ are the same as in Chap. IV, § 1.

In Chap. IV, Theorem 1, we stated the expansion

$$(15)\qquad \left| F_n(x) - \Phi(x) - \sum_{\nu=1}^{k-3} \frac{P_\nu(-\Phi)}{n^{\nu/2}} \right| \le \frac{\text{const.}}{n^{\frac{k-2}{2}}}, \qquad (k \text{ an integer} \ge 3),$$

provided that the condition (C) is satisfied. In the same chapter we found, however, that it is generally not possible to obtain anything better than

$$(16)\qquad | F_n(x) - \Phi(x) | \le \frac{\text{const.}}{\sqrt{n}},$$

without introducing a discontinuous term. Now we shall show that (15) always holds if $|x|$ is sufficiently large. At the same time we shall obtain the dependence on x of the remainder terms in (15) and (16). The method of proof is analogous to that of Theorem 2, Chap. II.

We begin by sketching the main features of the proofs. Our problem consists in the comparison of $F_n(x)$ with a certain function $G_n(x)$ (in the following $\Phi(x)$ or the terms in the Edgeworth expansion) satisfying the conditions:

$$(17)\qquad \begin{cases} G_n(x) \text{ is real and of bounded variation on the whole real axis,} \\ G_n(-\infty) = 0, \quad G_n(+\infty) = 1. \end{cases}$$

Further, let $Q(x)$ and $q(t)$ be two real even functions such that

$$(18)\qquad \begin{aligned} &1^\circ \quad Q(x) \ge 0,\ 0 \le q(t) \le 1,\ \int_{-\infty}^{\infty} Q(x)\,dx = 1,\ \int_{-\infty}^{\infty} q(t)\,dt < \infty; \\ &2^\circ \qquad q(t) = \int_{-\infty}^{\infty} e^{itx} Q(x)\,dx. \end{aligned}$$

Other properties of the two functions are later specified. As in the proof of Theorem 2, Chap. II,

$$\int_{-\infty}^{\infty} \lambda\, Q(\lambda x - \lambda y)\,[G_n(y) - F_n(y)]\,dy = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\frac{g_n(t) - f_n(t)}{-it}\,q\left(\frac{t}{\lambda}\right)dt, \tag{19}$$

$g_n(t)$ being the Fourier-Stieltjes transform of $G_n(x)$, x and λ two parameters. From (19) we wish to obtain an estimation of $|F_n(y) - G_n(y)|$ as a function of both n and y, assuming that it is already known that

$$|F_n(y) - G_n(y)| \leq d(n) \text{ for all } y. \tag{20}$$

We now put

$$\Delta(b) = \underset{|y|>b}{\mathrm{Max}}\,|F_n(y) - G_n(y)|, \tag{21}$$

(b a positive quantity). It is sufficient to treat the case $y > 0$ only. Let us suppose that this maximum occurs for $y = a \geq b$ and that $G_n(a) > F_n(a)$. (The case $G_n(a) < F_n(a)$ is treated in the same way). We wish to demonstrate the inequality

$$\Delta(b) \leq \text{const.} \cdot \varrho(a, n), \tag{22}$$

$\varrho(a, n)$ being a positive function which tends steadily to zero as a and n separately tend to infinity. Two cases may occur.

$$1^\circ \quad \Delta(b) \leq \varrho(a, n). \tag{23}$$

This is the desired inequality.

$$2^\circ \quad \Delta(b) > \varrho(a, n).$$

Now we determine a number $\xi > 0$ such that

$$G_n(a) - G_n(y) < \frac{\Delta(b)}{2} \text{ for } a - \xi \leq y \leq a. \tag{24}$$

Thus

$$G_n(y) - F_n(y) \geq G_n(a) - \frac{\Delta(b)}{2} - F_n(a) = \frac{\Delta(b)}{2} \text{ for } a - \xi \leq y \leq a. \tag{25}$$

As in the proof of Theorem 2, Chap. II, we now obtain from (19), (20) and (25)

$$\frac{\Delta(b)}{2}\int_{-\frac{\lambda\xi}{2}}^{\frac{\lambda\xi}{2}} Q(y)\,dy - 2\,d(n)\int_{\frac{\lambda\xi}{2}}^{\infty} Q(y)\,dy \leq \left|\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\frac{g_n(t) - f_n(t)}{-it}\,q\left(\frac{t}{\lambda}\right)dt\right|, \tag{26}$$

putting

$$x = a - \frac{\xi}{2}. \tag{27}$$

In the following the relation (26) is our main inequality. From (26) we generally obtain:

$$\Delta(b) \leq \text{const.} \cdot \varrho(a, n). \tag{28}$$

Hence from (23) and (28) the required result follows.

We first prove the following theorem.

Theorem 2. *If $X_1, X_2, \ldots, X_n$ are a sequence of independent r. v.'s with the same d. f. $F(x)$, the c. f. $f(t)$, the mean value zero, the dispersion $\sigma \neq 0$ and the finite absolute moment β_k, (k an integer ≥ 3), then*

$$\left| F_n(x) - \Phi(x) - \sum_{\nu=1}^{k-3} \frac{P_\nu(-\Phi)}{n^{\nu/}} \right| \leq \frac{c(\delta, \beta)}{1 + |x|^k} \cdot \frac{1}{n^{\frac{k-2}{2}}} \tag{29}$$

for

$$|x| \geq \sqrt{(1 + \delta)(k - 2) \log n}, \tag{30}$$

where δ is an assigned number such that $0 < \delta < 1$ and $c(\delta, \beta)$ is a finite positive constant, only depending on δ and the moments $\beta_3, \beta_3, \ldots, \beta_k$.

Remark. From (30) it is immediately seen that (29) may be written

$$|F_n(x) - \Phi(x)| \leq \frac{c'(\delta, \beta)}{1 + |x|^k} \cdot \frac{1}{n^{\frac{k-2}{2}}}$$

for $|x| \geq \sqrt{(1 + \delta)(k - 2) \log n}$, $c'(\delta, \beta)$ satisfying the same conditions as $c(\delta, \beta)$. We have, however, stated the theorem in the present form for purposes of comparison with the expansion (15).

Proof of Theorem 2. By $c_1, c_2, \ldots$ we denote a sequence of finite positive constants, only depending on δ and the moments β. The quantities a, b, ξ etc. are defined by (17)—(28). Further we put

$$G_n(x) = \Phi(x) + \sum_{\nu=1}^{k-3} \frac{P_\nu(-\Phi)}{n^{\nu/2}} \tag{31}$$

with the c. f.

$$g_n(t) = e^{-\frac{t^2}{2}} + \sum_{\nu=1}^{k-3} \frac{P_\nu(it)}{n^{\nu/2}} e^{-\frac{t^2}{2}}.$$

Our problem is to obtain an estimation of $\Delta(b)$ as a function of n and b. By hypothesis

$$b \geq \sqrt{(1+\delta)(k-2)\log n} \tag{32}$$

and we may without loss of generality suppose n to be so large that

$$b \geq 1. \tag{33}$$

From (32) it is easily seen that there exist two constants c_1 and c_2 such that

$$a - \xi = a(1 - c_1 \delta) \tag{34}$$

and

$$G_n(a) - G_n(y) \leq \frac{c_2}{1+a^k} \cdot \frac{1}{n^{\frac{k-2}{2}}} \tag{35}$$

for $a - \xi \leq y \leq a$. We now choose

$$\varrho(a, n) = \frac{2\, c_2}{1+a^k} \cdot \frac{1}{n^{\frac{k-2}{2}}} \tag{36}$$

and obtain two different cases:

1°. $\Delta(b) \leq \varrho(a, n)$.

Then the theorem is proved.

2°. $\Delta(b) > \varrho(a, n)$.

Then by (35) and (36) the inequality (24) holds. Hence from (26)

$$\frac{\Delta(b)}{2} \int\limits_{-\frac{\lambda c_1 \delta a}{2}}^{\frac{\lambda c_1 \delta a}{2}} Q(y)\, dy - 2\, d(n) \int\limits_{\frac{\lambda c_1 \delta a}{2}}^{\infty} Q(y)\, dy \leq \frac{1}{2\pi} \left| \int\limits_{-\infty}^{\infty} e^{-itx} \frac{g_n(t) - f_n(t)}{-it}\, q\left(\frac{t}{\lambda}\right) dt \right|. \tag{37}$$

Here $d(n)$ obviously is a finite quantity. By Lemma 1, Chap. II, it is possible to choose $Q(x)$ and $q(t)$ such that the conditions (18) are satisfied and at the same time

$$\begin{cases} Q(x) = 0 \text{ for } |x| \geq 1, \\ q(t) \leq \dfrac{c_3}{|t|^{3(k-2)}}, \displaystyle\int\limits_{-\infty}^{\infty} |t|^{3k-7} q(t)\, dt = c_4 < \infty. \end{cases} \tag{38}$$

We now choose $\lambda = \dfrac{2}{c_1 \delta a}$ and obtain from (18), (37) and (38):

$$(39)\qquad \frac{\Delta(b)}{2} \leq \frac{1}{2\pi}\int_{-\infty}^{\infty}\left|\frac{g_n(t)-f_n(t)}{t}\right| q\left(\frac{c_1\,\delta\,a\,t}{2}\right)dt = \varepsilon.$$

Applying Lemma 2 a, Chap. III, and putting

$$(40)\qquad \varepsilon = \int_{-\infty}^{-T_{kn}} + \int_{-T_{kn}}^{T_{kn}} + \int_{T_{kn}}^{\infty} = \varepsilon_1 + \varepsilon_2 + \varepsilon_3,$$

we have by (33) and (38):

$$\varepsilon_2 \leq \frac{c_5}{n^{\frac{k-2}{2}}}\int_{-\infty}^{\infty}(|t|^k + |t|^{3(k-2)})\, e^{-\frac{t^2}{4}} q\left(\frac{c_1\,\delta\,a\,t}{2}\right)\frac{dt}{|t|} \leq$$

$$(41)\qquad \leq \frac{c_5}{n^{\frac{k-2}{2}}}\int_{-\infty}^{\infty}\left[\left(\frac{2}{c_1\,\delta\,a}\right)^k |t|^k + \left(\frac{2}{c_1\,\delta\,a}\right)^{3(k-2)}|t|^{3(k-2)}\right] q(t)\frac{dt}{|t|} \leq \frac{c_6}{a^k}\cdot\frac{1}{n^{\frac{k-2}{2}}}.$$

Moreover, from (38)

$$(42)\qquad \varepsilon_1 + \varepsilon_3 \leq c_7\int_{c_8\sqrt{n}}^{\infty}\left(\frac{2}{c_1\,\delta\,a}\right)^{3(k-2)}\frac{1}{|t|^{3(k-2)}}\frac{dt}{|t|} \leq \frac{c_9}{a^k}\cdot\frac{1}{n^{\frac{k-2}{2}}}.$$

Summing up we obtain

$$(43)\qquad \Delta(b) \leq \frac{c_{10}}{a^k}\cdot\frac{1}{n^{\frac{k-2}{2}}} \leq c_{11}\cdot\varrho(a, n).$$

Thus in either case:

$$\Delta(b) \leq \frac{c(\delta, \beta)}{1 + a^k}\cdot\frac{1}{n^{\frac{k-2}{2}}} \leq \frac{c(\delta, \beta)}{1 + b^k}\cdot\frac{1}{n^{\frac{k-2}{2}}}$$

and the theorem is proved.

Theorem 2 holds for $|x| \geq \sqrt{(1+\delta)(k-2)\log n}$. We shall now state and prove a theorem valid in the remaining interval. By $c_1, c_2, \ldots$ we denote as before finite, positive constants only depending on δ and the absolute moments $\beta_2, \beta_3, \ldots, \beta_k$.

Theorem 3. *Let $X_1, X_2, \ldots, X_n$ be a sequence of independent r.v.'s with the same d.f. $F(x)$, the mean value zero, the dispersion $\sigma \neq 0$ and the finite absolute moment β_k, (k an integer ≥ 3). Then*

$$\left|F_n(x) - \Phi(x) - \sum_{\nu=1}^{k-3} \frac{P_\nu(-\Phi)}{n^{\nu/2}}\right| \leq \frac{c_1(\delta, \beta)}{\sqrt{n}}(1 + |x|^3)\, e^{-\frac{x^2}{2}} + \frac{c_2(\delta, \beta)}{n^{\frac{k-2}{2}}} \tag{44}$$

for an assigned number δ, $(0 < \delta < 1)$, *and*

$$|x| \leq \sqrt{(1 + \delta)(k-2)\log n}. \tag{45}$$

Remark. In view of (45) it is possible to replace (44) by

$$|F_n(x) - \Phi(x)| \leq \frac{c_1'(\delta, \beta)}{\sqrt{n}}(1 + |x|^3)\, e^{-\frac{x^2}{2}} + \frac{c_2(\delta, \beta)}{n^{\frac{k-2}{2}}}$$

$$\text{for } |x| \leq \sqrt{(1 + \delta)(k-2)\log n}.$$

Proof of Theorem 3. The quantities a, b, ξ etc. are defined by (17)—(28), $G_n(x)$ and its transform $g_n(t)$ by (31). The auxiliary functions $Q(x)$ and $q(t)$ satisfy (18), and further, by Lemma 1, Chap. II:

$$\begin{cases} q(t) = 0 \text{ for } |t| \geq 1 \\ Q(x) \leq c_3 \cdot e^{-|x|^{3/4}} \end{cases}. \tag{46}$$

By Theorem 1, Chap. III, we also have

$$|F_n(x) - G_n(x)| \leq d(n) = \frac{c_4}{\sqrt{n}}. \tag{47}$$

According to (45) and (47) it is sufficient to prove the theorem for

$$1 \leq c_5 \leq b \leq a \leq \sqrt{(1 + \delta)(k-2)\log n} \tag{48}$$

where n is so large that the above inequality is satisfied and c_5 is a constant later to be determined. Further we may suppose that $x > 0$ in (44). It is readily observed that there exist two positive constants c_6 and c_7 such that

$$\xi = c_6 \cdot \frac{a^3}{\sqrt{n}} \tag{49}$$

and

$$G_n(a) - G_n(y) \leq \frac{c_7}{\sqrt{n}} a^3 e^{-\frac{a^2}{2}} \text{ for } a - \xi \leq y \leq a. \tag{50}$$

We now assume that

$$\Delta(b) \geq \frac{2\, c_7}{\sqrt{n}} a^3 e^{-\frac{a^2}{2}}, \tag{51}$$

or else the theorem is proved. Then (24) holds and we obtain from (26), (46) and (18)

$$(52)\qquad \frac{\Delta(b)}{2}\int\limits_{-\frac{\lambda}{2}\frac{c_8 a^3}{\sqrt{n}}}^{\frac{\lambda}{2}\frac{c_8 a^3}{\sqrt{n}}} Q(y)\,dy-\frac{2\,c_4}{\sqrt{n}}\int\limits_{\frac{\lambda}{2}\frac{c_8 a^3}{\sqrt{n}}}^{\infty} Q(y)\,dy \le \frac{1}{2\pi}\int\limits_{-\lambda}^{\lambda}\left|\frac{g_n(t)-f_n(t)}{t}\right|dt.$$

Putting

$$(53)\qquad \lambda = T_{kn}$$

and applying Lemma 2 a, Chap. III, to the right-hand side of (52), we obtain in the usual way:

$$(54)\qquad \frac{\Delta(b)}{2}\int\limits_{-c_8 a^3}^{c_8 a^3} Q(y)\,dy-\frac{2\,c_4}{\sqrt{n}}\int\limits_{c_8 a^3}^{\infty} Q(y)\,dy \le \frac{c_9}{n^{\frac{k-2}{2}}}.$$

According to (46)

$$(55)\qquad \int\limits_{c_8 a^3}^{\infty} Q(y)\,dy \le c_{10}\,a^{3/4}\,e^{-c_8^{3/4}\cdot a^{9/4}}.$$

We now choose c_5 in (48) in such a manner that the following two conditions are satisfied:

$$(56)\qquad \begin{cases} \int\limits_{-c_8 c_5^3}^{c_8 c_5^3} Q(y)\,dy \ge \frac{1}{2}; \\ c_{10}\,a^{3/4}\,e^{-c_8^{3/4}\cdot a^{9/4}} \le a^3\,e^{-\frac{a^2}{2}} \text{ for } a \ge c_5. \end{cases}$$

Hence from (54), (55) and (56)

$$(57)\qquad \frac{\Delta(b)}{4}-\frac{2\,c_4}{\sqrt{n}}\,a^3\,e^{-\frac{a^2}{2}} \le \frac{c_9}{n^{\frac{k-2}{2}}}.$$

In view of (57) and the converse of (51) we finally obtain:

$$\Delta(b) \le \operatorname{Max}\left\{\frac{2\,c_7}{\sqrt{n}}\,a^3\,e^{-\frac{a^2}{2}};\ \frac{8\,c_4}{\sqrt{n}}\,a^3\,e^{-\frac{a^2}{2}}+\frac{4\,c_9}{n^{\frac{k-2}{2}}}\right\},$$

which proves the theorem.

It is interesting to compare Theorems 2 and 3, this chapter, with Theorems 1 and 4, Chap. IV. Owing to the discontinuous term in the lattice distribution case it is to be expected that the general remainder term will contain a function of the type $\frac{e^{-\frac{x^2}{2}}}{\sqrt{n}}$, and this is also the case in Theorem 3. As x becomes large the predominance of this term vanishes, and we obtain the expansion (29).

As in (29), it is possible to multiply the term $\frac{c_2}{n^{\frac{k-2}{2}}}$ in (44) by a function of x, tending to zero as $x \to \infty$. In order to do this we may proceed as in the proof of Theorem 3. The only difference is that we have to investigate the dependence on a and n of the integral

$$\int_{-\lambda}^{\lambda} e^{-it\left(a-\frac{1}{3}\frac{c_8 a^3}{\sqrt{n}}\right)} \frac{g_n(t)-f_n(t)}{-it} q\left(\frac{t}{\lambda}\right) dt.$$

This may be done by repeated partial integrations. I confine myself to stating the theorem for $k=3$, Theorem 3 giving nothing new in this case, since now

$$\frac{1}{n^{\frac{k-2}{2}}} = \frac{1}{\sqrt{n}}.$$

Theorem 4. *Under the same conditions as in Theorem 3 with β_3 finite the following inequality holds:*

$$|F_n(x) - \Phi(x)| \leq \frac{c(\sigma, \beta_3)}{\sqrt{n}} \frac{\log(2+|x|)}{1+|x|^3},$$

for all values of x, where $c(\sigma, \beta_3)$ is a finite, positive constant only depending on σ and β_3.

