1961

Limit theorems in probability and statistics

Michael Gannaway Billings

The University of Montana

LIMIT THEOREMS IN PROBABILITY AND STATISTICS

by

MICHAEL GANNAWAY BILLINGS

B.A. Montana State University, 1958

Presented in partial fulfillment of the requirements for degree of

Master of Arts

MONTANA STATE UNIVERSITY

1961

Approved by:

Howard E. Reinhardt

Chairman, Board of Examiners

Ellis Waldron

Dean, Graduate School

MAY 26 1961

Date

PREFACE

The author wishes to express his gratitude to those of the Mathematics Department at the University of Montana whose guidance made possible the completion of this work. I especially thank Professor Howard Reinhardt for his unlimited patience and learned instruction during the past two years. Also, my sincerest thanks go to Professors William Ballard and Joseph Hashisaki for their critical reading of the original drafts of this paper. Finally, I wish to express my gratitude to my wife. To her fell the task of transcribing my handwritten mathematics into typewritten legibility.

TABLE OF CONTENTS

INTRODUCTION

A mathematical theory, to be of practical value, must obviously provide theoretical counterparts for the real phenomena which it purports to describe. Mathematical probability theory is no exception. Since probabilistic methods are now being used extensively in varied applications to real situations, it seems reasonable to inquire as to just what links the theory of probability to the real world. The probability theoretic concept that provides the necessary link with reality is the Weak Law of Large Numbers, sometimes called the "law of averages," and is formulated in mathematically precise terms in Theorem III.10, Khintchine's Theorem, in the sequel.

Essentially, what the Weak Law says in real terms is this: If we perform an experiment a large number of times and observe the frequency of the occurrence of a particular event, then there is a very large probability that the fraction of times the event occurs will differ only slightly from the true probability of the event. Thus, if we toss a coin a large number of times and observe the fraction of the tosses which result in the occurrence of heads, we can be quite sure that this fraction will differ little from the true probability that a head will occur on a single toss.

It is not, however, the intent of this paper to consider the application of the mathematical "law of averages" to the real world. The basic purpose of our discussions of the Weak Law and related theorems is to indicate the conditions under which they apply to theoretical situations. Nevertheless, the importance of the theorems to be considered derives not only from the contribution they make to probability theory, but also from their wide applicability in statistical analyses of actual situations.

As evidence of important theoretical and real implications we consider briefly two of the theorems proved in Chapter III. Theorem III.15, the Lindberg-Lévy Theorem, shows that if we have a sequence of identically distributed independent random variables, then, subject to certain conditions, sums and means of sums of these variables have asymptotically normal distributions. This theorem enables us to draw inferences based on a consideration of the means, without unduly concerning ourselves with the distribution of the variables themselves.

Theorem III.17, the Liapounoff Theorem, states that, under certain very general conditions, the sums of random variables are approximately normally distributed, even if the variables do not necessarily have identical distributions. For example, I.Q. scores are based on answers to a sequence of questions, each of which is worth a specified number of points. An individual test score is the sum of the points a person receives for each question. If the results of different questions are independent, then the Liapounoff Theorem says that the test scores are approximately normally distributed, despite the fact that the number of points a person scores on question one, for example, may have a decidedly different distribution than the number scored on question two.

In the discussions that follow, the author assumes a knowledge of probability and statistics at the level of Feller[1] [3], Cramér [1], or comparable works; and a familiarity with measure and integration theory at the level of Munroe [6]. The notation employed in this paper conforms to the notation used in [2] and [3].

For ease of reference, several theorems and definitions are included in this introduction.

Let S be a measure space, with measure μ. (In the literature of probability theory, S is called a sample space if $\mu(S)=1$.) We define a random variable X on S as a real valued measurable function whose domain is S.

The random variable X induces a measure P on the class of measurable sets of real numbers; for the inverse image under X of any measurable set T of real numbers is a μ-measurable set of S. We define the measure of T by $P(T) = \mu[X^{-1}(T)]$.

---

[1]Numbers in square brackets refer to the numbers of the references in the list of references cited on page 42.