3. **On the Uniform Law of Great Numbers.** Consider a r. v. X with the mean value zero, the dispersion $\sigma \neq 0$ and *the finite fourth moment* β_4. Let us imagine a sequence of independent trials. In the first trial X assumes the value $X_1^{(1)}$ and we put

$$X_{(1)} = X_1^{1)}.$$

In the second trial, consisting of two independent trials, X first assumes the value $X_1^{(2)}$ and then $X_2^{(2)}$. We put

$$X_{(2)} = X_1^{(2)} + X_2^{(2)}.$$

We proceed in the same way: in the nth trial, consisting of n independent trials, X assumes the values $X_1^{(n)}, X_2^{(n)}, \ldots, X_n^{(n)}$, and we put

$$X_{(n)} = X_1^{(n)} + X_2^{(n)} + \cdots + X_n^{(n)}.$$

Then the following theorem holds:

Theorem 5. *If δ is an assigned arbitrary small positive number, the probability of*

$$|X_{(n)}| \leq \sqrt{2(1+\delta)\sigma^2 n \log n}$$

for all $n > n_0$ tends to 1 as $n_0 \to \infty$, while the probability of

$$|X_{(n)}| \leq \sqrt{2(1-\delta)\sigma^2 n \log n}$$

for all $n > n_0$ tends to 0 as $n_0 \to \infty$.

This is a form of the Uniform Law of Great Numbers which was proved by Cramér [4] on the assumption that the *fifth* moment is also finite. Theorem 5 easily follows from Theorems 2 and 3, the proofs of which are the main difficulty. For the method of passing over from these theorems to Theorem 5 the reader is referred to Cramér [4].

The conditions of Theorem 5 are as general as possible in so far as the theorem need not be true if $\beta_\mu < \infty$ for all $\mu < 4$, while $\beta_4 = \infty$. This is seen from examples.

Concluding Notes.

We have hitherto considered different forms of the Central Limit Theorem and have especially studied the remainder term problem. In order to avoid unnecessary complications we have often confined ourselves to the case of equal d. f.'s. It is, however, sometimes possible to escape from this condition; this is especially the case for Theorem 1, Chap. IV, and the theorems of Chap. V. Furthermore, we have supposed the r. v.'s to be mutually independent. Following the method of Bernstein [1] (see also Craig [1]) we may generalize the theorems to hold for a sum of variables dependent in a certain way. There is also a problem of a different kind which we have not touched upon, namely the problem of the convergence of frequency functions to the normal frequency function $\Phi'(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.

Most of the theorems stated are based on Theorem 2, Chap. II, or on the method of proof of this theorem. Part II of this work is devoted to probability distributions in a multi-dimensional space. Now, it has proved impossible to extend this method to the multi-dimensional case; we have been obliged to proceed in another manner. This new method may also be used in proving the theorems of Chapters III—V. Nevertheless, we have hitherto preferred the old method for two reasons: firstly, it is of interest to vary the methods, secondly, the old method is much more simple to use in the one-dimensional case.

PART II.

Probability Distributions in More Than One Dimension.

Chapter VI.

Random Variables in k Dimensions.

We have hitherto solely considered probability distributions of one-dimensional r. v.'s. We now proceed to the case of k-dimensional r. v.'s ($k \geq 2$). By R_k we always denote a k-dimensional euclidean space. The concept of a r. v. X in R_k was defined in § 1, Chap. III, where we found that the probability distribution of X is characterized by the probability function (pr. f.) $P(E)$. We further defined the distribution function of X, a concept extremely useful in the one-dimensional case. In more than one dimension, however, it is preferable to study the pr. f. We therefor start this part of the work by giving an account of the properties of the pr. f. For the proofs of several of the following theorems reference is made to Jessen-Wintner [1] and Cramér [5].

By definition a pr. f. $P(E)$ is a set function, determined for every Borel set E in R_k, and such that

1° $0 \leq P(E) \leq 1$,

2° $P(R_k) = 1$,

3° $P(E)$ is a completely additive set function.

In the following we solely consider Borel sets. By $x=(x_1, x_2, \ldots, x_k)$ we denote a variable point in R_k, while by $f(x)$ we always mean a B-measurable function. The notation for an integral with respect to $P(E)$ will be

$$\int_E f(x)\, d\, P(x),$$

the integral being taken in the Lebesgue-Radon sense (c. f. RADON [1]). By

$$\int_E f(x)\, d\, x$$

we denote the ordinary Lebesgue integral.

A set E is called a *continuity set* of P if $P(E')=P(E'')$ where E' denotes the interior points of E and E'' is the closure of E. There exists an at the most enumerable set of real numbers such that at least those intervals $a_i \leq x_i \leq b_i$, $(i=1, 2, \ldots, k)$, for which the numbers a_i, b_i do not belong to this set are continuity sets of $P(E)$. Such an interval is called a *continuity interval*. By the *point spectrum* $Q(P)$ of P[1] we understand the set of those points x for which $P(x)>0$. The point spectrum Q is at the most enumerable. It is often convenient to represent a probability distribution by a positive mass distribution of total amount 1, dispersed all over the space so that every set E is allotted the mass $P(E)$. Hence, in the following, we often speak of the *probability mass*. The point spectrum, for instance, is the set of points, each of which has a positive mass.

A pr. f. P is called *continuous* or *discontinuous* according to whether $Q(P)$ is empty or not. According to RADON [1] every pr. f. P can be written as a sum of three components

$$P=a_1 P_1+a_2 P_2+a_3 P_3, \tag{1}$$

where P_1, P_2 and P_3 are pr. f.'s and a_1, a_2 and a_3 non-negative numbers with the sum 1. Here

P_1 is *absolutely continuous*, i. e.

$$P_1(E)=\int_E D(x)\, d\, x$$

where $D(x)$ is a non-negative point function, the density function,

P_2 is *singular*, i. e. continuous and such that there exists a set E of measure zero, but yet $P_2(E)=1$,

P_3 is *purely discontinuous*, i. e. $P_3(Q(P_3))=1$.

[1] Or perhaps rather the point spectrum of the characteristic function of P.

The *convolution* (»Faltung») P of two pr. f.'s P_1 and P_2 is a very important operation. It is defined by

$$P(E) = P_1 * P_2 = \int_{R_k} P_1(E - x)\, d\, P_2(x)$$

where $E - x$ denotes the set obtained by E through the translation $-x$. The set function $P(E)$ is also a pr. f. and

$$P = P_1 * P_2 = P_2 * P_1.$$

The vectorial sum $A + B$ of two point sets A and B is defined as those points in R_k which may be represented in at least one way as the vector sum $a + b$, where a and b are points of A and B respectively. If either of A and B is empty, $A + B$ is by definition also empty. If Q_1, Q_2 and Q are the point spectra of P_1, P_2 and $P = P_1 * P_2$ respectively, then

$$Q = Q_1 + Q_2.$$

The concept of convolution of two functions is immediately extended to the convolution of n functions.

If $\{P_n(E)\}$ is a sequence of pr. f.'s and $P(E)$ is another pr. f. and if

$$\lim_{n \to \infty} P_n(I) = P(I)$$

for every continuity interval I of $P(E)$, then we say that $P_n(E)$ *converges* to $P(E)$.

The *characteristic function* (c. f.) $f(t_1, t_2, \ldots, t_k)$ of $P(E)$ is defined as the Fourier-Radon transform of $P(E)$:

$$(2) \qquad f(t_1, t_2, \ldots, t_k) = \int_{R_k} e^{i(t_1 x_1 + t_2 x_2 + \cdots + t_k x_k)}\, d\, P(x).$$

We always assume that $t_1, t_2, \ldots, t_k$ are real numbers. Then by $t = (t_1, \ldots, t_k)$ and $x = (x_1, \ldots, x_k)$ we may denote two vectors in R_k with the origin at $0 = (0, 0, \ldots, 0)$ and the components $t_1, t_2, \ldots, t_k$ and $x_1, x_2, \ldots, x_k$ respectively. By $|t|$ and $|x|$ we denote the lengths of the vectors t and x. Then the expression $t_1 x_1 + t_2 x_2 + \cdots + t_k x_k$ is the scalar product $t x$ of t and x. Hence we may write

$$(2') \qquad f(t_1, t_2, \ldots, t_k) = f(t) = \int_{R_k} e^{itx}\, d\, P(x),$$

and we often use this notation when there is no danger of error. The function $f(t)$ is uniformly bounded and continuous:

$$|f(t)| \leq f(0) = 1.$$

According to (1) we obtain

$$f(t) = a_1 f_1(t) + a_2 f_2(t) + a_3 f_3(t), \tag{3}$$

where the functions $f_i(t)$ also are c. f.'s, $(i = 1, 2, 3)$. By the generalized Riemann-Lebesgue theorem it follows:

$$\lim_{|t|\to\infty} |f_1(t)| = 0. \tag{4}$$

From (3) and (4) it results:

If $P(E)$ has an absolutely continuous component, then

$$\overline{\lim_{|t|\to\infty}} |f(t)| < 1.$$

According to (2) the c. f. $f(t)$ is determined by $P(E)$. Conversely, $P(E)$ is determined by the knowledge of $f(t)$. This is a consequence of the following well-known *inversion theorem.*

If the k-dimensional interval J, defined by $x_i \le \xi_i \le x_i + h_i$, $(i = 1, 2, \ldots, k)$, is a continuity interval of the pr. f. $P(E)$ with the c. f. $f(t)$, then

$$P(J) = \lim_{T\to\infty} \frac{1}{(2\pi)^k} \int_{-T}^{T} \cdots \int_{-T}^{T} \frac{1 - e^{-it_1 h_1}}{i t_1} \cdot \ldots \cdot \frac{1 - e^{-it_k h_k}}{i t_k} \cdot e^{-i(x_1 t_1 + x_2 t_2 + \cdots + x_k t_k)} f(t_1, \ldots, t_k)\, dt_1 \ldots dt_k. \tag{5}$$

By the inversion formula $P(E)$ is determined for all continuity intervals and hence for all Borel sets. In this work we shall mainly consider circles, spheres and hyper spheres instead of intervals. Then it may be of some interest to obtain the probability mass of such a region expressed as a functional of $f(t)$.

Theorem 1. *If S, $\left(\sum_{\nu=1}^{k} (x_\nu - \xi_\nu)^2 \le R^2\right)$, is a sphere in R_k with the centre $\xi = (\xi_1, \xi_2, \ldots, \xi_k)$ and the radius R and if S is a continuity set of the pr. f. $P(E)$ with the c. f. $f(t)$, then*

$$P(S) = \lim_{a\to\infty} \left(\frac{R}{2\pi}\right)^{k/2} \int \cdots \int_{|t| \le a} e^{-i\xi t} \frac{J_{k/2}(R|t|)}{|t|^{k/2}} f(t)\, dt_1 \ldots dt_k, \tag{6}$$

where $J_{k/2}(z)$ is the Bessel function of order $k/2$ and the vector notation is used.

Proof of Theorem 2. We start from the right-hand side $P_a(S)$ of (6) with a finite value of a, and from (2) we obtain after some transformations and easy evaluations of integrals:

$$P_a(S) = \int_{R_k} K_a(u)\, d\, P(x),$$

where

$$u = \sqrt{\sum_{\nu=1}^{k} (x_\nu - \xi_\nu)^2}$$

and

$$K_a(u) = \frac{R^{k/2}}{u^{\frac{k-2}{2}}} \int_0^a J_{k/2}(R\,s)\, J_{\frac{k-2}{2}}(u\,s)\, d\,s.$$

By a well-known formula[1] we have

$$\lim_{a \to \infty} K_a(u) = \begin{cases} 1 \text{ if } u < R \\ \frac{1}{2} \text{ if } u = R\,. \\ 0 \text{ if } u > R \end{cases}$$

The remainder of the proof is immediately clear.

If the point spectrum $Q(P)$ of the pr. f. $P(E)$ is not empty, it is sometimes of interest to express the probability mass at a point ξ as a functional of $f(t)$. Let D be an assigned k-dimensional parallelogram of positive volume with its centre at 0 and let D_T denote that parallelogram which is obtained from D by the magnification to the scale $T : 1$. Then the following theorem holds.

If $P(E)$ is a pr. f. with the c. f. $f(t)$ and $\xi = (\xi_1, \xi_2, \ldots, \xi_k)$ is a point in R_k, which is to be understood as the Borel set consisting of the point ξ alone, then

$$(7) \qquad P(\xi) = \lim_{T \to \infty} \frac{1}{D_T} \int_{D_T} e^{-i\xi t} f(t)\, d t_1\, d t_2 \ldots d t_k.$$

Here D_T denotes both the volume of the parallelogram and the region of integration, and the vector notation is used.

The proof is analogous to that of formula (6), Chap. I.

The connection between the c. f.'s of a sequence of pr. f.'s and the c. f. of their convolution is expressed by the *convolution theorem.*

If $f_1(t), f_2(t), \ldots, f_n(t)$ are the c. f.'s of the pr. f.'s $P_1, P_2, \ldots, P_n$ respectively and $f(t)$ is the c. f. of the convolution

---

[1] Watson [1], p. 406.

$$P = P_1 * P_2 * \cdots * P_n,$$

then

$$f(t) = \prod_{\nu=1}^{n} f_\nu(t).$$

Finally let us consider the *moments* of r. v. $X = (X_1, X_2, \ldots, X_k)$ in R_k with the pr. f. $P(E)$ and the c. f. $f(t)$. We use the symbolic notation

$$\alpha_1^{\nu_1} \alpha_2^{\nu_2} \ldots \alpha_k^{\nu_k} = \int_{R_k} x_1^{\nu_1} x_2^{\nu_2} \ldots x_k^{\nu_k} \, d\, P(x), \tag{8}$$

$$\beta_1^{\nu_1} \beta_2^{\nu_2} \ldots \beta_k^{\nu_k} = \int_{R_k} |x_1|^{\nu_1} |x_2|^{\nu_2} \ldots |x_k|^{\nu_k} \, d\, P(x), \tag{9}$$

where $\nu_1, \nu_2, \ldots, \nu_k$ are non-negative integers. Here $\alpha_1^{\nu_1} \alpha_2^{\nu_2} \ldots \alpha_k^{\nu_k}$ is interpreted as a symbolic product, so that the following relation holds:

$$(\alpha_1 t_1 + \alpha_2 t_2 + \cdots + \alpha_k t_k)^\nu = \int_{R_k} (t_1 x_1 + t_2 x_2 + \cdots + t_k x_k)^\nu \, d\, P(x), \tag{10}$$

(ν a positive integer). The same applies to (9). If r is a positive integer and the absolute moments (9) are finite for $\nu_1 + \nu_2 + \cdots + \nu_k \leq r$, then we have from (2), (10) and the expansion of e^{itx} in series:

$$f(t) = 1 + \sum_{\nu=1}^{r} \frac{i^\nu}{\nu!} (\alpha_1 t_1 + \alpha_2 t_2 + \cdots + \alpha_k t_k)^\nu + o(|t|^r)$$

for small values of $|t|$. We also observe that if all β_i^r, $(i = 1, 2, \ldots, k)$, are finite, then all moments $\alpha_1^{\nu_1} \alpha_2^{\nu_2} \ldots \alpha_k^{\nu_k}$ and $\beta_1^{\nu_1} \beta_2^{\nu_2} \ldots \beta_k^{\nu_k}$ with $\nu_1 + \nu_2 + \cdots + \nu_k \leq r$ are finite. This follows from the inequality

$$|\alpha_1^{\nu_1} \alpha_2^{\nu_2} \ldots \alpha_k^{\nu_k}| \leq \beta_1^{\nu_1} \beta_2^{\nu_2} \ldots \beta_k^{\nu_k} \leq \{\beta_1^r\}^{\frac{\nu_1}{r}} \cdot \{\beta_2^r\}^{\frac{\nu_2}{r}} \cdot \cdots \cdot \{\beta_k^r\}^{\frac{\nu_k}{r}}, \tag{11}$$

an immediate consequence of the Hölder inequality.

A special importance is attached to the first moments or mean values m_i and the quantities μ_{ij} defined by

$$m_i = \int_{R_k} x_i \, d\, P(x), \qquad (i = 1, 2, \ldots, k), \tag{12}$$

$$\mu_{ij} = \int_{R_k} (x_i - m_i)(x_j - m_j) \, d\, P(x), \qquad (i, j = 1, 2, \ldots, k). \tag{13}$$

We call the quantities μ_{ij} the *translated second order moments*. We further put

$$\sigma_i^2 = \mu_{ii} = \int\limits_{R_k} (x_i - m_i)^2 \, dP(x) \tag{14}$$

and

$$r_{ij} = \frac{\mu_{ij}}{\sigma_i \sigma_j}. \tag{15}$$

The quantity m_i is the *mean value*, the quantity σ_i the *dispersion* of the component X_i of X. The quantity r_{ij} is called the *correlation coefficient* between the components X_i and X_j. It satisfies the inequality $-1 \leq r_{ij} \leq 1$. If $i \neq j$ we call the μ_{ij}'s the *mixed translated second order moments.*

Consider the quadratic form

$$\sum_{i,j=1}^{k} \mu_{ij} u_i u_j = \int\limits_{R_k} \left(\sum_{i=1}^{k} u_i (x_i - m_i) \right)^2 dP(x) \tag{16}$$

with the determinant

$$\Delta = \| \mu_{ij} \|. \tag{17}$$

Obviously the form (16) is definite positive or semi-definite according as $\Delta > 0$ or $\Delta = 0$. If $\Delta = 0$, it is easily found by the theory of quadratic forms that the probability mass is concentrated to a sub-space of R_k. Then the problem is reduced to the study of a r. v. in R_{k_1} with $k_1 < k$. Thus we may neglect the case $\Delta = 0$ and always assume the form (16) to be positive definite.

The subsequent investigations are formally simplified by a certain transformation.

Lemma 1. *If $X = (X_1, X_2, \ldots, X_k)$ is a r.v. in R_k with the mean values $m_i = 0$ and the finite second order moments μ_{ij} and $\Delta = \| \mu_{ij} \| > 0$, then by a linear, non-singular real transformation*

$$Y_i = a_{1i} X_1 + a_{2i} X_2 + \cdots + a_{ki} X_k, \qquad (i = 1, 2, \ldots, k),$$

it is possible to obtain a new r.v. $Y = (Y_1, Y_2, \ldots, Y_k)$ such that

$$\begin{cases} 1^\circ \ \text{the mean value of } Y_i \text{ is zero}, \ (i = 1, 2, \ldots, k), \\ 2^\circ \ \text{the dispersion of } Y_i, \ (i = 1, 2, \ldots, k), \text{ is equal to } 1, \\ 3^\circ \ \text{the mixed moments of the second order are zero.} \end{cases} \tag{18}$$

Further

$$\Big| \| a_{ij} \| \Big| = \frac{1}{\sqrt{\Delta}}.$$

If Δ_{ij} denotes the algebraic complement of Δ with respect to μ_{ij}, then

$$\sum_{i=1}^{k} Y_i^2 = \sum_{i,j=1}^{k} \frac{\Delta_{ij}}{\Delta} X_i X_j. \tag{19}$$

Proof of Lemma 1. Consider the definite positive quadratic form

$$\sum_{i,j=1}^{k} \mu_{ij} u_i u_j = \int\limits_{R_k} \left(\sum_{i=1}^{k} u_i x_i\right)^2 dP(x). \tag{20}$$

Since $\Delta > 0$ there exists, according to the theory of quadratic forms, a non-singular real linear transformation

$$u_i = a_{i1} u_1' + a_{i2} u_2' + \cdots + a_{ik} u_k', \qquad (i = 1, 2, \ldots, k), \tag{21}$$

such that

$$\sum_{i,j=1}^{k} \mu_{ij} u_i u_j = \sum_{i=1}^{k} (u_i')^2. \tag{22}$$

By a well-known theorem $\left| \| a_{ij} \| \right| = \frac{1}{\sqrt{\Delta}}$. Substituting (21) in (20) and introducing the variables

$$Y_i = a_{1i} X_1 + a_{2i} X_2 + \cdots + a_{ki} X_k, \qquad (i = 1, 2, \ldots, k), \tag{23}$$

we obtain from (22)

$$\sum_{i=1}^{k} (u_i')^2 = \int\limits_{R_k} \left(\sum_{i=1}^{k} u_i' y_i\right)^2 dP_1(y), \tag{24}$$

where $P_1(E)$ is the pr. f. of the r. v. $Y = (Y_1, Y_2, \ldots, Y_k)$. From (23) and (24) it immediately follows that the variable Y satisfies the conditions (18). Only the relation (19) remains to be proved. From (23) we obtain

$$\sum_{i=1}^{k} Y_i^2 = \sum_{i,j=1}^{k} b_{ij} X_i X_j, \tag{25}$$

where the coefficients b_{ij} are to be determined. We introduce the matrices

$$\begin{aligned} \mathfrak{D} &= \begin{pmatrix} \mu_{11} & \mu_{12} & \ldots & \mu_{1k} \\ \cdot & \cdot & \cdot & \cdot \\ \mu_{k1} & \mu_{k2} & \ldots & \mu_{kk} \end{pmatrix}, & \mathfrak{A} &= \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1k} \\ \cdot & \cdot & \cdot & \cdot \\ a_{k1} & a_{k2} & \ldots & a_{kk} \end{pmatrix}, \\ \mathfrak{B} &= \begin{pmatrix} b_{11} & b_{12} & \ldots & b_{1k} \\ \cdot & \cdot & \cdot & \cdot \\ b_{k1} & b_{k2} & \ldots & b_{kk} \end{pmatrix}, & \mathfrak{E} &= \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ \cdot & \cdot & \cdot & \ldots & \cdot \\ 0 & 0 & 0 & \ldots & 1 \end{pmatrix}. \end{aligned} \tag{26}$$

By $\mathfrak{C}^*$ we denote the transpose of a matrix $\mathfrak{C}$, by $\mathfrak{C}^{-1}$ the inverse of $\mathfrak{C}$. Now (22) and (25) may be written

$$\mathfrak{A}\,\mathfrak{D}\,\mathfrak{A} = \mathfrak{E} \tag{27}$$

and

$$\mathfrak{A}^*\,\mathfrak{E}\,\mathfrak{A}^* = \mathfrak{B} \tag{28}$$

respectively. Hence from (28) $\mathfrak{B}^* = \mathfrak{A}\,\mathfrak{A} = \mathfrak{A}\,\mathfrak{A}\,\mathfrak{D}\,\mathfrak{D}^{-1}$. From (27) we obtain $\mathfrak{A}\,\mathfrak{D} = \mathfrak{A}^{-1}$ and hence $\mathfrak{B}^* = \mathfrak{A}\,\mathfrak{A}^{-1}\,\mathfrak{D}^{-1} = \mathfrak{D}^{-1}$, and the lemma is proved.

In the following chapters we shall mainly occupy ourselves with the addition of independent r. v.'s. Let $X^{(1)} = (X_1^{(1)}, X_2^{(1)}, \ldots, X_k^{(1)})$ and $X^{(2)} = (X_1^{(2)}, X_2^{(2)}, \ldots, X_k^{(2)})$ be two r. v.'s in R_k. By the sum $X = X^{(1)} + X^{(2)}$ we understand the variable

$$X = (X_1^{(1)} + X_1^{(2)},\ X_2^{(1)} + X_2^{(2)}, \ldots,\ X_k^{(1)} + X_k^{(2)}).$$

The following well-known *addition theorem* holds:

If $X^{(1)}$ and $X^{(2)}$ are two independent r. v.'s in R_k with the pr. f.'s $P_1(E)$ and $P_2(E)$ respectively and the corresponding c. f.'s $f_1(t)$ and $f_2(t)$, then the sum $X = X^{(1)} + X^{(2)}$ has the pr. f.

$$P(E) = P_1 * P_2 = \int_{R_k} P_1(E - x)\, d P_2(x)$$

and the c. f.

$$f(t) = f_1(t) \cdot f_2(t).$$

The generalization to a sum of n independent r. v.'s is immediate. If the variables have the mean values $m_i^{(1)}$ and $m_i^{(2)}$ and the dispersions $\sigma_i^{(1)}$ and $\sigma_i^{(2)}$, $(i = 1, 2, \ldots, k)$, we also observe:

$$m_i = m_i^{(1)} + m_i^{(2)} \tag{29}$$

and

$$\sigma_i^2 = (\sigma_i^{(1)})^2 + (\sigma_i^{(2)})^2, \tag{30}$$

where m_i and σ_i are the mean values and the dispersions of X.

In the one-dimensional case the normal distribution function $\Phi\left(\frac{x-m}{\sigma}\right)$ with the mean value m and the dispersion σ has the c. f. $e^{imt - \frac{1}{2}\sigma^2 t^2}$. In the multi-dimensional case the normal distribution is defined in the following way.

A r. v. $X = (X_1, X_2, \ldots, X_k)$ in R_k with the mean values m_r, $(r = 1, 2, \ldots, k)$, and the translated second order moments μ_{rs}, $(r, s = 1, 2, \ldots, k)$, is normally distributed, if the pr. f. is absolutely continuous with the density function

$$D(x) = \frac{1}{(2\pi)^{k/2}\sqrt{\Delta}} e^{-\frac{1}{2}q(x_1, x_2, \ldots, x_k)}, \tag{31}$$

where

$$q(x_1, x_2, \ldots, x_k) = \sum_{r,s=1}^{k} \frac{\Delta_{rs}}{\Delta}(x_r - m_r)(x_s - m_s)$$

and $\Delta = \| \mu_{rs} \|$ *is the determinant of the quadratic form* $\sum_{r,s=1}^{k} \mu_{rs} t_r t_s$. *Here* Δ_{rs} *is the algebraic complement of* μ_{rs} *with respect to* Δ *and* Δ *is supposed to be* > 0. *The c.f.* $f(t)$ *is expressed by*

$$f(t) = e^{i\sum_{r=1}^{k} m_r t_r - \frac{1}{2}\sum_{r,s=1}^{k} \mu_{rs} t_r t_s} \tag{32}$$

If $\Delta = 0$, $f(t)$ may be considered as the c. f. of an improper normal distribution with the probability mass concentrated to a sub-space of R_k. We neglect this case.

If a r. v. $X = (X_1, X_2, \ldots, X_k)$ is normally distributed according to (31) let us consider the variable $\bar{X} = (X_1 - m_1, X_2 - m_2, \ldots, X_k - m_k)$. The variable $\bar{X}$ is also normally distributed with the mean values zero and the ordinary second moments μ_{ij}. By Lemma 1 it is possible to form a r. v. $Y = (Y_1, Y_2, \ldots, Y_k)$ with the mean values zero, the dispersions 1 and the mixed second order moments zero. It is easily seen that Y has the frequency function

$$\varphi(y_1, y_2, \ldots, y_k) = \frac{1}{(2\pi)^{k/2}} e^{-\frac{1}{2}(y_1^2 + y_2^2 + \cdots + y_k^2)}. \tag{33}$$

The probability $\psi(a, k)$ of Y being situated within a sphere with its centre at $0 = (0, 0, \ldots, 0)$ and the radius a, is expressed by

$$\psi(a, k) = \int \cdots \int_{y_1^2 + \cdots + y_k^2 \leq a^2} \frac{1}{(2\pi)^{k/2}} e^{-\frac{1}{2}(y_1^2 + \cdots + y_k^2)} dy_1 \ldots dy_k. \tag{34}$$

It is easily found that

$$\psi(a, k) = \begin{cases} 1 - \sqrt{\frac{2}{\pi}} \int_a^\infty e^{-\frac{1}{2}r^2} dr - \sqrt{\frac{2}{\pi}} \left[\frac{a}{1} + \frac{a^3}{1 \cdot 3} + \cdots + \frac{a^{k-2}}{1 \cdot 3 \cdot 5 \cdots (k-2)} \right] e^{-\frac{a^2}{2}} & \text{for } k \text{ odd,} \\ 1 - \left[1 + \frac{a^2}{2} + \cdots + \frac{a^{k-2}}{2 \cdot 4 \cdots (k-2)} \right] e^{-\frac{a^2}{2}} & \text{for } k \text{ even.} \end{cases} \tag{35}$$

Chapter VII.

On the Central Limit Theorem in R_k. Estimation of the Remainder Term.

Introduction. Let us consider a sequence of independent r. v.'s $X^{(1)}$, $X^{(2)}$, $\ldots$, $X^{(n)}$ in R_k, $(k \geq 2)$. As in the one-dimensional case it is of great importance to the theory of probability and its applications to study the distribution of the sum of a large number of such variables. If $'P_n(E)$ denotes the pr. f. of the variable

$$^{(n)}X = \frac{X^{(1)} + X^{(2)} + \cdots + X^{(n)}}{\sqrt{n}},$$

then under certain conditions[1] $P_n(E)$ converges to the normal pr. f. as n tends to infinity. This is the *Central Limit Theorem* in R_k. How large is the error involved when the process ceases at a finite value of n? The only result hitherto obtained in this direction is due to Joubavsky [1], who, however, only gives a rough estimation of the error term.[2]

Being mainly interested in principles we confine ourselves to the case of equal distributions; there is, however, no difficulty in generalizing the subsequent theorems. Thus, consider a sequence of independent r. v.'s

$$X^{(1)}, X^{(2)}, \ldots, X^{(n)} \tag{1}$$

in R_k with the same pr. f. Let an arbitrary variable $X = (X_1, X_2, \ldots, X_k)$ of the sequence have the properties:

$$\begin{cases} 1^\circ \text{ the mean value of every component is equal to zero;} \\ 2^\circ \text{ the determinant } \Delta = \|\mu_{ij}\| > 0 \text{ where } \mu_{ij} \text{ are the moments of the second order;} \\ 3^\circ \text{ the fourth moments are all finite.} \end{cases} \tag{2}$$

Our problem is to study the distribution of the variable

$$^{(n)}X = (^{(n)}X_1, {}^{(n)}X_2, \ldots, {}^{(n)}X_k) = \frac{X^{(1)} + X^{(2)} + \cdots + X^{(n)}}{\sqrt{n}}. \tag{3}$$

---

[1] Bernstein [1], Cramér [5], p. 113, Joubavsky [1] and others.

[2] Considering the probability of $^{(n)}X$ belonging to a k-dimensional interval and supposing that the absolute moments of order $2+\delta$, $(0<\delta<1)$, are finite, he obtains a remainder term $=$ $= O\left(\frac{1}{n^{\frac{\delta}{2(k+3+\delta)}}}\right)$.

In order to facilitate the calculation we make a transformation according to Lemma 1, Chap. VI, of each variable X in the sequence (1), obtaining hereby a new r. v. $Y=(Y_1, Y_2, \ldots, Y_k)$ such that

(4) $\begin{cases} 1^\circ & \text{the mean value of each component is equal to zero,} \\ 2^\circ & \text{the dispersion of each component is equal to 1,} \\ 3^\circ & \text{the mixed moments of the second order are equal to zero,} \\ 4^\circ & \text{the fourth moments are finite.} \end{cases}$

Now form the variable

$$ {}^{(n)}Y = ({}^{(n)}Y_1, {}^{(n)}Y_2, \ldots, {}^{(n)}Y_k) = \frac{Y^{(1)} + Y^{(2)} + \cdots + Y^{(n)}}{\sqrt{n}}, \tag{5} $$

where by Lemma 1, Chap. VI,

$$ {}^{(n)}Y_i = a_{1i}\, {}^{(n)}X_1 + a_{2i}\, {}^{(n)}X_2 + \cdots + a_{ki}\, {}^{(n)}X_k, \qquad (i = 1, 2, \ldots, k), \tag{6} $$

and

$$ \sum_{i=1}^{k} ({}^{(n)}Y_i)^2 = \sum_{i,j=1}^{k} \frac{\Delta_{ij}}{\Delta}\, {}^{(n)}X_i\, {}^{(n)}X_j, \tag{7} $$

Δ_{ij} being the algebraic complement of μ_{ij} with respect to Δ. From (6) it follows that the probability distribution of ${}^{(n)}X$ is easily obtained if it is known for the variable ${}^{(n)}Y$. Hence we may confine ourselves to the case that the conditions (2) are identical with the conditions (4).

N o t a t i o n s. By $P_n(E)$ we denote the pr. f. of ${}^{(n)}Y$, by $\mu_n(a)$ the probability of ${}^{(n)}Y$ lying within the sphere

$$ S: \sum_{i=1}^{k} y_i^2 \leq a^2. \tag{8} $$

Further $\varphi(y) = \varphi(y_1, y_2, \ldots, y_k)$ denotes the density function of the normalized normal distribution:

$$ \varphi(y_1, y_2, \ldots, y_k) = \frac{1}{(2\pi)^{k/2}} e^{-\frac{1}{2}(y_1^2 + y_2^2 + \cdots + y_k^2)}. $$

The normal pr. f. $\Pi(E)$ is expressed by

$$ \Pi(E) = \int_E \varphi(y)\, dy. $$

By (34), Chap. VI,

$$ \Pi(S) = \psi(a, k) = \int_S \varphi(y)\, dy. $$

We wish to estimate the quantity $|P_n(E) - \Pi(E)|$ as a function of n. In order to do this we have to impose certain conditions on E. With regard to the applications it is natural to let E be a k-dimensional interval or a hyper sphere with its centre at $0 = (0, \ldots, 0)$, this case giving particularly interesting results with application to the χ^2 method. The main problem of this chapter is the estimation of $|P_n(S) - \Pi(S)| = |\mu_n(a) - \psi(a, k)|$. The following theorem holds:

Theorem 1. *Let* $Y^{(1)}, Y^{(2)}, \ldots, Y^{(n)}$ *be a sequence of independent r.v.'s in* R_k, $(k \geq 2)$, *with the same pr. f.* $P(E)$ *and c. f.* $f(t)$. *Further let an arbitrary variable* $Y = (Y_1, Y_2, \ldots, Y_k)$ *of the sequence satisfy the conditions:*

1° *the mean value of every component* Y_i *is equal to zero;*
2° *the dispersion of every component* Y_i *is equal to* 1;
3° *the mixed moments of the second order are equal to zero;*
4° *the fourth moments* β_i^4 *are finite.*

$(i = 1, 2, \ldots, k.)$

If

$$^{(n)}Y = (^{(n)}Y_1, {}^{(n)}Y_2, \ldots, {}^{(n)}Y_k) = \frac{Y^{(1)} + Y^{(2)} + \cdots + Y^{(n)}}{\sqrt{n}}$$

and $\mu_n(a)$ *denotes the probability of* $\sum_{i=1}^{k} (^{(n)}Y_i)^2 \leq a^2$, *then*

$$|\mu_n(a) - \psi(a, k)| \leq c(k) \cdot \frac{\beta_4^{3/2}}{n^{\frac{k}{k+1}}} \tag{9}$$

for all a, *where* $c(k)$ *is a finite, positive constant only depending on* k, $\beta_4 = \sum_{i=1}^{k} \beta_i^4$ *and*

$$\psi(a, k) = \frac{1}{(2\pi)^{k/2}} \int\limits_{y_1^2 + \cdots + y_k^2 \leq a^2} e^{-\frac{1}{2}(y_1^2 + y_2^2 + \cdots + y_k^2)} \, dy_1 \ldots dy_k.$$

Corollary. Let $X^{(1)}, X^{(2)}, \ldots, X^{(n)}$ be a sequence of independent r. v.'s satisfying the conditions (2) and let $^{(n)}X$ be defined by (3). Then from (7) the function $\mu_n(a)$ in Theorem 1 is also the probability of

$$\sum_{i,j=1}^{k} \frac{\Delta_{ij}}{\Delta} {}^{(n)}X_i \, {}^{(n)}X_j \leq a^2$$

and the inequality (9) still holds. We further observe that $\psi(a, k)$ also may be expressed by

$$\frac{1}{(2\pi)^{k/2}\sqrt{\Delta}} \int\limits_{\sum_{i,j=1}^{k} \frac{\Delta_{ij}}{\Delta} x_i x_j \leq a^2} e^{-\frac{1}{2}\sum_{i,j=1}^{k} \frac{\Delta_{ij}}{\Delta} x_i x_j} \, dx_1 \ldots dx_k. \tag{10}$$

We briefly sketch the proof of Theorem 1. It is based on a convolution method which is, however, not the same as in the one-dimensional case. As before we need some lemmata concerning the behaviour of the c. f. $f_n(t)$ of ${}^{(n)}Y$ about $t = 0$. These are given in § 2. The essential point of the proof is, however, an investigation of the value distribution of the modulus of the c. f. $f(t)$. This question is studied in § 1. In § 3 we form an auxiliary function. After these preliminaries the proof of Theorem 1 follows, (§ 4). In § 5 we apply the theorem to the χ^2 method. In the next and last chapter we study the k-dimensional lattice distribution, especially its connection with the general lattice point problem for ellipsoids.