Let P be the measure defined above. We associate with P a non-decreasing point function F(x), called a distribution function, such that for any finite interval (a,b] we have $F(b) - F(a) = P(a < x \leq b)$. It can be shown that F is uniquely determined except for an additive constant; further, F(x) is everywhere continuous to the right. If P is a probability measure, F may be chosen such that $\lim_{x\to\infty} F(x) = 1$, $\lim_{x\to-\infty} F(x) = 0$.

Thus, with each random variable X we associate a distribution F(x) defined by $F(x) = P[X \leq x]$.

Theorem A ([2] page 66): If $\lim_{v} g_v(x) = g(x)$ exists, except on a set of P-measure zero, in a measurable set S of real numbers, and if $|g_v(x)| < G(x)$ for all v and all x in S, where G(x) is a function integrable with respect to F on S, then g(x) is integrable with respect to F on S, and

$$\lim_{v\to\infty} \int_S g_v(x)dF = \int_S g(x)dF.$$

Theorem B ([2] pages 73-74): If F and G are non-decreasing functions in [a,b], have no common discontinuities, and have only a finite number of discontinuities all of which are in (a,b), then

$$\int_a^b dFG = \int_a^b FdG + \int_a^b GdF.$$

Theorem C ([8] page 138): Let $g(x)$ be continuous on $[a,b]$ and let $\{F_n\}$ be a sequence of functions such that F_n is a distribution function for each n, and $\lim_n F_n(x) = F(x)$. Then

$$\lim_{n\to\infty} \int_a^b g(x)dF_n = \int_a^b g(x)dF.$$

Definition D: An interval (a,b) will be called a continuity interval of the function $F(x)$ if a and b are continuity points of $F(x)$.

CHAPTER I

We consider in this chapter some of the modes of convergence of sequences of random variables.

Let $\{X_n\}$ be a sequence of random variables defined on the same sample space S for which a probability measure P has been defined. Let X be another random variable defined on S.

Definition I.1: We say that $\{X_n\}$ converges to X with probability one if

$$P[\lim_{n \to \infty} X_n = X] = 1.$$

Definition I.2: The sequence $\{X_n\}$ converges in mean square to the random variable X if $\lim_{n \to \infty} E[(X_n - X)^2] = 0.$

Definition I.3: The sequence $\{X_n\}$ converges in probability to the random variable X if, for every $\epsilon > 0$,

$$\lim_{n \to \infty} P[|X_n - X| > \epsilon] = 0.$$

An important property of convergence in probability is that no moments need exist before it can be considered, which is quite evidently not the case with convergence in mean square. (Our interest centers on convergence in probability and convergence in distribution--defined in 1.4.)

In measure theory language, convergence with probability one is almost everywhere convergence, while convergence in probability is convergence in measure.

Applying one form of Chebychev's inequality we obtain that, for any $\epsilon > 0$,

$$P\left[\,|X_n - X| > \epsilon\,\right] \leq \frac{1}{\epsilon^2} \cdot E\left[(X_n - X)^2\right].$$

Thus convergence in mean square implies convergence in probability.

We are primarily concerned with convergence in distribution, defined as follows:

Definition I.4: A sequence of distribution functions $\{F_n(x)\}$ is said to be convergent if there exists a non-decreasing function $F(x)$ such that

$$\lim_{n \to \infty} F_n(x) = F(x)$$

in every continuity point of $F(x)$. (This form of convergence is sometimes called essential convergence.)

Theorem I.5 [2]: Let $\{X_n\}$ be a sequence of random variables with corresponding distribution functions $\{F_n(x)\}$. Suppose $\lim_{n \to \infty} F_n(x) = F(x)$ is a distribution function. Let Y_n be another sequence of random variables, and suppose Y_n converges in probability to a constant c. Put

$$\xi_n = X_n + Y_n,\ \eta_n = X_n \cdot Y_n,\ \gamma_n = \frac{X_n}{Y_n}\,.$$

Then the distribution function of ξ_n tends to $F(x-c)$. Further, if $c > 0$, the distribution function of η_n tends to $F(\frac{x}{c})$, while that of γ_n tends to $F(cx)$. If $c < 0$, the distribution function of η_n tends to $1 - F(\frac{x}{c})$, while that of γ_n tends to $1 - F(cx)$.