1. On the approach towards 1 of the modulus of a characteristic function.

Consider the pr. f. $P(E)$ of the r. v. $X = (X_1, X_2, \ldots, X_k)$ in R_k, $(k \geq 2)$. Throughout this section we assume that the following conditions hold:

$$\begin{cases} \int\limits_{R_k} x_\nu \, dP(x) = 0; \quad \int\limits_{R_k} x_\nu^2 \, dP(x) = 1; \quad \beta_\nu^3 = \int\limits_{R_k} |x_\nu|^3 \, dP(x) < \infty, \quad (\nu = 1, 2, \ldots, k); \\ \int\limits_{R_k} x_r x_s \, dP(x) = 0 \text{ for } r, s = 1, 2, \ldots, k \text{ and } r \neq s. \end{cases} \tag{11}$$

If a variable satisfies the two first and the last condition of (11) we call it *normalized*. By β_3 we denote the quantity

$$\beta_3 = \sum_{\nu=1}^{k} \beta_\nu^3. \tag{12}$$

Consider the c. f.

$$f(t) = f(t_1, t_2, \ldots, t_k) = \int\limits_{R_k} e^{i(t_1 x_1 + t_2 x_2 + \cdots + t_k x_k)} \, dP(x). \tag{13}$$

The problem of this section is to study the approach of $|f(t)|$ towards 1. It is, however, easier to treat the function

$$g(t) = g(t_1, t_2, \ldots, t_k) = |f(t)|^2. \tag{14}$$

Now

$$\overline{f(t)} = \int_{R_k} e^{-i(t_1\xi_1 + t_2\xi_2 + \cdots + t_k\xi_k)}\, d\,P(\xi)$$

and hence

$$(15)\qquad g(t) = f(t)\overline{f(t)} = \int_{R_k}\int_{R_k} \cos\,[t_1(x_1-\xi_1) + \cdots + t_k(x_k-\xi_k)]\, d\,P(x)\, d\,P(\xi).$$

By S_ϱ we denote an arbitrary k-dimensional sphere of radius ϱ, and by $m s_\varrho(A)$ the measure of those t-points belonging to S_ϱ for which a certain property A is satisfied.

Theorem 2. *If the pr. f. $P(E)$ is normalized, $\beta_3 < \infty$ and*

$$(16)\qquad \varrho = \frac{1}{6\,(1+\sqrt[3]{2})\,k^{1/2}\beta_3},$$

then

$$(17)\qquad m s_\varrho\,\{g(t) \geq 1-\varepsilon\} \leq \frac{1}{\Gamma\left(1+\frac{k}{2}\right)}\{3\,\pi\,(1+\sqrt[3]{2})\,\varepsilon\}^{k/2},$$

where ε is an arbitrary number such that $0 < \varepsilon \leq 1$.

We begin by proving the following lemma.

Lemma 1. *If $a = (a_1, a_2, \ldots, a_k)$ is a point in R_k and*

$$r = \sqrt{\sum_{\nu=1}^{k}(t_\nu - a_\nu)^2},$$

then under the conditions of Theorem 2

$$(18)\qquad g(t) \leq g(a) + (t_1-a_1)\left(\frac{\partial g}{\partial t_1}\right)_a + \cdots + (t_k-a_k)\left(\frac{\partial g}{\partial t_k}\right)_a - \\ - r^2\{1 - 6\,k^{1/3}\beta_3^{2/3}(1-g(a))^{1/3}\} + \tfrac{4}{3}k^{1/2}\beta_3 r^3.$$

Proof of Lemma 1. Expanding $\cos\,[t_1(x_1-\xi_1) + \cdots + t_k(x_k-\xi_k)]$ about $t=a$, we obtain from (15):

$$g(t) = g(a) + (t_1-a_1)\left(\frac{\partial g}{\partial t_1}\right)_a + \cdots + (t_k-a_k)\left(\frac{\partial g}{\partial t_k}\right)_a - \\ - \tfrac{1}{2}\int_{R_k}\int_{R_k}[(t_1-a_1)(x_1-\xi_1) + \cdots + (t_k-a_k)(x_k-\xi_k)]^2 \cos\,[a_1(x_1-\xi_1) + \\ + \cdots + a_k(x_k-\xi_k)]\, d\,P(x)\, d\,P(\xi) + \\ + \frac{\theta}{6}\int_{R_k}\int_{R_k}|(t_1-a_1)(x_1-\xi_1) + \cdots + (t_k-a_k)(x_k-\xi_k)|^3\, d\,P(x)\, d\,P(\xi),$$

where $0 < |\theta| < 1$. By (11)

$$\int_{R_k}\int_{R_k} [(t_1 - a_1)(x_1 - \xi_1) + \cdots + (t_k - a_k)(x_k - \xi_k)]^2 \, dP(x)\, dP(\xi) = 2r^2,$$

and hence

$$(19) \qquad g(t) = g(a) + (t_1 - a_1)\left(\frac{\partial g}{\partial t_1}\right)_a + \cdots + (t_k - a_k)\left(\frac{\partial g}{\partial t_k}\right)_a - r^2 + J_1 + J_2,$$

where

$$(20) \quad J_1 = \tfrac{1}{2}\int_{R_k}\int_{R_k} [(t_1 - a_1)(x_1 - \xi_1) + \cdots + (t_k - a_k)(x_k - \xi_k)]^2 \cdot$$
$$\cdot [1 - \cos(a_1(x_1 - \xi_1) + \cdots + a_k(x_k - \xi_k))]\, dP(x)\, dP(\xi)$$

and

$$(21) \qquad J_2 = \frac{\theta}{6}\int_{R_k}\int_{R_k} |(t_1 - a_1)(x_1 - \xi_1) + \cdots + (t_k - a_k)(x_k - \xi_k)|^3 \, dP(x)\, dP(\xi).$$

We first estimate J_2. From Cauchy's inequality it follows that

$$|J_2| \leq \frac{r^3}{6}\int_{R_k}\int_{R_k} |(x_1 - \xi_1)^2 + \cdots + (x_k - \xi_k)^2|^{3/2}\, dP(x)\, dP(\xi).$$

We now apply the inequality

$$|(x_1 - \xi_1)^2 + \cdots + (x_k - \xi_k)^2|^{3/2} \leq k^{1/2}[|x_1 - \xi_1|^3 + \cdots + |x_k - \xi_k|^3] \leq$$
$$\leq 4k^{1/2}[|x_1|^3 + |\xi_1|^3 + \cdots + |x_k|^3 + |\xi_k|^3],$$

and then from (11) and (12):

$$(22) \qquad |J_2| \leq \tfrac{4}{3} k^{1/2} \beta_3 r^3.$$

The estimation of J_1 is somewhat more laborious. We first apply the Cauchy inequality and obtain

$$J_1 \leq \tfrac{1}{2} r^2 \int_{R_k}\int_{R_k} [(x_1 - \xi_1)^2 + \cdots + (x_k - \xi_k)^2] \cdot$$
$$\cdot [1 - \cos(a_1(x_1 - \xi_1) + \cdots + a_k(x_k - \xi_k))]\, dP(x)\, dP(\xi) =$$
$$(23) \qquad = \tfrac{1}{2} r^2 \cdot \iint_{(x_1-\xi_1)^2 + \cdots + (x_k - \xi_k)^2 \leq \lambda^2} + \tfrac{1}{2} r^2 \cdot \iint_{(x_1-\xi_1)^2 + \cdots + (x_k - \xi_k)^2 > \lambda^2} = \tfrac{1}{2} r^2 (J_3 + J_4),$$

λ being a positive number later to be determined. Now

$$J_3 \leq \lambda^2 \int_{R_k}\int_{R_k} [1 - \cos(a_1(x_1 - \xi_1) + \cdots + a_k(x_k - \xi_k))]\, dP(x)\, dP(\xi)$$

and hence from (15)

$$J_3 \leq \lambda^2 (1 - g(a)). \tag{24}$$

In order to estimate J_4 we observe the following inequalities:

$$8\, k^{1/2} \beta_3 \geq \int_{R_k} \int_{R_k} k^{1/2} [|x_1 - \xi_1|^3 + \cdots + |x_k - \xi_k|^3]\, dP(x)\, dP(\xi) \geq$$

$$\geq \int_{R_k} \int_{R_k} [(x_1 - \xi_1)^2 + \cdots + (x_k - \xi_k)^2]^{3/2}\, dP(x)\, dP(\xi) \geq$$

$$\geq \lambda \cdot \int\int_{(x_1-\xi_1)^2 + \cdots + (x_k - \xi_k)^2 > \lambda^2} [(x_1 - \xi_1)^2 + \cdots + (x_k - \xi_k)^2]\, dP(x)\, dP(\xi).$$

But now

$$J_4 \leq 2 \cdot \int\int_{(x_1-\xi_1)^2 + \cdots + (x_k - \xi_k)^2 > \lambda^2} [(x_1 - \xi_1)^2 + \cdots + (x_k - \xi_k)^2]\, dP(x)\, dP(\xi)$$

and hence

$$J_4 \leq 16 \frac{k^{1/2} \beta_3}{\lambda}. \tag{25}$$

Summing up we obtain from (23), (24) and (25):

$$J_1 \leq \tfrac{1}{2} r^2 \left[\lambda^2 (1 - g(a)) + 16 \frac{k^{1/2} \beta_3}{\lambda}\right]. \tag{26}$$

We now choose λ so that the right-hand side of (26) becomes as small as possible. This occurs for

$$\lambda = \frac{2\, k^{1/6} \beta_3^{1/3}}{(1 - g(a))^{1/3}}$$

and then

$$J_1 \leq 6\, r^2 k^{1/3} \beta_3^{2/3} (1 - g(a))^{1/3}. \tag{27}$$

From (19), (22) and (27) the desired inequality follows, and the lemma is proved.

Proof of Theorem 2.

A. First suppose that $g(t)$ has a maximum for $t = a = (a_1, a_2, \ldots, a_k)$ and that

$$\begin{cases} 1^\circ \;\; g(a) \geq 1 - \varepsilon, \\ 2^\circ \;\; 0 < \varepsilon \leq \dfrac{(1 - \lambda)^3}{6^3 k \beta_3^2}, \end{cases} \tag{28}$$

where

$$\lambda = \frac{1}{1 + \sqrt[3]{2}}.$$

Then from Lemma 1

$$g(t) \leq 1 - r^2 \{1 - 6\, k^{1/3} \beta_3^{2/3} \varepsilon^{1/3}\} + \tfrac{4}{3}\, k^{1/2} \beta_3 r^3,$$

where

$$r = \sqrt{\sum_{i=1}^{k} (t_i - a_i)^2}.$$

From $(28: 2^\circ)$ it follows:

$$g(t) \leq 1 - \lambda r^2 + \tfrac{4}{3}\, k^{1/2} \beta_3 r^3. \tag{29}$$

The function $1 - \lambda r^2 + \frac{4}{3} k^{1/2} \beta_3 r^3$ steadily decreases as r increases from 0 to

$$r = \varrho_0 = \frac{\lambda}{2\, k^{1/2} \beta_3}. \tag{30}$$

Hence from (29) and (30):

$$g(t) \leq 1 - \frac{\lambda}{3} r^2 \text{ for } 0 \leq r \leq \varrho_0. \tag{31}$$

According to (31) the set of t-points about $t = a$, for which $g(t) \geq 1 - \varepsilon$, is situated within a sphere of radius

$$\varrho_1 = \sqrt{\frac{3\varepsilon}{\lambda}}. \tag{32}$$

By the choice of λ, $\varrho_1 < \varrho_0$.

B. Now consider an arbitrary sphere S_ϱ in R_k of radius

$$\varrho = \tfrac{1}{2}\left\{\varrho_0 - \sqrt{\frac{3}{\lambda} \frac{(1-\lambda)^3}{6^3\, k\, \beta_3^2}}\right\} = \frac{1}{6(1 + \sqrt[3]{2})\, k^{1/2} \beta_3}. \tag{33}$$

We still assume that the condition $(28: 2^\circ)$ is satisfied. Then obviously

$$\varrho \leq \tfrac{1}{2}\left\{\varrho_0 - \sqrt{\frac{3\varepsilon}{\lambda}}\right\}.$$

Three cases may occur.

1. There exists a point a in S_ϱ for which $g(a)$ is maximum and $g(a) \geq 1 - \varepsilon$. According to the choice of ϱ, S_ϱ is entirely situated within a sphere with its centre at a and of radius ϱ_0. Hence according to (32) the set of t-points in S_ϱ, satisfying the condition $g(t) \geq 1 - \varepsilon$, is entirely situated within a sphere of radius $\sqrt{\frac{3\varepsilon}{\lambda}}$, i. e.

$$m_{S_\varrho}\{g(t) \geq 1 - \varepsilon\} \leq \frac{\pi^{k/2}}{\Gamma\left(1 + \frac{k}{2}\right)} \left(\frac{3\varepsilon}{\lambda}\right)^{k/2}. \tag{34}$$

2. There is no maximum in S_ϱ but there exists a point p in S_ϱ such that $g(p) \geq 1 - \varepsilon$. It is easily seen from Lemma 1 that there must exist a point a in the neighbourhood of p for which $g(t)$ is maximum, and from A it follows that a is at the most at the distance $\sqrt{\frac{3\varepsilon}{\lambda}}$ from p. Owing to the choice of ϱ, S_ϱ still is situated within a sphere with its centre at a and of radius ϱ_0. The inequality (34) is still valid.

3. There is no point t in S_ϱ for which $g(t) \geq 1 - \varepsilon$. The validity of (34) is immediately clear.

C. Now we make the contradictory assumption to (28: 2°), i. e.

$$\varepsilon > \frac{(1-\lambda)^3}{6^3 k \beta_3^2}, \tag{35}$$

and consider an arbitrary sphere S_ϱ in R_k of radius (33). Our aim is the proof of the inequality

$$m_{S_\varrho}\{g(t) \geq 1 - \varepsilon\} \leq K \cdot \varepsilon^{k/2}, \tag{36}$$

K being a constant. The smallest possible value of K on the assumption (35) obviously occurs if the left-hand side of (36) is replaced by the volume of S_ϱ and if

$$\varepsilon = \frac{(1-\lambda)^3}{6^3 k \beta_3^2}.$$

Hence

$$m_{S_\varrho}\{g(t) \geq 1 - \varepsilon\} \leq \frac{\pi^{k/2} \varrho^k}{\Gamma\left(1 + \frac{k}{2}\right)} \left\{\frac{6^3 k \beta_3^2}{(1-\lambda)^3}\right\}^{k/2} \cdot \varepsilon^{k/2}. \tag{37}$$

D. Comparing the inequalities (34) and (37) and observing that the right-hand sides are equal for $\lambda = \frac{1}{1 + \sqrt[3]{2}}$, we obtain the desired result.

Remarks.

1°. In Theorem 2 the quantity β_3 occurs; later on, however, we are most interested in the fourth moments, i. e. the quantity

$$\beta_4 = \sum_{i=1}^{k} \beta_i^4.$$

Now

$$\beta_3 = \sum_{i=1}^{k} \beta_i^3 \leq \sum_{i=1}^{k} (\beta_i^4)^{3/4} \leq k^{1/4} \left(\sum_{i=1}^{k} \beta_i^4 \right)^{3/4} = k^{1/4} \beta_4^{3/4}.$$

Hence in Theorem 2 we may replace $k^{1/2} \beta_3$ with $(k\beta_4)^{3/4}$ and obtain the result: If S_ϱ is an arbitrary k-dimensional sphere of radius

$$\varrho = \frac{1}{6(1 + \sqrt[3]{2})(k\beta_4)^{3/4}}$$

then (17) holds.

2°. Theorem 2 also holds for $k = 1$. Hence by a simple transformation we obtain Theorem 6, Chap. I.

2. **Some lemmata concerning $f_n(t)$.** The notations and hypotheses of Theorem 1, the main theorem, remain unaltered in this section; the same remark holds concerning the symbolic notation of the moments, introduced in Chap. VI. By $f_n(t)$ we denote the c. f. of ${}^{(n)}Y$. Then by the addition theorem, Chap. VI,

$$f_n(t) = \left\{ f\left(\frac{t}{\sqrt{n}} \right) \right\}^n. \tag{38}$$

If $t = (t_1, t_2, \ldots, t_k)$ is a point in R_k, the quantity r is defined by

$$r = \sqrt{t_1^2 + t_2^2 + \cdots + t_k^2}.$$

The following lemma is easily proved as in the one-dimensional case.

Lemma 2.

$$\left| f_n(t) - e^{-\frac{r^2}{2}} + \frac{i}{6\sqrt{n}} (\alpha_1 t_1 + \cdots + \alpha_k t_k)^3 e^{-\frac{r^2}{2}} \right| \leq c(k) \frac{\beta_4^{3/2}}{n} (r^4 + r^6) e^{-\frac{r^2}{3}}$$

for $r \leq \frac{\sqrt{n}}{(k\beta_4)^{3/4}}$, $c(k)$ *being a finite constant only depending on* k.

We further observe that the function

$$e^{-\frac{r^2}{2}} - \frac{i}{6\sqrt{n}} (\alpha_1 t_1 + \cdots + \alpha_k t_k)^3 e^{-\frac{r^2}{2}}$$

is the *Fourier transform* of the »frequency function»

$$\omega(x_1, x_2, \ldots, x_k) = \frac{1}{(2\pi)^{k/2}} e^{-\frac{\varrho^2}{2}} - \frac{1}{(2\pi)^{k/2}} \frac{1}{6\sqrt{n}} \left(\alpha_1 \frac{\partial}{\partial x_1} + \cdots + \alpha_k \frac{\partial}{\partial x_k} \right)^3 e^{-\frac{1}{2}\varrho^2}, \tag{39}$$

($\varrho^2 = x_1^2 + \cdots + x_k^2$). Let S be a sphere in R_k with its centre at $0 = (0, 0, \ldots, 0)$ and of radius a. With regard to subsequent applications we also notice that

$$(40) \qquad \int\limits_S \omega(x_1, x_2, \ldots, x_k)\, dx = \int\limits_S \frac{1}{(2\pi)^{k/2}} e^{-\frac{1}{2}\varrho^2}\, dx = \psi(a, k).$$

By S_ϱ we denote the sphere of radius

$$(41) \qquad \varrho = \frac{1}{6(1+\sqrt{2})(k\beta_4)^{3/4}}$$

introduced in Theorem 2. (Compare Remark 1, § 1). This theorem makes it possible to prove the following lemma.