Proof: We prove the assertion for γ_n, $c > 0$. The other proofs are similar. Let x be a continuity point of

$F(cx)$, and denote by $F_n(x)$ the distribution function of γ_n; i.e. $F_n(x) = P(\gamma_n \leq x)$. Then we must show that

$$\lim_{n\to\infty} F_n(x) = \lim_{n\to\infty} P[\frac{X_n}{Y_n} \leq x] = F(cx).$$

We consider the set S_n of all points in the (X_n, Y_n) plane such that $\frac{X_n}{Y_n} \leq x$. This set can be written as the union of two disjoint sets, S_{n_1} and S_{n_2}, defined by the inequalities

$$S_{n_1}\text{: } \frac{X_n}{Y_n} \leq x, |Y_n - c| \leq \epsilon;$$

$$S_{n_2}\text{: } \frac{X_n}{Y_n} \leq x, |Y_n - c| > \epsilon.$$

Thus,

$$P(S_n) = P(S_{n_1}) + P(S_{n_2}).$$

Clearly, S_{n_2} is a subset of the set of points on the plane for which $|Y_n - c| > \epsilon$. By hypothesis, $\lim_{n\to\infty} P[|Y_n - c| > \epsilon] = 0$ for any $\epsilon > 0$. Hence for any $\epsilon > 0$, $P(S_{n_2}) \to 0$ with increasing n.

Further, it is seen that $P(S_{n_1})$ is enclosed between the bounds $P[X_n \leq (c-\epsilon)x, |Y_n - c| \leq \epsilon]$ and $P[X_n \leq (c+\epsilon)x, |Y_n - c| \leq \epsilon]$. (Figure 1).

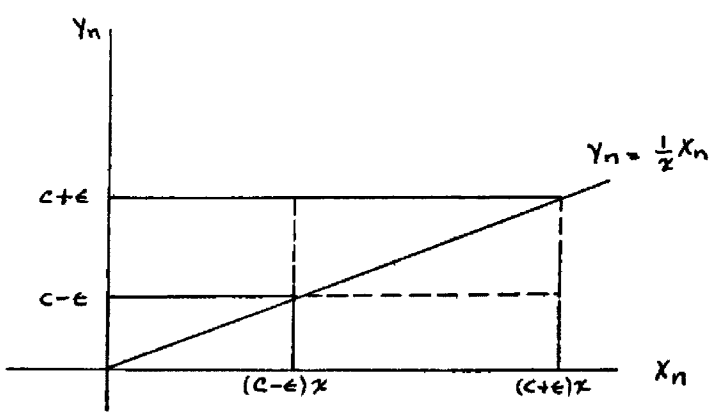

Figure 1

Now $P[X_n \leq (c \pm \epsilon)x] = F_n[(c \pm \epsilon)x]$. Each of these bounds differs from $F_n[(c \pm \epsilon)x]$ by a quantity less than $P[|Y_n - c| > \epsilon]$. For, we have that $0 \leq F_n[(c - \epsilon)x] - P[X_n \leq (c - \epsilon)x, |Y_n - c| > \epsilon] = P[X_n \leq (c - \epsilon)x, |Y_n - c| > \epsilon] \leq P[|Y_n - c| > \epsilon]$. A similar consideration gives the corresponding relationship for $F_n[(c + \epsilon)x]$. But $P[|Y_n - c| > \epsilon]$ tends to zero for any ϵ by hypothesis; hence we may make $P(S_{n_1})$ as close as we please to $F(cx)$, and the theorem is proved.

Theorem I.6 [2] : Every sequence $\{F_n(x)\}$ of distribution functions contains a convergent subsequence. The limit function $F(x)$ can always be determined so as to be everywhere continuous on the right.