Lemma 3. *Let S_ϱ be an arbitrary sphere in R_k of radius ϱ determined by* (41). *Then*

$$\int\limits_{S_\varrho} |f(t)|^n\, dt_1 \ldots dt_k \leq \frac{c(k)}{n^{k/2}},$$

$c(k)$ being a constant only depending on k.

Proof of Lemma 3. Dividing up the region of integration in the following way, we obtain from Theorem 2:

$$m_{S_\varrho}\left\{0 \leq |f(t)|^n < \left(1 - \frac{1}{2}\right)^{n/2}\right\} \leq K \cdot 1^{k/2},$$

$$m_{S_\varrho}\left\{\left(1 - \frac{1}{2}\right)^{n/2} \leq |f(t)|^n < \left(1 - \frac{1}{2^2}\right)^{n/2}\right\} \leq K \cdot \left(\frac{1}{2}\right)^{k/2},$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$m_{S_\varrho}\left\{\left(1 - \frac{1}{2^\nu}\right)^{n/2} \leq |f(t)|^n < \left(1 - \frac{1}{2^{\nu+1}}\right)^{n/2}\right\} \leq K \cdot \left(\frac{1}{2^\nu}\right)^{k/2},$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots,$$

K being the constant of the right-hand side of (17). Hence

$$J = \int\limits_{S_\varrho} |f(t)|^n\, dt_1 \ldots dt_k \leq K \sum_{\nu=0}^{\infty} \left(1 - \frac{1}{2^{\nu+1}}\right)^{n/2} \left(\frac{1}{2^\nu}\right)^{k/2}.$$

By comparing the series with an integral it is easily found that $J \leq \frac{c(k)}{n^{k/2}}$, which proves the lemma.

3. **Construction of an auxiliary function.** Let the function $Q_a(x_1, x_2, \ldots, x_k)$ in R_k, $(k \geq 2)$, be defined in the following way:

$$Q_a(x_1, x_2, \ldots, x_k) = \begin{cases} 1 \text{ for } \sqrt{x_1^2 + x_2^2 + \cdots + x_k^2} \leq a \\ 0 \text{ for } \sqrt{x_1^2 + x_2^2 + \cdots + x_k^2} > a \end{cases}, \tag{43}$$

where a is an assigned positive number. The Fourier transform

$$q_a(t_1, t_2, \ldots, t_k) = \int_{R_k} e^{i(t_1 x_1 + t_2 x_2 + \cdots + t_k x_k)} Q_a(x_1, x_2, \ldots, x_k)\, dx_1 \ldots dx_k$$

is easily evaluated according to well-known methods.[1] It is found that

$$q_a(t_1, t_2, \ldots, t_k) = \left(\frac{2\pi a}{r}\right)^{k/2} J_{k/2}(ar), \qquad (r = \sqrt{t_1^2 + t_2^2 + \cdots + t_k^2}), \tag{44}$$

$J_{k/2}(z)$ denoting the Bessel function of order $k/2$. Now consider the convolution function

$$M(x_1, x_2, \ldots, x_k) = \frac{\Gamma\left(1 + \frac{k}{2}\right)}{\pi^{k/2} \varepsilon^k} \int_{R_k} Q_a(x_1 - \xi_1, \ldots, x_k - \xi_k)\, Q_\varepsilon(\xi_1, \ldots, \xi_k)\, d\xi_1 \ldots d\xi_k, \tag{45}$$

where $0 < \varepsilon < a$. From (43) it is easily seen that

$$\begin{cases} M(x_1, x_2, \ldots, x_k) = \begin{cases} 1 \text{ for } \sqrt{x_1^2 + x_2^2 + \cdots + x_k^2} \leq a - \varepsilon \\ 0 \text{ for } \sqrt{x_1^2 + x_2^2 + \cdots + x_k^2} \geq a + \varepsilon \end{cases}, \\ \qquad |M(x)| \leq 1 \text{ for all } x. \end{cases} \tag{46}$$

The Fourier transform of M, $m(t_1, t_2, \ldots, t_k)$, is obtained by

$$m(t_1, t_2, \ldots, t_k) = \left(\frac{2\pi a}{r}\right)^{k/2} J_{k/2}(ar)\, 2^{k/2}\, \Gamma\left(1 + \frac{k}{2}\right) \frac{J_{k/2}(\varepsilon r)}{(\varepsilon r)^{k/2}}, \tag{47}$$

owing to the fact that the Fourier transform of a convolution is equal to the product of the transforms corresponding to the functions in the convolution.

From (46) and (47) we obtain by a simple transformation: the function

$$\left(\frac{2\pi\left(a + \frac{\varepsilon}{2}\right)}{r}\right)^{k/2} J_{k/2}\left[\left(a + \frac{\varepsilon}{2}\right) r\right] 2^{k/2}\, \Gamma\left(1 + \frac{k}{2}\right) \frac{J_{k/2}\left(\frac{\varepsilon r}{2}\right)}{\left(\frac{\varepsilon r}{2}\right)^{k/2}} \tag{48}$$

[1] See for instance BOCHNER [1], § 43.

is the Fourier transform of a function =

$$= \begin{cases} 1 \text{ for } \sqrt{x_1^2 + x_2^2 + \cdots + x_k^2} \le a \\ 0 \text{ for } \sqrt{x_1^2 + x_2^2 + \cdots + x_k^2} \ge a + \varepsilon \end{cases},$$

the modulus of which is bounded by 1 for all x.

In the same way: the function

$$(49) \qquad \left(\frac{2\pi\left(a - \frac{\varepsilon}{2}\right)}{r}\right)^{k/2} J_{k/2}\left[\left(a - \frac{\varepsilon}{2}\right) r\right] 2^{k/2} \Gamma\left(1 + \frac{k}{2}\right) \frac{J_{k/2}\left(\frac{\varepsilon r}{2}\right)}{\left(\frac{\varepsilon r}{2}\right)^{k/2}}$$

is the Fourier transform of a function =

$$= \begin{cases} 1 \text{ for } \sqrt{x_1^2 + x_2^2 + \cdots + x_k^2} \le a - \varepsilon \\ 0 \text{ for } \sqrt{x_1^2 + x_2^2 + \cdots + x_k^2} \ge a \end{cases},$$

the modulus of which is bounded by 1 for all x.

By $c_1, c_2, \ldots$ we denote a sequence of finite positive constants only depending on k. We now use the following well-known properties of the Bessel functions:

$$(50) \qquad \begin{cases} 1^\circ \quad \left|\frac{J_{k/2}(z)}{z^{k/2}}\right| \le c_1 \text{ for all positive } z \\ 2^\circ \quad |J_{k/2}(z)| \le \frac{c_2}{\sqrt{z}} \text{ for all positive } z \end{cases}$$

The relations (48), (49) and (50) imply the validity of

Lemma 4. *Let a and ε be two assigned constants and $0 < \varepsilon < a$. There exists a function $H(x_1, x_2, \ldots, x_k, a, \varepsilon) = H(\varrho, a, \varepsilon)$ only depending on*

$$\varrho = \sqrt{x_1^2 + x_2^2 + \cdots + x_k^2}$$

such that

$$1^\circ \quad H(\varrho, a, \varepsilon) = \begin{cases} 1 \textit{ for } 0 \le \varrho \le a \\ 0 \textit{ for } \varrho \ge a + \varepsilon \end{cases}, \textit{ and } |H(\varrho, a, \varepsilon)| \le 1$$

for all ϱ.

Further the Fourier transform of H, $h(t_1, t_2, \ldots, t_k, a, \varepsilon) = h(r, a, \varepsilon)$, is only dependent on

$$r = \sqrt{t_1^2 + t_2^2 + \cdots + t_k^2}$$

and

$$2^\circ \quad |h(r, a, \varepsilon)| \le c_3 \cdot \frac{a^{\frac{k-1}{2}}}{r^{\frac{k+1}{2}}},$$

$$3^\circ \quad |h(r, a, \varepsilon)| \leq c_4 \cdot \frac{a^{\frac{k-1}{2}}}{\varepsilon^{\frac{k}{2}} r^{\frac{2k+1}{2}}}.$$

There also exists a function $H(\varrho, a, -\varepsilon)$ *such that*

$$4^\circ \quad H(\varrho, a, -\varepsilon) = \begin{cases} 1 \textit{ for } 0 \leq \varrho \leq a - \varepsilon \\ 0 \textit{ for } \quad \varrho \geq a \end{cases}, \textit{ and } |H \varrho, a, -\varepsilon)| \leq 1$$

for all ϱ, *the Fourier transform of which,* $h(r, a, -\varepsilon)$, *satisfies the inequalities* 2° *and* 3°.

4. **Proof of the main theorem.** We use the same *notation* as in the statement of Theorem 1 but repeat it here for the sake of lucidity. $P_n(E)$ denotes the pr. f. of the sum

$$(51) \qquad {}^{(n)}Y = \frac{Y^{(1)} + Y^{(2)} + \cdots Y^{(n)}}{\sqrt{n}};$$

$f_n(t)$ is the corresponding c. f. and

$$f_n(t) = \left\{ f\left(\frac{t}{\sqrt{n}}\right)\right\}^n,$$

$f(t)$ being the c. f. of the variable $Y^{(\nu)}$, $(\nu = 1, 2, \ldots, n)$. $\mu_n(a)$ is the probability of ${}^{(n)}Y$ being situated within the sphere S with its centre $0 = (0, 0, \ldots, 0)$ and of radius a.

$$\varphi(x_1, x_2, \ldots, x_k) = \frac{1}{(2\pi)^{k/2}} e^{-\frac{1}{2}(x_1^2 + x_2^2 + \cdots + x_k^2)},$$

the normal frequency function.

$$\psi(a, k) = \int_S \varphi(x_1, x_2, \ldots, x_k)\, dx_1 \ldots dx_k; \text{ see Chap. VI, (35).}$$

$$\omega(x_1, x_2, \ldots, x_k) = \varphi(x_1, x_2, \ldots, x_k) - \\ - \frac{1}{(2\pi)^{k/2}} \frac{1}{6\sqrt{n}} \left(\alpha_1 \frac{\partial}{\partial x_1} + \cdots + \alpha_k \frac{\partial}{\partial x_k}\right)^3 e^{-\frac{1}{2}(x_1^2 + \cdots + x_k^2)};$$

$$r = \sqrt{t_1^2 + t_2^2 + \cdots + t_k^2}; \qquad \varrho = \sqrt{x_1^2 + x_2^2 + \cdots + x_k^2}.$$

Further,

$$e^{-\frac{r^2}{2}} - \frac{i}{6\sqrt{n}} (\alpha_1 t_1 + \cdots + \alpha_k t_k)^3 e^{-\frac{r^2}{2}}$$

is the Fourier transform of $\omega(x_1, x_2, \ldots, x_k)$, see Lemma 2.

$$U_n(E) = \int_E \omega(x_1, x_2, \ldots, x_k)\, dx_1\, dx_2 \ldots dx_k.$$

$H(\varrho, a, \varepsilon)$, $H(\varrho, a, -\varepsilon)$, $h(r, a, \varepsilon)$ and $h(r, a, -\varepsilon)$ are defined in Lemma 4. Finally we remark that the function

$$(52) \quad \Delta_n(t_1, t_2, \ldots, t_k) = f_n(t_1, t_2, \ldots, t_k) - e^{-\frac{r^2}{2}} + \frac{i}{6\sqrt{n}}(\alpha_1 t_1 + \cdots + \alpha_k t_k)^3 e^{-\frac{r^2}{2}}$$

is the c. f. of the set function $P_n(E) - U_n(E)$.

By $c_1, c_2, \ldots$ we denote a sequence of finite positive constants only depending on k.

The starting point of the proof is the formula

$$(53) \quad \int_{R_k} H(\varrho, a, \varepsilon)\, d\{P_n(x) - U_n(x)\} = \frac{1}{(2\pi)^k}\int_{R_k} \Delta_n(t_1, t_2, \ldots, t_k)\, h(r, a, \varepsilon)\, dt_1 \ldots dt_k,$$

the validity of which is immediately clear according to the Fourier inversion formula. Owing to the properties of H, (Lemma 4: 1°), we obtain from (53) and (40):

$$(54) \quad \mu_n(a) - \psi(a+\varepsilon, k) - \frac{1}{(2\pi)^{k/2}}\frac{1}{6\sqrt{n}}\int\limits_{a \le \varrho \le a+\varepsilon}\left|\left(\alpha_1\frac{\partial}{\partial x_1} + \cdots + \alpha_k\frac{\partial}{\partial x_k}\right)^3 e^{-\frac{\varrho^2}{2}}\right| dx_1 \ldots dx_k \le$$
$$\le \frac{1}{(2\pi)^k}\int_{R_k} \Delta_n(t)\, h(r, a, \varepsilon)\, dt.$$

Using $H(\varrho, a, -\varepsilon)$ in (53) instead of $H(\varrho, a, \varepsilon)$ we obtain in the same way:

$$(55) \quad \mu_n(a) - \psi(a-\varepsilon, k) + \frac{1}{(2\pi)^{k/2}}\frac{1}{6\sqrt{n}}\int\limits_{a-\varepsilon \le \varrho \le a}\left|\left(\alpha_1\frac{\partial}{\partial x_1} + \cdots + \alpha_k\frac{\partial}{\partial x_k}\right)^3 e^{-\frac{\varrho^2}{2}}\right| dx_1 \ldots dx_k \ge$$
$$\ge \frac{1}{(2\pi)^k}\int_{R_k} \Delta_n(t)\, h(r, a, -\varepsilon)\, dt.$$

Hence from (54) and (55):

$$(56) \qquad |\mu_n(a) - \psi(a, k)| \le \operatorname{Max}(A_1, A_2),$$

where

$$(57) \quad A_1 = |\psi(a+\varepsilon, k) - \psi(a, k)| +$$
$$+ \frac{1}{(2\pi)^{k/2}}\frac{1}{6\sqrt{n}}\int\limits_{a \le \varrho \le a+\varepsilon}\left|\left(\alpha_1\frac{\partial}{\partial x_1} + \cdots + \alpha_k\frac{\partial}{\partial x_k}\right)^3 e^{-\frac{\varrho^2}{2}}\right| dx_1 \ldots dx_k +$$
$$+ \frac{1}{(2\pi)^k}\int_{R_k} |\Delta_n(t)\, h(r, a, \varepsilon)|\, dt,$$

$$(58)\quad A_2 = |\psi(a,k) - \psi(a-\varepsilon,k)| + \\ + \frac{1}{(2\pi)^{k/2}}\frac{1}{6\sqrt{n}}\int\limits_{a-\varepsilon\le\varrho\le a}\left|\left(\alpha_1\frac{\partial}{\partial x_1}+\cdots+\alpha_k\frac{\partial}{\partial x_k}\right)^3 e^{-\frac{\varrho^2}{2}}\right| dx_1\ldots dx_k + \\ + \frac{1}{(2\pi)^k}\int\limits_{R_k}|\Delta_n(t)\,h(r,a,-\varepsilon)|\,dt.$$

The relation (56), together with (57) and (58), is the main inequality of the subsequent estimations.

Without loss of generality we may make the following assumptions.

$$(59)\qquad 1^\circ\quad a < \log(2+n),$$

or else we choose $\varepsilon = a/2$ and proceed as in the subsequent estimations.

$$(60)\qquad 2^\circ\quad \frac{\beta_4^{3/2}}{n^{\frac{k}{k+1}}} \le \tfrac{1}{2},$$

or else Theorem 1 is true with $c(k) = 2$.

We may confine ourselves to the estimation of A_1, A_2 being treated in a similar way. Now choose

$$(61)\qquad \varepsilon = a\cdot\frac{\beta_4^{3/2}}{n^{\frac{k}{k+1}}}.$$

Hence $0 < \varepsilon < a$. It is immediately found that

$$(62)\quad |\psi(a+\varepsilon,k)-\psi(a,k)| + \frac{1}{(2\pi)^{k/2}}\frac{1}{6\sqrt{n}}\int\limits_{a\le\varrho\le a+\varepsilon}\left|\left(\alpha_1\frac{\partial}{\partial x_1}+\cdots+\alpha_k\frac{\partial}{\partial x_k}\right)^3 e^{-\frac{\varrho^2}{2}}\right| dx_1\ldots dx_k \le \\ \le c_1\cdot\frac{\beta_4^{3/2}}{n^{\frac{k}{k+1}}}.$$

We proceed to the estimation of the last term of A_1 and begin by dividing up the region of integration:

$$(63)\quad I = \frac{1}{(2\pi)^k}\int\limits_{R_k}|\Delta_n(t)\,h(r,a,\varepsilon)|\,dt = \frac{1}{(2\pi)^k}\int\limits_{0\le r\le\frac{\sqrt{n}}{(k\beta_4)^{3/4}}} + \frac{1}{(2\pi)^k}\int\limits_{r>\frac{\sqrt{n}}{(k\beta_4)^{3/4}}} = I_1 + I_2.$$

A. **Estimation of I_1.** By virtue of Lemma 2, Lemma 4: 2° and (59) it follows:

$$I_1 \leq c_2 \int\limits_{R_k} \frac{\beta_4^{3/2}}{n}(r^4 + r^6)e^{-\frac{r^2}{3}} \frac{a^{\frac{k-1}{2}}}{r^{\frac{k+1}{2}}} d t_1 \ldots d t_k \leq c_2 \beta_4^{3/2} \frac{2\pi^{k/2}}{\Gamma\left(\frac{k}{2}\right)} \cdot$$

$$\cdot \frac{(\log(2+n))^{\frac{k-1}{2}}}{n} \int\limits_0^\infty (r^4 + r^6) r^{\frac{k-3}{2}} e^{-\frac{r^2}{3}} d r \leq c_3 \cdot \frac{\beta_4^{3/2}}{n^{\frac{k}{k+1}}}.$$

Hence

$$I_1 \leq c_3 \cdot \frac{\beta_4^{3/2}}{n^{\frac{k}{k+1}}}. \tag{64}$$

B. **Estimation of I_2.** This is the main point of the proof. We use the earlier result regarding the value distribution of $|f(t)|$, (Theorem 2 and Lemma 3).

$$I_2 \leq \frac{1}{(2\pi)^k} \int\limits_{r > \frac{\sqrt{n}}{(k\beta_4)^{3/4}}} \left| f\left(\frac{t_1}{\sqrt{n}}, \ldots, \frac{t_k}{\sqrt{n}}\right)\right|^n |h(r, a, \varepsilon)|\, d t + c_4 \cdot \frac{\beta_4^{3/2}}{n^{\frac{k}{k+1}}} =$$

$$= \frac{n^{k/2}}{(2\pi)^k} \int\limits_{r > \frac{1}{(k\beta_4)^{3/4}}} |f(t_1, t_2, \ldots, t_k)|^n |h(r\sqrt{n}, a, \varepsilon)|\, d t + c_4 \cdot \frac{\beta_4^{3/2}}{n^{\frac{k}{k+1}}} =$$

$$= \frac{n^{k/2}}{(2\pi)^k} \int\limits_{\frac{1}{(k\beta_4)^{3/4}} \leq r \leq \frac{1}{\varepsilon\sqrt{n}}} + \frac{n^{k/2}}{(2\pi)^k} \int\limits_{r > \frac{1}{\varepsilon\sqrt{n}}} + c_4 \cdot \frac{\beta_4^{3/2}}{n^{\frac{k}{k+1}}} = I_3 + I_4 + c_4 \cdot \frac{\beta_4^{3/2}}{n^{\frac{k}{k+1}}}. \tag{65}$$

$\left(\text{Without loss of generality we may suppose } \frac{1}{\varepsilon\sqrt{n}} > \frac{1}{(k\beta_4)^{3/4}}\right)$. From Lemma 4: 2° we first obtain:

$$I_3 \leq c_5 \cdot \frac{a^{\frac{k-1}{2}} n^{k/2}}{n^{\frac{k+1}{4}}} \int\limits_{\frac{1}{(k\beta_4)^{3/4}} \leq r \leq \frac{1}{\varepsilon\sqrt{n}}} \frac{|f(t_1, t_2, \ldots, t_k)|^n}{r^{\frac{k+1}{2}}} d t_1 \ldots d t_k,$$

or

$$\begin{cases} I_3 \leq c_5\, a^{\frac{k-1}{2}} n^{\frac{k-1}{4}} I_5, \text{ where} \\ I_5 = \displaystyle\int\limits_{\frac{1}{(k\beta_4)^{3/4}} \leq r \leq \frac{1}{\varepsilon\sqrt{n}}} \frac{|f(t_1, t_2, \ldots, t_k)|^n}{r^{\frac{k+1}{2}}} d t_1 \ldots d t_k. \end{cases} \tag{66}$$

From Lemma 4: 3° we obtain in the same way:

$$(67)\qquad \begin{cases} I_4 \leq c_6 \cdot \dfrac{a^{\frac{k-1}{2}}}{\varepsilon^{k/2}\, n^{1/4}} I_6, \text{ where} \\ I_6 = \displaystyle\int\limits_{r > \frac{1}{\varepsilon \sqrt{n}}} \frac{|f(t_1, t_2, \ldots, t_k)|^n}{r^{\frac{2k+1}{2}}}\, d\,t_1 \ldots d\,t_k. \end{cases}$$

B: a. Estimation of I_5 and I_3. By D we denote the region between two k-dimensional cubes, the edges of which are parallel with the coordinate axeses. Furthermore the interior cube is inscribed in a sphere with its centre at $o = (0, 0, \ldots, 0)$ and of radius $\frac{1}{(k\beta_4)^{3/4}}$, while the exterior cube is circumscribed around a sphere with its centre at $o = (0, 0, \ldots, 0)$ and of radius $\frac{1}{\varepsilon\sqrt{n}}$. Then

$$(68)\qquad I_5 < \int\limits_D \frac{|f(t_1, t_2, \ldots, t_k)|^n}{r^{\frac{k+1}{2}}}\, d\,t_1 \ldots d\,t_k = I_7.$$

The interior cube has the edge-length

$$(69)\qquad 2\,s = \frac{2}{\sqrt{k}\,(k\beta_4)^{3/4}}.$$

By K_ϱ we always denote a cube with the edges parallel with the coordinate axeses, which is inscribed in a sphere S_ϱ of radius

$$\varrho = \frac{1}{6(1 + \sqrt[3]{2})\,(k\beta_4)^{3/4}}.$$

The edge-length b of K_ϱ is calculated to be

$$(70)\qquad b = \frac{2}{\sqrt{k}} \cdot \frac{1}{6(1 + \sqrt[3]{2})\,(k\beta_4)^{3/4}}.$$

By Lemma 3 we have:

$$(71)\qquad \int\limits_{K_\varrho} |f(t_1, t_2, \ldots, t_k)|^n\, d\,t_1 \ldots d\,t_k \leq \frac{c_7}{n^{k/2}}.$$

Now consider a sequence of cubes with their centres at $o = (0, 0, \ldots, 0)$, their edges parallel with the coordinate axeses and their edge-lengths $2(s + \nu b)$,

$\left(\nu = 0, 1, 2, \ldots, \nu_0, \nu_0 \text{ being so determined that } s + (\nu_0 - 1)b \leq \frac{1}{\varepsilon\sqrt{n}} < s + \nu_0 b\right)$.
The number of cubes K_ϱ without common part which may be situated in the space between two cubes of the sequence with the edge-lengths $2\{s + \nu b\}$ and $2\{s + (\nu + 1)b\}$ respectively is less than or equal to

$$\left(2\frac{s + (\nu + 1)b}{b} + 1\right)^k - \left(2\frac{s + \nu b}{b} - 1\right)^k = \left(2\frac{s}{b} + 2\nu + 3\right)^k -$$
$$- \left(2\frac{s}{b} + 2\nu - 1\right)^k \leq \frac{4k \cdot 2^{k-1}}{b^{k-1}}\left(s + \nu b + \frac{3b}{2}\right)^{k-1}.$$

Hence according to (71) the contribution of this region of integration to I_7 is

$$\leq \frac{c_8}{n^{k/2}} \cdot \frac{1}{b^{k-1}} \frac{\left(s + \nu b + \frac{3b}{2}\right)^{k-1}}{(s + \nu b)^{\frac{k+1}{2}}} \leq \frac{c_9}{n^{k/2}} \cdot \frac{1}{b^{k-1}} (s + \nu b)^{\frac{k-3}{2}}.$$

Thus

$$I_5 \leq \frac{c_9}{n^{k/2}} \cdot \frac{1}{b^{k-1}} \sum_{\nu=0}^{\nu_0} (s + \nu b)^{\frac{k-3}{2}} \leq \frac{c_9}{n^{k/2} \cdot b^k} \int_0^{\frac{2}{\varepsilon\sqrt{n}}} y^{\frac{k-3}{2}} dy$$

or

$$I_5 \leq \frac{c_{10}}{n^{k/2}} \cdot \frac{1}{b^k \cdot \varepsilon^{\frac{k-1}{2}} n^{\frac{k-1}{4}}}. \tag{72}$$

From (66), (72), (70) and (61) we finally obtain:

$$I_3 \leq c_{11} \cdot a^{\frac{k-1}{2}} n^{\frac{k-1}{4}} \frac{1}{n^{k/2} n^{\frac{k-1}{4}}} \cdot \frac{\beta_4^{\frac{3k}{4}} n^{\frac{k-1}{2} \cdot \frac{k}{k+1}}}{a^{\frac{k-1}{2}} \beta_4^{\frac{3}{2} \cdot \frac{k-1}{2}}} = \frac{c_{11} \cdot \beta_4^{3/4}}{n^{\frac{k}{k+1}}}$$

or

$$I_3 \leq c_{11} \cdot \frac{\beta_4^{3/4}}{n^{\frac{k}{k+1}}}. \tag{73}$$

(Since the dispersions are all 1, $\beta_4 > 1$; compare Chap. VI, (11)).

B: b. Estimation of I_6 and I_4. The estimation of I_6 proceeds in exactly the same way as that of I_5. Dividing up the region of integration into a sequence of cubes, we apply Lemma 3 and find:

$$I_6 \leq \frac{c_{12}}{n^{k/2}\, b^k} \int\limits_{\frac{1}{2\sqrt{k}\,\varepsilon\sqrt{n}}}^{\infty} \frac{y^{k-1}}{y^{\frac{2k+1}{2}}}\, dy = \frac{c_{13}}{n^{k/2}} \cdot \frac{n^{1/4}\, \varepsilon^{1/2}}{b^k}. \tag{74}$$

Hence from (67), (74), (70) and (61):

$$I_4 \leq c_{14} \cdot \frac{a^{\frac{k-1}{2}}\, n^{\frac{k-1}{2} \cdot \frac{k}{k+1}}}{a^{\frac{k-1}{2}} \cdot \beta_4^{\frac{3}{4} \cdot (k-1)}\, n^{1/4}} \cdot \frac{\beta_4^{\frac{3k}{4}}\, n^{1/4}}{n^{k/2}} = c_{14} \cdot \frac{\beta_4^{3/4}}{n^{\frac{k}{k+1}}}$$

or

$$I_4 \leq c_{14} \cdot \frac{\beta_4^{3/2}}{n^{\frac{k}{k+1}}}. \tag{75}$$

The final result of Section B follows from (65), (73) and (75):

$$I_2 \leq c_{15} \cdot \frac{\beta_4^{3/2}}{n^{\frac{k}{k+1}}}. \tag{76}$$

Conclusion. From (63), (64) and (76) we obtain:

$$I \leq c_{16} \cdot \frac{\beta_4^{3/2}}{n^{\frac{k}{k+1}}}, \tag{77}$$

and hence from (57), (62), (63) and (77):

$$A_1 \leq c_{17} \cdot \frac{\beta_4^{3/2}}{n^{\frac{k}{k+1}}}.$$

This proves the theorem.

Remarks.

1. We have proved Theorem 1 on the assumption that all the r. v.'s have the same probability distribution. If this is not the case it is necessary to modify the proof a little. The main difference consists in the estimation of the integral

$$I = \int\limits_{S_\varrho} |f_1(t) f_2(t) \ldots f_n(t)|\, d t_1 \ldots d t_k.$$

Using the inequality of Hölder in a suitable manner we may estimate I by a product of integrals of type

$$\int\limits_{S_\varrho} |f_\nu(t)|^{n_\nu}\, d t_1 \ldots d t_k.$$

The Lemma 3 is applicable to every such integral. In this way it is possible to prove all the theorems of Part I concerning probability distributions, for instance the inequality

$$|\overline{F_n(x)} - \Phi(x)| \le c \cdot \frac{\varrho_{3n}}{\sqrt{n}}; \text{ (Chap. III, (15)).}$$

2. In Theorem 1 we have made the least restrictive assumptions possible. Now suppose that $\overline{\lim\limits_{|t|\to\infty}} |f(t_1, t_2, \ldots, t_k)| < 1$. Then it is easily seen that the remainder term has the order of magnitude $O\left(\frac{1}{n}\right)$. Absolute moments of order greater than 4 being supposed to be finite, it is also possible to obtain an asymptotic expansion as in the one-dimensional case.

3. In the one-dimensional case we have found that the absolute third moment plays an important part in the problem of obtaining the general true order of magnitude of the remainder term. In the multi-dimensional case the same applies to the fourth moments (compare the next chapter).

4. In Theorem 1 we have confined ourselves to the case where ${}^{(n)}Y$ belongs to a sphere S with its centre at $0 = (0, 0, \ldots, 0)$. It is, indeed, possible to escape from this restriction; but the form of the theorem given here is especially simple.

5. **Application to the χ^2 method.** In this section we shall briefly apply Theorem 1 to the so-called χ^2 method; we confine ourselves to the most simple case which may concretely be illustrated by the drawing of balls from an urn. Let an urn contain a collection of balls, white, black etc. and let in all $k + 1$ different colours be represented. Suppose that the probability of drawing a white ball is p_1, of drawing a black p_2 etc. In every trial a ball is drawn, its colour is noted and then it is replaced. In all we suppose that n trials are made, the results of which consist of m_1' white, m_2' black balls etc. Obviously

$$m_1' + m_2' + \cdots + m_{k+1}' = n.$$

If the drawing were so performed as to give an exact representation of the distribution of the balls among the different colours, these numbers would be the mean values:

$$m_1 = p_1 n,\ m_2 = p_2 n,\ \ldots,\ m_{k+1} = p_{k+1} n.$$

Now we ask: What is the probability $p(\chi)$ of

(78) $$\sum_{\nu=1}^{k+1} \frac{(m'_\nu - m_\nu)^2}{m_\nu} > \chi^2?$$

The result of every trial may be characterized by a k-dimensional r. v. $X = (X_1, X_2, \ldots, X_k)$, the component X_1 assuming the value 1 if a white ball is drawn, otherwise the value 0 etc. The mean value of the component X_ν is equal to p_ν and the fourth moments are finite.

Let the results of the n trials be represented by a sequence of variables $X^{(1)}, X^{(2)}, \ldots, X^{(n)}$ and form

(79) $${}^{(n)}X = ({}^{(n)}X_1, {}^{(n)}X_2, \ldots, {}^{(n)}X_k) = \frac{X^{(1)} + X^{(2)} + \cdots + X^{(n)} - M^{(n)}}{\sqrt{n}},$$

$M^{(n)}$ being a point in R_k with the coordinates $(np_1, np_2, \ldots, np_k)$. It is possible to show[1] that

$$\sum_{\nu=1}^{k+1} \frac{(m'_\nu - m_\nu)^2}{m_\nu} = \sum_{r,s=1}^{k} \frac{\Delta_{rs}}{\Delta} {}^{(n)}X_r \, {}^{(n)}X_s,$$

where $\Delta = ||\mu_{rs}||$ and Δ_{rs} is the algebraic complement of μ_{rs} with respect to Δ. By the corollary of Theorem 1 it follows:

(80) $$p(\chi) = 1 - \psi(\chi, k) + \frac{\theta(p_1, p_2, \ldots, p_k)}{n^{\frac{k}{k+1}}},$$

where $\theta(p_1, p_2, \ldots, p_k)$ is a finite quantity only depending on $p_1, p_2, \ldots, p_k$ and

$$\psi(\chi, k) = \frac{1}{(2\pi)^{k/2}} \int\limits_{x_1^2 + \cdots + x_k^2 \leq \chi^2} e^{-\frac{1}{2}(x_1^2 + \cdots + x_k^2)} \, dx_1 \ldots dx_k.$$

The relation (80) answers our question.

The so-called χ^2 method, applied to this case, consists of taking $p(\chi) = 1 - \psi(\chi, k)$; thus the remainder term of (80) is neglected. By our methods we have, however, been able to estimate the order of magnitude of the remainder. I hope to have the opportunity of returning to these questions at a later date.

---

[1] PEARSON [1].

Chapter VIII.

Lattice Distributions. Connection with the Lattice Point Problem.

In the one-dimensional case we have found that it is generally impossible to obtain an expansion of the d. f. $F_n(x)$ corresponding to the normalized sum of a large number n of independent r. v.'s in a series of continuous functions, and with a remainder term of order of magnitude less than $\frac{1}{\sqrt{n}}$. If every variable of the sum has the same d. f. and this is of lattice type, then $F_n(x)$ has discontinuities of order of magnitude $\frac{1}{\sqrt{n}}$. In the multi-dimensional case the situation is analogous. In the last chapter we obtained the remainder $O\left(\frac{1}{n^{\frac{k}{k+1}}}\right)$. The pr. f. being subject to certain conditions, we also remarked that this estimation may be improved and that it is possible to obtain an asymptotic expansion, provided that absolute moments of order greater than 4 are finite. However, as we shall see in this chapter, this is generally not possible even if all moments are finite, and we shall show that the remainder term $O\left(\frac{1}{n^{\frac{k}{k+1}}}\right)$ is intimately connected with a certain kind of probability distribution, the lattice distribution. We shall compare this remainder with that of the lattice point problem in the analytic theory of numbers.

1. **On characteristic functions having the modulus equal to 1 at a sequence of points.** It is to be expected that the remainder term in Theorem 1, Chap. VII, will be of as large order of magnitude as possible when the modulus of the c. f. $f(t_1, t_2, \ldots, t_k)$ is equal to 1 at a sequence of points different from $(0, 0, \ldots, 0)$ (cf. the one-dimensional case, Chap. IV). Thus, let us find out when this case may occur. For the sake of simplicity we only treat the two-dimensional case.

Consider a c. f. $f(t_1, t_2)$ corresponding to a two-dimensional pr. f. $P(E)$, and suppose that there exists a finite point $(t_1^{(0)}, t_2^{(0)}) \neq (0, 0)$ such that

$$(1) \qquad |f(t_1^{(0)}, t_2^{(0)})| = 1.$$

As in the one-dimensional case (Theorem 5, Chap. I) it follows from (1) that the *probability mass necessarily is concentrated to the straight lines*

$$(2) \qquad t_1^{(0)} x_1 + t_2^{(0)} x_2 - \theta_0 = \nu \cdot 2\pi, \qquad (\nu = 0, \pm 1, \pm 2, \pm \cdots).$$

It is readily observed that only the following cases may occur:

a. The probability mass is concentrated to one single line. We neglect this case being of a one-dimensional nature.

b The probability mass is concentrated to at least two parallel lines, and there is a point $(t_1^{(1)}, t_2^{(1)})$ different from (0, 0) and $(t_1^{(0)}, t_2^{(0)})$ such that $\frac{t_1^{(1)}}{t_2^{(1)}} \neq \frac{t_1^{(0)}}{t_2^{(0)}}$ and $|f(t_1^{(1)}, t_2^{(1)})| = 1$. Then the probability mass is also situated in the lines

$$t_1^{(1)} x_1 + t_2^{(1)} x_2 - \theta_1 = \nu \cdot 2\pi, \qquad (\nu = 0, \pm 1, \pm 2, \pm \cdots), \tag{3}$$

and hence is concentrated to the points of intersection of the families of lines (2) and (3), i. e. is situated in a set of lattice points. We call such a distribution a *lattice distribution.*

c. The probability mass is concentrated to at least two parallel lines, and the distribution is not of lattice type. It is easily found that all points (t_1, t_2), for which $|f(t_1, t_2)| = 1$, form a set of equidistant points belonging to one single straight line through (0, 0).