Proof: Let r_1, r_2,....be an enumeration of the rationals, and consider the sequence $\{F_n(r_1)\}$. Since $0 \leq F_n(r_1) \leq 1$, the Bolzano-Weierstrass property of the real numbers tells us that the sequence has a limit point. Thus $\{F_n(r_1)\}$ contains a convergent subsequence. This is the same as saying that the sequence $\{F_n(x)\}$ contains a subsequence Z_1 which converges for the particular value $x = r_1$. By the same argument, Z_1 contains a subsequence Z_2 which converges for $x = r_1$ and $x = r_2$. We find that similarly, Z_2 contains a subsequence Z_3 convergent for $x = r_1$, $x = r_2$ and $x = r_3$. We continue in this manner, obtaining a contracting sequence $\{Z_n\}$ such that Z_n always contains a subsequence Z_{n+1} convergent for $x = r_1, \ldots,$ $x = r_{n+1}$. We may form a sequence Z whose nth element is the nth element of Z_n; clearly, Z converges for all rational values r_i of x.

Reindexing, for simplicity, let the members of Z be $F_1(x)$, $F_2(x)$,.., and put $\lim_{n\to\infty} F_n(r_i) = c_i$, i = 1, 2,... We always will have $c_i \leq 1$, and since F_n is a non-decreasing function for every n, we will have $F_n(r_i) \leq F_n(r_k)$ if $r_i \leq r_k$, hence that $\lim_{n\to\infty} F_n(r_i) \leq \lim_{n\to\infty} F_n(r_k)$; i.e. $c_i \leq c_k$ whenever $r_i \leq r_k$.

We now define a function F(x) for all real x by

$$F(x) = \inf c_i \text{ for all } r_i > x.$$

That F(x) is bounded follows from the fact that $\{c_i\}$ is bounded. That F(x) is monotone non-decreasing is seen from the following: Let x' > x. Then $\{r_i | r_i > x\}$ contains $\{r_i | r_i > x'\}$. Hence $F(x) = \{\inf c_i | r_i > x\} \leq \{\inf c_i | r_i > x'\} = F(x')$, which is the desired result.

F(x) is also everywhere continuous to the right. To show this, let $\epsilon > 0$ be given. Since $F(x_o) = \{\inf c_i | r_i > x_o\}$, we can choose r_{i_o} such that $c_{i_o} - F(x_o) < \epsilon$. Let $\delta = r_{i_o} - x_o$; when $x - x_o < \delta$, $x \in (x_o, r_{i_o})$, and, since F(x) is monotone non-decreasing, $F(x_o) \leq F(x) \leq c_{i_o}$. Thus, $F(x) - F(x_o) < \epsilon$, and right continuity is established.

We now show that in every continuity point of F(x), we have

$$\lim_{n\to\infty} F_n(x) = F(x),$$

so that the subsequence Z is convergent. If x is a continuity point of F(x), then for a given $\epsilon > 0$ we may choose

$h > 0$ such that

$$F(x+h) - F(x) < \frac{\epsilon}{2},$$

and

$$F(x) - F(x-h) < \frac{\epsilon}{2}.$$

Let r_i and r_k be rational points situated in the open intervals $(x-h,x)$ and $(x,x+h)$ respectively; we find, making use of the monotonicity of F, that

$$F(x-h) \leq c_i \leq F(x) \leq c_k \leq F(x+h).$$

For every n, we have

$$\text{(I.7)} \qquad F_n(r_i) \leq F_n(x) \leq F_n(r_k),$$

by the monotonicity of F_n; but

$$\lim_{n\to\infty} F_n(r_i) = c_i, \text{ and } \lim_{n\to\infty} F_n(r_k) = c_k.$$

Since $F(x+h) - F(x-h) < \epsilon$ and $F(x-h) \leq c_i \leq F(x) \leq c_k \leq F(x+h)$, $c_i - c_k < \epsilon$. By (I.7) we have

$$\lim_{n\to\infty} F_n(r_i) \leq \lim_{n\to\infty} F_n(x) \leq \lim_{n\to\infty} F_n(r_k),$$

or

$$c_i \leq \lim_{n\to\infty} F_n(x) \leq c_k.$$

Thus, both $F(x)$ and $\lim_{n\to\infty} F_n(x)$ lie between c_i and c_k, hence

$$\left|F(x) - \lim_{n\to\infty} F_n(x)\right| < \epsilon;$$

but ϵ is arbitrary, so the sequence Z is convergent.

CHAPTER II

In our discussion of characteristic functions and their application we shall encounter integrals of the type

$$(II.1) \qquad u(t) = \int_S g(x,t)dF(x),$$

where t is a real parameter, S a given measurable set. We shall require certain theorems concerning continuity, differentiation, and integration of such functions of t. In what follows we assume that g(x,t) is a complex valued function and that, for each fixed t considered, the real and imaginary parts of g(x,t) are Borel measurable functions of x which are integrable over S with respect to F(x). Proofs of the following three theorems can be found in [2], pages 66-70.

Theorem II.2 - Continuity: If, for almost all values of x in S, the function g(x,t) is continuous with respect to t at t_o, and if for all t in some neighborhood of t_o we have $|g(x,t)| < G_1(x)$, then u(t) (II.1) is continuous for $t = t_o$; i.e.,

$$\lim_{t \to t_o} \int_S g(x,t)dF(x) = \int_S g(x,t_o)dF(x).$$

Theorem II.3 - Differentiation: If, for almost all values of x in S and for a fixed value of t, the following conditions are satisfied:

1) The partial derivative $\frac{\partial g(x,t)}{\partial t}$ exists,

2) $\left|\frac{g(x,t+h)-g(x,t)}{h}\right| < G_2(x)$ for $0 < |h| < h_o$, h_o independent of x,

then

$$u'(t) = \frac{d}{dt}\int_S g(x,t)dF(x)$$

exists and equals
$$\int_S \frac{\partial[g(x,t)]}{\partial t}\, dF(x).$$

Theorem II.4 - Integration: If, for almost all values of x in S, the function g(x,t) is continuous with respect to t in the finite open interval (a,b) and satisfies the condition $|g(x,t)| < G_3(x)$ for all t in (a,b), then

$$\int_a^b u(t)dt = \int_a^b \left[\int_S g(x,t)dF(x)\right] dt$$

exists and equals

$$\int_S \left[\int_a^b g(x,t)dt\right]dF(x).$$

Further, if the above conditions are satisfied for every finite interval (a,b), and if, in addition, we have

$$\int_{-\infty}^{\infty} |g(x,t)| \, dt < G_4(x),$$

then

$$\int_{-\infty}^{\infty} u(t)dt = \int_S \left[\int_{-\infty}^{\infty} g(x,t)dt\right] dF(x).$$

We proceed now to a consideration of the characteristic function of a random variable.

Definition II.5: Let X be a random variable, let F(x) be its probability distribution function, and let t be a real number. The function

$$\varphi(t) = E(e^{itX}) = \int_{-\infty}^{\infty} e^{itx} dF(x)$$

is called the characteristic function of the random variable X.

As we shall see, from a knowledge of the characteristic function of a random variable one may obtain complete knowledge of its distribution function. The purpose of this section is to examine the nature of this correspondence.

The characteristic function has the following properties:

1) $|\varphi(t)| \leq \int_{-\infty}^{\infty} dF(x) = 1.$

2) $\varphi(-t) = \overline{\varphi(t)}$.

3) $\varphi(0) = 1.$

4) $\varphi(t)$ is continuous.

5) If the kth order moment of the distribution exists, then $\varphi(t)$ is k times differentiable with respect to t, and, for $0 \leq v \leq k$,

$$\varphi^{(v)}(t) = i^v \int_{-\infty}^{\infty} x^v e^{itx} dF(x).$$

6) Under the conditions of property 5), $\varphi^{(v)}(t)$ is continuous for all real t, and

$$\varphi^{(v)}(0) = i^v \int_{-\infty}^{\infty} x^v dF(x), \quad v = 1,2,..,k.$$

Hence, in the neighborhood of t = 0, we have, by the Maclaurin expansion,

$$\varphi(t) = 1 + \sum_{v=1}^{k} \frac{\alpha_v}{v!} (it)^v + o(t^k), \tag{II.6}$$

where $\lim_{t \to 0} \frac{o(t^k)}{t^k} = 0$, and $\alpha_v = \int_{-\infty}^{\infty} x^v dF(x)$.