Let us consider the lattice distributions more closely. It is convenient to use vectors. Let $P(E)$ be a lattice distribution in R_2, i. e. let there exist three vectors $\mathfrak{z}_0$, $\mathfrak{a}_1$ and $\mathfrak{a}_2$ ($\mathfrak{a}_1$ not parallel with $\mathfrak{a}_2$) such that the probability mass is concentrated to the points[1]

$$(x_1^{(\mu)}, x_2^{(\nu)}) = \mathfrak{z}_{\mu\nu} = \mathfrak{z}_0 + \mu\,\mathfrak{a}_1 + \nu\,\mathfrak{a}_2, \quad (\mu, \nu = 0, \pm 1, \pm 2, \pm \cdots). \tag{4}$$

$(x_1^{(\mu)}, x_2^{(\nu)})$ are the rectangular components of the vector $\mathfrak{z}_{\mu\nu}$ having the origin at (0, 0). Let the probability mass at $\mathfrak{z}_{\mu\nu}$ be $a_{\mu\nu} \geq 0$. Thus $\sum\limits_{\mu,\nu} a_{\mu\nu} = 1$. The c. f. of $P(E)$ is expressed by

$$f(t_1, t_2) = \sum_{\mu,\nu} a_{\mu\nu}\, e^{i(t_1 x_1^{(\mu)} + t_2 x_2^{(\nu)})}. \tag{5}$$

By $\mathfrak{t} = (t_1, t_2)$ we denote a vector having the rectangular components (t_1, t_2) and the origin at (0, 0). Further we put $f(t_1, t_2) = f(\mathfrak{t})$. Using vector notation and the concept of scalar product we obtain from (4) and (5):

$$f(\mathfrak{t}) = \sum_{\mu,\nu} a_{\mu\nu}\, e^{i\mathfrak{t}(\mathfrak{z}_0 + \mu\mathfrak{a}_1 + \nu\mathfrak{a}_2)} = e^{i\mathfrak{t}\mathfrak{z}_0} \sum_{\mu,\nu} a_{\mu\nu}\, e^{i\mathfrak{t}(\mu\mathfrak{a}_1 + \nu\mathfrak{a}_2)}. \tag{6}$$

From (6) it is readily observed, that the necessary and sufficient condition for

$$|f(\mathfrak{t})| = 1 \tag{7}$$

[1] It is always supposed that $\mathfrak{a}_1$ and $\mathfrak{a}_2$ are the greatest possible of their kind.

is that

$$(8) \qquad \mathfrak{t} = N_1 \mathfrak{u}_1 + N_2 \mathfrak{u}_2, \quad (N_1, N_2 = 0, \pm 1, \pm 2, \pm \cdots),$$

where $\mathfrak{u}_1$ and $\mathfrak{u}_2$ are determined by

$$(9) \qquad \left.\begin{matrix} \mathfrak{u}_1 \mathfrak{a}_1 = 2\pi \\ \mathfrak{u}_1 \mathfrak{a}_2 = 0 \end{matrix}\right\}, \qquad \left.\begin{matrix} \mathfrak{u}_2 \mathfrak{a}_1 = 0 \\ \mathfrak{u}_2 \mathfrak{a}_2 = 2\pi \end{matrix}\right\}.$$

Further $|f(\mathfrak{z} + \mathfrak{t})| = |f(\mathfrak{z})|$ for every $\mathfrak{t}$ satisfying (8). The area p of the parallelogram of periodicity formed by the vectors $\mathfrak{u}_1$ and $\mathfrak{u}_2$ is easily calculated. If the rectangular components of $\mathfrak{a}_1$ and $\mathfrak{a}_2$ are (a_{11}, a_{12}) and (a_{21}, a_{22}) respectively, then

$$p = \frac{(2\pi)^2}{\left\| \begin{matrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{matrix} \right\|}.$$

From (8) it follows that the points (t_1, t_2) for which $|f(t_1, t_2)| = 1$, $f(t_1, t_2)$ being the c. f. of a lattice distribution, also form a set of lattice points. Let us for a moment return to Theorem 1, Chap. VII. The estimation of the remainder term $O\left(\frac{1}{n^{\frac{k}{k+1}}}\right)$, $(k = 2)$, is mainly based on Theorem 2, Chap. VII. The purport of this theorem is, roughly speaking, that the modulus of the c. f. may approach the value 1 only at points which do not lie closer than a set of lattice points. The most unfavourable case with regard to our method of proof is thus the lattice distribution and it is of special interest to study the remainder term problem in this case. In the following two sections we shall see that this problem is connected with the difficult estimation of the remainder occurring in the lattice point problem of the analytic theory of numbers.

The preceding results are easily extended to the multi-dimensional case. By a *lattice distribution* in R_k we understand a probability distribution $P(E)$, the probability mass of which is concentrated to the lattice points

$$(10) \qquad \mathfrak{z}_0 + \nu_1 \mathfrak{a}_1 + \cdots + \nu_k \mathfrak{a}_k, \quad (\nu_1, \nu_2, \ldots, \nu_k = 0, \pm 1, \pm 2, \pm \cdots),$$

where $\mathfrak{z}_0, \mathfrak{a}_1, \ldots, \mathfrak{a}_k$ are vectors in R_k and the volume of the parallelogram formed by the vectors $\mathfrak{a}$ is $\neq 0$. If $\mathfrak{a}_\nu$ has the rectangular components $(a_{\nu 1}, a_{\nu 2}, \ldots, a_{\nu k})$, this means that the determinant $A = \|a_{\mu\nu}\| \neq 0$. If $f(\mathfrak{t}) = f(t_1, t_2, \ldots, t_k)$ is the c. f. of $P(E)$, where $\mathfrak{t} = (t_1, t_2, \ldots, t_k)$, then there exist vectors $\mathfrak{u}_1, \mathfrak{u}_2, \ldots, \mathfrak{u}_k$, determined by

$$\left.\begin{array}{l} \mathfrak{u}_1\,\mathfrak{a}_1 = 2\pi \\ \mathfrak{u}_1\,\mathfrak{a}_2 = 0 \\ \cdots\cdots \\ \mathfrak{u}_1\,\mathfrak{a}_k = 0 \end{array}\right\}, \cdots\cdots, \left.\begin{array}{l} \mathfrak{u}_k\,\mathfrak{a}_1 = 0 \\ \mathfrak{u}_k\,\mathfrak{a}_2 = 0 \\ \cdots\cdots \\ \mathfrak{u}_k\,\mathfrak{a}_k = 2\pi \end{array}\right\},$$

such that every $\mathfrak{t}$ for which $|f(\mathfrak{t})| = 1$ is expressed by

$$(11) \qquad \mathfrak{t} = N_1\mathfrak{u}_1 + N_2\mathfrak{u}_2 + \cdots + N_k\mathfrak{u}_k, \quad (N_1, N_2, \ldots, N_k = 0, \pm 1, \pm 2, \pm \cdots),$$

and conversely. Furthermore $|f(\mathfrak{s} + \mathfrak{t})| = |f(\mathfrak{s})|$ for all $\mathfrak{s}$ and every $\mathfrak{t}$ satisfying (11). The volume p of the k-dimensional parallelogram of periodicity formed by the vectors $\mathfrak{u}$ is equal to

$$(12) \qquad p = \frac{(2\pi)^k}{|A|}.$$

2. **On the probability mass at a discontinuity point.** Let $X^{(1)}, X^{(2)}, \ldots, X^{(n)}$ be a sequence of independent r. v.'s in R_k with the same pr. f. $P(E)$ and the c. f. $f(t_1, t_2, \ldots, t_k)$. We further suppose that $P(E)$ is a lattice distribution defined by (10). We form the variable

$$(13) \qquad {}^{(n)}X = \frac{X^{(1)} + X^{(2)} + \cdots + X^{(n)}}{\sqrt{n}}$$

with the pr. f. $P_n(E)$ and the c. f.

$$(14) \qquad f_n(t_1, t_2, \ldots, t_k) = \left\{f\left(\frac{t_1}{\sqrt{n}}, \frac{t_2}{\sqrt{n}}, \ldots, \frac{t_k}{\sqrt{n}}\right)\right\}^n.$$

By Chap. VI $P_n(E)$ is also a lattice distribution with the point spectrum situated in

$$(15) \qquad \frac{n\,\mathfrak{z}_0 + \nu_1\mathfrak{a}_2 + \cdots + \nu_k\mathfrak{a}_k}{\sqrt{n}}, \quad (\nu_1, \nu_2, \ldots, \nu_k = 0, \pm 1, \pm 2, \pm \cdots).$$

We wish to express the probability mass $q^{(n)}(\xi_1, \xi_2, \ldots, \xi_k)$ at a discontinuity point $(\xi_1, \xi_2, \ldots, \xi_k)$ of $P_n(E)$ as a functional of $f_n(t_1, t_2, \ldots, t_k)$. By p we denote both the volume of the parallelogram of periodicity, (cf. (12)), and the region of integration formed by that parallelogram when it is moved parallel to itself so that the origin and the centre of the parallelogram coincide. By $p(\sqrt{n})$ we understand the parallelogram p magnified to the scale $\sqrt{n} : 1$.

By a proof similar to that of Lemma 2, Chap. IV, we obtain:

Lemma 1.

$$q^{(n)}(\xi_1, \xi_2, \ldots, \xi_k) = \frac{1}{p\, n^{k/2}} \int\limits_{p(\sqrt{n})} e^{-i(\xi_1 t_1 + \cdots + \xi_k t_k)} f_n(t_1, \ldots, t_k)\, d t_1 \ldots d t_k,$$

or by (12),

$$q^{(n)}(\xi_1, \xi_2, \ldots, \xi_k) = \frac{|A|}{(2\pi)^k n^{k/2}} \int\limits_{p(\sqrt{n})} e^{-i(\xi_1 t_1 + \cdots + \xi_k t_k)} f_n(t_1, \ldots, t_k)\, d t_1 \ldots d t_k.$$

Lemma 1 may be applied to the proof of

Theorem 1. *Let* $X^{(1)}, X^{(2)}, \ldots, X^{(n)}$ *be a sequence of independent r.v.'s in* R_k *with the same pr.f.* $P(E)$ *and let* $P(E)$ *be a lattice distribution defined by* (10). *Let an arbitrary r.v. of the sequence have the properties:*

1° *the mean values are equal to zero;*
2° *the dispersions are equal to* 1*;*
3° *the mixed moments of the second order are zero;*
4° *the fourth moments are finite.*

Then the probability function $P_n(E)$ *of* ${}^{(n)}X$ *is also a lattice distribution and the probability mass* $q^{(n)}(\xi_1, \xi_2, \ldots, \xi_k)$ *at a discontinuity point* $(\xi_1, \xi_2, \ldots, \xi_k)$ *of* $P_n(E)$ *is expressed by*

$$q^{(n)}(\xi_1, \xi_2, \ldots, \xi_k) = \frac{|A|}{(2\pi n)^{k/2}} \left\{ e^{-\frac{\varrho^2}{2}} - \frac{1}{6\sqrt{n}} \left(\alpha_1 \frac{\partial}{\partial \xi_1} + \cdots + \alpha_k \frac{\partial}{\partial \xi_k} \right)^3 e^{-\frac{\varrho^2}{2}} \right\} + O\left(\frac{1}{n^{\frac{k+2}{2}}} \right),$$

where

$$\varrho^2 = \xi_1^2 + \xi_2^2 + \cdots + \xi_k^2 \quad \textit{and} \quad \left(\alpha_1 \frac{\partial}{\partial \xi_1} + \cdots + \alpha_k \frac{\partial}{\partial \xi_k} \right)^3$$

is taken in the symbolic sense, (cf. (10), Chap. VI).

The proof of Theorem 1, which is an immediate consequence of Lemma 1, this chapter, and Lemma 2, Chap. VII, is similar to that of the corresponding one-dimensional case, (Theorem 5, Chap. IV).

3. **The connection with the lattice point problem.** We begin by giving an account of the lattice point problem concerning a k-dimensional ellipsoid.[1] Let $k \geq 2$ be an integer and

$$(16) \qquad Q = Q(y) = \sum_{\mu, \nu = 1}^{k} a_{\mu\nu}\, y_\mu y_\nu, \quad (a_{\mu\nu} = a_{\nu\mu}),$$

[1] For further information, cf. Jarník [1] and [2].

a definite positive quadratic form with the determinant $D = ||a_{\mu\nu}||$. The form Q is called *rational* if there exists a number α such that $a_{\mu\nu} = \alpha b_{\mu\nu}$, where the $b_{\mu\nu}$'s are integers; otherwise Q is *irrational*. For $x > 0$, $B(x) = B_Q(x)$ denotes the number of lattice points (i. e. points in R_k having integers $(m_1, m_2, \ldots, m_k)$ as coordinates) in the closed k-dimensional ellipsoid $Q(y) \leq x$. The volume of this ellipsoid is equal to

$$V(x) = V_Q(x) = \frac{\pi^{k/2}\, x^{k/2}}{\sqrt{D}\, \Gamma\left(1 + \frac{k}{2}\right)}. \tag{17}$$

We put

$$P(x) = P_Q(x) = B_Q(x) - V_Q(x), \tag{18}$$

where $P_Q(x)$ is called the lattice remainder. We also put

$$R(x) = R_Q(x) = \frac{1}{x} \int_0^x |P_Q(z)|\, dz. \tag{19}$$

For all forms Q and all $k \geq 2$ the following result holds, (Landau):

$$\begin{cases} P(x) = O\left(x^{\frac{k}{2} - \frac{k}{k+1}}\right), \\ P(x) = \Omega\left(x^{\frac{k-1}{4}}\right). \end{cases} \tag{20}$$

If Q is *rational* and $k > 4$ the true order of magnitude of $P(x)$ is known, (Landau, Walfisz, Jarník):

$$P(x) = O\left(x^{\frac{k}{2} - 1}\right) \text{ for } k > 4, \tag{21}$$

$$P(x) = \Omega\left(x^{\frac{k}{2} - 1}\right) \text{ for } k \geq 2. \tag{22}$$

Even if $k \leq 4$, estimations similar to (21) are known. Let, for instance, $k=2$ and consider the circular case: $Q = y_1^2 + y_2^2$. Then

$$\begin{cases} P(x) = O(x^{1/3}), \\ P(x) = \Omega(x^{1/4} \log^{1/4} x), \\ R(x) = O(x^{1/4}). \end{cases} \tag{23}$$

By very deep methods the exponent $\frac{1}{3}$ in (23) may be diminished a little, for example replaced by $\frac{27}{82}$, (Nieland). Later, slightly better results have been obtained. The true value is not known.

If Q has the form $a_1 y_1^2 + a_2 y_2^2 + \cdots + a_k y_k^2$ then, (JARNÍK):

$$(24)\qquad \begin{cases} R(x) = O(x^{1/4}\log^2 x), & k = 2, \\ R(x) = O(x^{1/2}\log\ x), & k = 3, \\ R(x) = O\left(x^{\frac{k}{2}-1}\right), & k \geq 4. \end{cases}$$

The estimation (20) always holds, the estimations (21), (22) and (24) are proved only for special cases and lie very deep.

Finally we remark that the number of integer solutions $(y_1, y_2, \ldots, y_k)$ of the equation

$$(25)\qquad m = y_1^2 + y_2^2 + \cdots + y_k^2,$$

($k > 4$, m a positive integer), is asymptotically equal to

$$(26)\qquad \text{const. } m^{\frac{k}{2}-1}, \quad (m \to \infty).$$

(HARDY, MORDELL).

Now consider a sequence of independent r. v.'s $X^{(1)}, X^{(2)}, \ldots, X^{(n)}$ in R_k, ($k \geq 2$), all having the same pr. f. $P(E)$ and suppose that $P(E)$ is a *lattice distribution* with the probability mass concentrated in the points

$$(27)\qquad \nu_1 \mathfrak{a}_1 + \nu_2 \mathfrak{a}_2 + \cdots + \nu_k \mathfrak{a}_k, \qquad (\nu_1, \nu_2, \ldots, \nu_k = 0, \pm 1, \pm 2, \pm \cdots).$$

The vector $\mathfrak{a}_\nu$ has the rectangular components $(a_{\nu 1}, a_{\nu 2}, \ldots, a_{\nu k})$, $(\nu = 1, 2, \ldots, k)$. Further we suppose that the determinant

$$A = ||a_{\nu\mu}|| \neq 0.$$

Let an arbitrary r. v. X of the sequence have the properties:

$$(28)\qquad \begin{cases} 1^\circ & \text{the mean values are equal to zero;} \\ 2^\circ & \text{the dispersions are equal to } 1; \\ 3^\circ & \text{the mixed moments of the second order are equal to zero;} \\ 4^\circ & \text{the third moments are equal to zero;} \\ 5^\circ & \text{the fourth moments } \beta_\nu^4 \text{ are finite and } \beta_4 = \sum_{\nu=1}^{k} \beta_\nu^4. \end{cases}$$

We form the variable

$$^{(n)}X = \frac{X^{(1)} + X^{(2)} + \cdots + X^{(n)}}{\sqrt{n}}$$

and denote by $\mu_n(a)$ the probability of $^{(n)}X$ belonging to a sphere with its centre at $(0, 0, \ldots, 0)$ and of radius a. Furthermore, as usual,

$$\psi(a, k) = \frac{1}{(2\pi)^{k/2}} \int\limits_{x_1^2 + \cdots + x_k^2 \leq a^2} e^{-\frac{1}{2}(x_1^2 + \cdots + x_k^2)} dx_1 \ldots dx_k.$$

By Theorem 1, Chap. VII,

(29) $$|\mu_n(a) - \psi(a, k)| \leq c \cdot \frac{\beta_4^{3/2}}{n^{\frac{k}{k+1}}},$$

c being a constant. However, in the case under consideration we may find an explicit expression of $\mu_n(a)$, thus making it possible to discuss the remainder term in greater detail.