To show the validity of (II.6), we first let f(t) be a real valued function of the real variable t, and we suppose that f(t) has k continuous derivatives in some neighborhood of t = 0. Then, by the Maclaurin expansion, we obtain that

$$f(t) = \sum_{n=0}^{k-1} \frac{f^{(n)}(0)}{n!} t^n + \frac{f^{(k)}(\theta t)}{k!} t^k, \quad 0 < \theta < 1.$$

Thus,

$$f(t) = \sum_{n=0}^{k} \frac{f^{(n)}(0)t^n}{n!} + \frac{f^{(k)}(\theta t) - f^{(k)}(0)}{k!} t^k.$$

But $f^{(k)}(t)$ is continuous at $t=0$, so $f^{(k)}(\theta t) - f^{(k)}(0) \to 0$ as $t \to 0$. We may thus write

$$\frac{f^k(\theta x)-f^k(0)}{k!} t^k = o(t^k),$$

where $\lim_{t\to 0} \frac{o(t^k)}{t^k} = 0.$

For the case where $f(t) = \varphi(t)$ is a complex valued function of the real variable t we decompose f(t) into real and imaginary parts; i.e. $f(t) = f_1(t) + i f_2(t)$. We then apply the above argument for real valued functions to $f_1(t)$ and $f_2(t)$ separately, and the validity of II.6 follows directly.

Properties 1), 2), and 3) follow from the definition of the integral. Property 5) follows from Theorem II.3, while 4) and 6) follow from Theorem II.2.

From the above discussion it follows that the characteristic function of any function g(X) which is a random variable is the mean value of $e^{itg(X)}$. In case $g(X) = aX + b$, the characteristic function becomes

(II.7) $$E(e^{it(aX+b)}) = E(e^{itb} \cdot e^{itaX}) = e^{itb}\varphi(at).$$

Further, the "standardized variable" $\frac{X-m}{\sigma}$ has the characteristic function

(11.8) $$E(e^{it(\frac{X-m}{\sigma})}) = E(e^{\frac{itX}{\sigma}} \cdot e^{-itm}) = e^{-itm}\varphi(\tfrac{t}{\sigma}).$$

Suppose now that we have n independent random variables $X_1, X_2,\ldots,X_n$, with characteristic functions $\varphi_1(t),\ldots,$

$\varphi_n(t)$. Then the characteristic function of the sum

$X = \sum_{k=1}^{n} X_k$ is

$$\varphi(t) = E(e^{it \sum_{k=1}^{n} X_k}) = E(\prod_{k=1}^{n} e^{itX_k}).$$

Since the X_k are independent,

(II.9) $$\varphi(t) = \prod_{k=1}^{n} E(e^{itX_k}) = \prod_{k=1}^{n} \varphi_k(t).$$

In the discussion which follows we will make use of some of the properties of the following functions:

(II.10) $$s(h,T) = \frac{2}{\pi} \int_0^T \frac{\sin ht}{t}\, dt, \text{ and}$$

(II.11) $$c(h,T) = \frac{2}{\pi} \int_0^T \frac{1-\cos ht}{t^2}\, dt,$$

where h is real and $T > 0$. It is immediate that $c(h,T) \geq 0$, and

$$s(-h,T) = -s(h,T), \quad c(-h,T) = c(h,T).$$

If we let $u = ht$, then, for $h > 0$,

$$s(h,T) = \frac{2}{\pi} \int_0^{hT} \frac{h \sin u}{u} \frac{du}{h} = \frac{2}{\pi} \int_0^{hT} \frac{\sin t}{t}\, dt.$$

Standard treatises ([5] pages 247-8) show that

$$\int_0^u \frac{\sin t}{t}\, dt$$

is bounded for all $u > 0$; furthermore,

$$\lim_{u\to\infty} \int_0^u \frac{\sin t}{t} = \frac{\pi}{2} .$$

Lemma II.12: $s(h,T)$ is bounded for all real h and $T>0$, and for arbitrary fixed δ and δ',

$$\lim_{T\to\infty} s(h,T) = \begin{cases} 1 \text{ uniformly for } h \geqslant \delta > 0, \\ 0 \text{ for } h = 0, \\ -1 \text{ uniformly for } h \leqslant \delta' < 0. \end{cases}$$

Lemma II.13: For each real h, $\lim_{T\to\infty} c(h,T) = |h|$.