As before the pr. f. $P_n(E)$ of ${}^{(n)}X$ is a lattice distribution with the point spectrum belonging to

(30) $$\frac{\nu_1 \mathfrak{a}_1 + \nu_2 \mathfrak{a}_2 + \cdots + \nu_k \mathfrak{a}_k}{\sqrt{n}}, \quad (\nu_1, \nu_2, \ldots, \nu_k = 0, \pm 1, \pm 2, \pm \cdots).$$

If $(\xi_1, \xi_2, \ldots, \xi_k)$ is a discontinuity point of $P_n(E)$ and

(31) $$\varrho^2 = \xi_1^2 + \xi_2^2 + \cdots + \xi_k^2,$$

then by Theorem 1 and (28: 4°) the probability mass at $(\xi_1, \xi_2, \ldots, \xi_k)$ is expressed by

(32) $$q^{(n)}(\xi_1, \xi_2, \ldots, \xi_k) = \frac{|A|}{(2\pi n)^{k/2}} e^{-\frac{\varrho^2}{2}} + O\left(\frac{1}{n^{\frac{k+2}{2}}}\right),$$

and $\mu_n(a)$ is equal to the sum

(33) $$\sum_{\xi_1^2 + \cdots + \xi_k^2 \leq a^2} q^{(n)}(\xi_1, \xi_2, \ldots, \xi_k).$$

Consider the quadratic form

(34) $$Q(y) = (a_{11} y_1 + a_{21} y_2 + \cdots + a_{k1} y_k)^2 + \cdots + (a_{1k} y_1 + a_{2k} y_2 + \cdots + a_{kk} y_k)^2$$

with the determinant

(35) $$D = A^2,$$

the coefficients $a_{\nu\mu}$ of which are the components of the vectors $\mathfrak{a}$ in (30). Let the functions $B(x)$, $V(x)$, $P(x)$ and $R(x)$ be defined by (17)—(19), Q being the form (34). By $U(x)$ we denote the number of integer solutions $(y_1, y_2, \ldots, y_k)$ of the equation

$$Q(y) = x.$$

From (30) and (34) it is seen that the number of discontinuity points of $P_n(E)$ lying on the surface of a sphere with its centre at $(0, 0, \ldots, 0)$ and of radius ϱ, is equal to $U(\varrho^2 n)$. Hence from (32) the contribution of the probability masses belonging to this surface is equal to

$$\frac{|A|}{(2\pi n)^{k/2}} e^{-\frac{\varrho^2}{2}} U(\varrho^2 n) + O\left(\frac{U(\varrho^2 n)}{n^{\frac{k+2}{2}}}\right), \tag{36}$$

or from (33)

$$\mu_n(a) = \sum_{\varrho \leq a} \left\{ \frac{|A|}{(2\pi n)^{k/2}} e^{-\frac{\varrho^2}{2}} U(\varrho^2 n) + O\left(\frac{U(\varrho^2 n)}{n^{\frac{k+2}{2}}}\right)\right\}.$$

Since $B(x) = \sum_{y \leq x} U(y)$, we have:

$$\mu_n(a) = \int_0^a \frac{|A| e^{-\frac{\varrho^2}{2}}}{(2\pi n)^{k/2}} d\, B(\varrho^2 n) + O\left\{\frac{B(a^2 n)}{n^{\frac{k+2}{2}}}\right\}. \tag{37}$$

Here we introduce

$$B(\varrho^2 n) = V(\varrho^2 n) + P(\varrho^2 n) = \frac{\pi^{k/2} n^{k/2} \varrho^k}{|A| \Gamma\left(1 + \frac{k}{2}\right)} + P(\varrho^2 n),$$

and obtain after a simple calculation:

$$\mu_n(a) - \psi(a, k) = \frac{|A|}{(2\pi n)^{k/2}} \int_0^a e^{-\frac{\varrho^2}{2}} d\, P(\varrho^2 n) + O\left\{\frac{B(a^2 n)}{n^{\frac{k+2}{2}}}\right\}.$$

Integrating by parts we have:

$$\begin{aligned} \mu_n(a) - \psi(a, k) = \frac{|A|}{(2\pi n)^{k/2}} e^{-\frac{a^2}{2}} P(a^2 n) + \\ + \frac{|A|}{(2\pi n)^{k/2}} \int_0^a \varrho e^{-\frac{\varrho^2}{2}} P(\varrho^2 n)\, d\varrho + O\left\{\frac{B(a^2 n)}{n^{\frac{k+2}{2}}}\right\}. \end{aligned} \tag{38}$$

We now use the immediate estimation

$$B(x) = O(x^{k/2})$$

and observe that

$$\left| \int_0^a \varrho e^{-\frac{\varrho^2}{2}} P(\varrho^2 n)\, d\varrho \right| \leq \int_0^a \varrho |P(\varrho^2 n)|\, d\varrho = \frac{1}{2n} \int_0^{a^2 n} |P(y)|\, dy = \frac{a^2}{2} R(a^2 n).$$

Hence from (38):

$$(39) \qquad \mu_n(a) - \psi(a, k) = \frac{|A|}{(2\pi n)^{k/2}} e^{-\frac{a^2}{2}} P(a^2 n) + \theta_1 \frac{a^2}{n^{k/2}} R(a^2 n) + \theta_2 \frac{a^k}{n},$$

where $|\theta_1|$ and $|\theta_2|$ are bounded by constants independent of a and n[1]. The radius a occurs in the remainder term of formula (39). This is, however, of no importance with regard to our purpose, which is to study the order of magnitude in n. In the sequel we suppose that a is bounded. The relations (38) and (39) express the connection between the discontinuities of $\mu_n(a)$ and the remainders of the lattice point problem.

As we have mentioned earlier, the estimation (20) is the only one valid in the general case. If

$$P(x) = O\left(x^{\frac{k}{2} - \frac{k}{k+1}}\right),$$

then obviously

$$R(x) = O\left(x^{\frac{k}{2} - \frac{k}{k+1}}\right).$$

According to this we obtain from (39):

$$(40) \qquad \mu_n(a) - \psi(a, k) = O\left(\frac{1}{n^{\frac{k}{k+1}}}\right),$$

or the same order of magnitude as in (29). Conversely, it is possible to prove (20) by methods similar to those of the proof of Theorem 1, Chap. VII. I confine myself to this indication. Thus we may say, that Theorem 1, Chap. VII, and the estimation (20) are of the same depth. The estimations (21) and (24) lie deeper but are only valid in special cases. If $Q(y)$ in (34) is rational and $k > 4$, then from (21) and (39) we obtain the improvement

$$(41) \qquad \mu_n(a) - \psi(a, k) = O\left(\frac{1}{n}\right).$$

Hitherto we have only obtained O-estimations. Is it not possible also to get Ω-estimations? This does not follow from (39) since the remainder terms may eventually compensate each other. Consider, however, the following example: $Q(y)$ has the form $y_1^2 + y_2^2 + \cdots + y_k^2$ and $k > 4$. If S_ϱ is a sphere with its centre at $(0, 0, \ldots, 0)$ and of radius ϱ, having discontinuity points of $P_n(E)$ on its sur-

[1] Since $B(x)$ is not generally zero for $x = 0$ we suppose that $a^2 n \geq 1$, or else a trivial change in (39) has to be performed.

face, then from (25), (26) and (36) the probability mass on S_ϱ is asymptotically equal to

$$\text{const. } \varrho^{2\left(\frac{k}{2}-1\right)} e^{-\frac{\varrho^2}{2}} \frac{n^{\frac{k}{2}-1}}{n^{k/2}} = \text{const. } \varrho^{k-2} e^{-\frac{\varrho^2}{2}} \frac{1}{n}. \tag{42}$$

From (41) it follows that the remainder term is $O\left(\frac{1}{n}\right)$ and from (42) that it cannot be improved. Thus, even if all moments are finite, *an expansion like*

$$\mu_n(a) = \psi_1(a) + \frac{\psi_2(a)}{n} + o\left(\frac{1}{n}\right),$$

ψ_1 *and* ψ_1 *being continuous, is generally impossible.* There must enter into the expansion a discontinuous function which, however, is much more complicated than in the one-dimensional case.

To conclude we give an example which well illustrates the connection between our remainder term problem in the two-dimensional case and the lattice point problem for a circle. Consider the two-dimensional lattice distribution having the probability mass $\frac{1}{4}$ at the points $(\pm 1, \pm 1)$. Obviously the conditions (28) are satisfied. If n is even, it is easily seen that the point spectrum of $P_n(E)$ is situated in

$$\left(\nu_1 \cdot \frac{2}{\sqrt{n}};\ \nu_2 \cdot \frac{2}{\sqrt{n}}\right), \qquad (\nu_1, \nu_2 = 0, \pm 1, \pm 2, \pm \cdots).$$

Obviously

$$Q(y) \equiv 4y_1^2 + 4y_2^2.$$

We prefer to use the function

$$Q_1(y) \equiv y_1^2 + y_2^2.$$

Hence $P_Q(x) = P_{Q_1}(x/4)$, $R_Q(x) = R_{Q_1}(x/4)$, where P_{Q_1} and R_{Q_1} are the remainders of the lattice point problem for a circle (cf. (23)). Then from (39):

$$\mu_n(a) - \psi(a, 2) = \frac{2}{\pi} \frac{e^{-\frac{a^2}{2}}}{n} P_{Q_1}\left(\frac{a^2 n}{4}\right) + \theta_1 \frac{a^2}{n} R_{Q_1}\left(\frac{a^2 n}{4}\right) + \theta_2 \frac{a^2}{n}.$$

According to (23) we have:

$$\mu_n(a) - \psi(a, 2) = \frac{2}{\pi n} e^{-\frac{a^2}{2}} P_{Q_1}\left(\frac{a^2 n}{4}\right) + \theta \frac{a^{5/3}}{n^{5/6}}, \tag{43}$$

where $|\theta|$ is bounded. Using (23) again we obtain

$$\mu_n(a) - \psi(a, 2) = O\left(\frac{1}{n^{2/3}}\right), \tag{44}$$

or the order of magnitude of Theorem 1, Chap. VII. As was mentioned in connection with the circular lattice point problem, it is possible to replace the exponent $\frac{2}{3}$ by a somewhat greater number, but then very deep methods must be used. For instance, according to the result of NIELAND, we have

$$\mu_n(a) - \psi(a, 2) = O\left(\frac{1}{n^{1-\frac{27}{82}}}\right).$$

On the other hand it follows from $P_{Q_1}(x) = \Omega(x^{1/4} \log^{1/4} x)$ (cf. (23)), that the remainder term of (44) cannot be replaced by $O\left(\frac{1}{n^{3/4}}\right)$. *Thus in the two-dimensional case it is impossible to attain a better general result than*

$$|\mu_n(a) - \psi(a, 2)| \leq \frac{\text{const.}}{n^{\alpha}},$$

where $\frac{2}{3} \leq \alpha < \frac{3}{4}$.

Remark. We have hitherto exclusively studied the probability of ${}^{(n)}X$ belonging to a sphere about the origin. It must be observed that the remainder term is dependent on the region considered. In the two-dimensional case we obtained a remainder term of order of magnitude $\frac{1}{n^{2/3}}$ if the region was a circle, but if the region is a square with its centre at the origin and the sides parallel with the coordinate axeses, there may occur discontinuities of order of magnitude $\frac{1}{\sqrt{n}}$. This is, for instance, the case if $P(E)$ is a lattice distribution with the probability mass $\frac{1}{4}$ at $(\pm 1, \pm 1)$.

Bibliography.

S. BERNSTEIN, [1] Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes. *Math. Ann.* 97 (1927).

A. C. BERRY, [1] The accuracy of the Gaussian approximation to the sum of independent variates. *Trans. Amer. Math. Soc.* 49 (1941).

A. BEURLING, [1] Sur les intégrales de Fourier absolument convergentes et leur application à une transformation fonctionnelle. *Neuvième congrès des mathématiciens scandinaves.* Helsingfors 1938.

S. Bochner, [1] *Vorlesungen über Fouriersche Integrale.* Leipzig 1932.

——, [2] A theorem on Fourier-Stieltjes integrals. *Bull. Amer. Math. Soc.* 40 (1934).

H. Bohr, [1] Ein allgemeiner Satz über die Integration eines trigonometrischen Polynoms. *Prace mat.-fiz.* 43 (1935).

H. Bruns, [1] *Wahrscheinlichkeitsrechnung und Kollektivmasslehre.* Leipzig 1906.

T. Carleman, [1] *Les fonctions quasi analytiques.* Collection de monographies sur la théorie des fonctions, publiée sous la direction de M. Émile Borel. Paris 1926.

——, [2] *Sur les équations intégrales . . .* Uppsala Universitets Årsskrift 1923.

C. V. L. Charlier, [1] Über das Fehlergesetz. *Ark. Mat., Astr. o. Fysik* 2 (1905—06).

——, [2] Die strenge Form des Bernoullischen Theorems. *Ark. Mat., Astr. o. Fysik* 5 (1909).

——, [3] Contributions to the mathematical theory of statistics 3. *Ark. Mat., Astr. o. Fysik* 8 (1912—13).

——, [4] Contributions to the mathematical theory of statistics 5 and 6. *Ark. Mat., Astr. o. Fysik* 9 (1913—14).

C. C. Craig, [1] On the composition of dependent elementary errors. *Ann. Math.* 33 (1932).

H. Cramér, [1] Das Gesetz von Gauss und die Theorie des Risikos. *Skand. Aktuarietidskr.* 6 (1923).

——, [2] On some classes of series used in mathematical statistics. *Sixième congrès des mathématiciens scandinaves.* Copenhague 1925.

——, [3] On the composition of elementary errors, 1. *Skand. Aktuarietidskr.* 11 (1928).

——, [4] Su un teorema relativo alla legge uniforme dei grandi numeri. *Giornale dell'Istituto Italiano degli Attuari* (1934).

——, [5] *Random variables and probability distributions.* Cambridge Tracts No. 36 (1937).

——, [6] *Sur un nouveau théorème-limite de la théorie des probabilités.* Actualités scientifiques et industrielles No. 736. Paris 1938.

——, [7] Entwicklungslinien der Wahrscheinlichkeitsrechnung. *Neuvième congrès des mathématiciens scandinaves.* Helsingfors 1938.

F. Y. Edgeworth, [1] The law of error. *Cambridge Phil. Trans.* 20 (1904).

C.-G. Esseen, [1] On the Liapounoff limit of error in the theory of probability. *Ark. Mat., Astr. o. Fysik* 28 A (1942).

——, [2] Determination of the maximum deviation from the Gaussian law. *Ark. Mat., Astr. o. Fysik* 29 A (1943).

B. Gnedenko, [1] Sur les fonctions caractéristiques. *Bull. de l'Université d'État à Moscou.* A, Vol. 1, Fasc. 5 (1937).

G. H. Hardy—J. E. Littlewood—G. Pólya, [1] *Inequalities.* Cambridge 1934.

E. Hille—J. D. Tamarkin, [1] On the absolute integrability of Fourier transforms. *Fundamenta Math.* 25 (1935).

A. E. Ingham, [1] A note on Fourier transforms. *J. London Math. Soc.* 9 (1934)

V. Jarník, [1] Über Gitterpunkte in mehrdimensionalen Ellipsoiden. *Math. Ann.* 100 (1928).

V. JARNÍK, [2] Über die Mittelwertsätze der Gitterpunktlehre I, II, *Math. Z.* 33 (1931).

B. JESSEN—A. WINTNER, [1] Distribution functions and the Riemann zeta function. *Trans. Amer. Math. Soc.* 38 (1935).

A. JOURAVSKY, [1] Sur la théorème limite du calcul des probabilités (in Russian). *Travaux de l'Institut Physico-Mathématique Stekloff.* Section Math. 1933.

A. KHINTCHINE, [1] *Asymptotische Gesetze der Wahrscheinlichkeitsrechnung.* Berlin 1933.

A. KOLMOGOROFF, [1] *Grundbegriffe der Wahrscheinlichkeitsrechnung.* Berlin 1933.

P. S. LAPLACE, [1] *Théorie analytique des probabilités.* Oeuvres VII, Paris 1847.

P. LÉVY, [1] *Calcul des probabilités.* Paris 1925.

——, [2] *Théorie de l'addition des variables aléatoires.* Monographies des probabilités, publiées sous la direction de M. Émile Borel. Fasc. 1, Paris 1937.

——, [3] Sur les séries dont les termes sont des variables éventuelles indépendentes. *Studia Mathematica* 3 (1931).

A. LIAPOUNOFF, [1] Sur une proposition de la théorie des probabilités. *Bull. Acad. Sci. St-Pétersbourg*, (5), 13 (1900).

——, [2] Nouvelle forme du théorème sur la limite de probabilité. *Mém. Acad. Sci. St-Pétersbourg*, (8), 12 (1901).

J. W. LINDEBERG, [1] Eine neue Herleitung des Exponentialgesetzes der Wahrscheinlichkeitsrechnung. *Math. Z.* 15 (1922).

J. E. LITTLEWOOD, [1] On the Fourier coefficients of functions of bounded variation. *Quart. J. Math. Oxford Ser.* 7 (1936).

D. MENCHOFF, [1] Sur l'unicité du développement trigonométrique. *C. R. Acad. Sci. Paris* 163 (1916).

B. v. SZ. NAGY—A. STRAUSZ, [1] Über einen Satz von H. Bohr. *Math. u. Naturwiss. Anz. Ungar. Akad. Wiss.* 57 (1938).

K. PEARSON, [1] On deviations from the probable in a correlated system of variables. *Phil. Mag.* 50 (1900).

J. RADON, [1] Theorie und Anwendung der absolut additiven Mengenfunktionen. *Sitzungsber. Akad. Wien* 122 (1913).

F. RIESZ, [1] Über die Fourierkoeffizienten einer stetigen Funktion von beschränkter Schwankung. *Math. Z.* 2 (1918).

L. SCHWARTZ, [1] Sur le module de la fonction caractéristique du calcul des probabilités. *C. R. Acad. Sci. Paris* 212 (1941).

E. C. TITCHMARSH, [1] *Introduction to the theory of Fourier integrals.* Oxford 1937.

G. VALIRON, [1] Sur la formule d'interpolation de Lagrange. *Bull. Sci. Math.* 49 (1925).

G. N. WATSON, [1] *Theory of Bessel functions.* Cambridge 1922.

A. WINTNER, [1] On the addition of independent distributions. *Amer. J. Math.* 56 (1934).

——, [2] On a class of Fourier transforms. *Amer. J. Math.* 58 (1936).