Proof: By an integration by parts,

$$c(h,T) = \frac{2h}{\pi} \int_0^{hT} \frac{\sin t}{t}\, dt - \frac{2}{\pi} \cdot \frac{1-\cos hT}{T} .$$

Thus,

$$\lim_{T\to\infty} c(h,T) = h \lim_{T\to\infty} s(h,T) = |h| .$$

We now proceed to the first theorem showing the relation between the characteristic function of a random variable and its distribution function.

Theorem II.14: If $(a-h,a+h)$ is a continuity interval of the distribution function $F(x)$, then

$$F(a+h) - F(a-h) = \lim_{T\to\infty} \frac{1}{\pi} \int_{-T}^{T} \frac{\sin ht}{t} e^{-ita}\varphi(t)dt.$$

Proof: We write

$$J = \frac{1}{\pi} \int_{-T}^{T} \frac{\sin ht}{t} e^{-iat}\varphi(t)dt$$

$$= \frac{1}{\pi} \int_{-T}^{T} \frac{\sin ht}{t} e^{-iat}\, dt \int_{-\infty}^{\infty} e^{itx}\, dF(x).$$

Since

$$\left| \frac{\sin ht}{t} \quad e^{it(x-a)} \right| \leq \left| \frac{\sin ht}{t} \right| \leq |h|$$

for all x,t, the conditions of Theorem II.4 are satisfied, and we may interchange the order of integration. Thus

$$J = \frac{1}{\pi} \int_{-\infty}^{\infty} \left[\int_{-T}^{T} \frac{\sin ht}{t} e^{it(x-a)}\, dt \right] dF$$

$$= \frac{1}{\pi} \int_{-\infty}^{\infty} \left[\int_{-T}^{T} \frac{\sin ht}{t} [\cos t(x-a) + i \sin t(x-a)] dt \right] dF$$

$$= \frac{2}{\pi} \int_{-\infty}^{\infty} \left[\int_{0}^{T} \frac{\sin ht}{t} \cos t(x-a) dt \right] dF = \int_{-\infty}^{\infty} g(x,T) dF,$$

where

$$g(x,T) = \frac{2}{\pi} \int_{0}^{T} \frac{\sin ht}{t} \cos(x-a)t\, dt.$$

By a standard trigonometric substitution, we have

$$g(x,T) = \frac{1}{\pi} \int_{0}^{T} \frac{\sin(x-a+h)t}{t}\, dt - \frac{1}{\pi} \int_{0}^{T} \frac{\sin(x-a-h)t}{t}\, dt$$

$$= \frac{1}{2} S(x-a+h,T) - \frac{1}{2} S(x-a-h,T).$$

We have, by Lemma II.12,

$$\lim_{T\to\infty} g(x,T) = \begin{cases} 0 \text{ for } x < a-h, \\ \frac{1}{2} \text{ for } x = a-h, \\ 1 \text{ for } a-h < x < a+h, \\ \frac{1}{2} \text{ for } x = a+h, \\ 0 \text{ for } x > a+h. \end{cases}$$

Hence $|g(x,T)|$ is bounded by some absolute constant, and we may apply Theorem A. Thus, since $F(x)$ is continuous at $x = a \pm h$, we obtain from a consideration of the Darboux sums that

$$\lim_{T\to\infty} J = \int_{a-h}^{a+h} dF(x) = F(a+h) - F(a-h),$$

and the theorem is proved.

The theorem shows that if two distributions have the same characteristic function, then they have the same variation on every interval which is a continuity interval for both distributions; but then, since the functions are both bounded by 1, they are necessarily identical ([2] page 58).

We have established the one to one correspondence which exists between distribution functions and characteristic functions. We now proceed to show that, under certain conditions, the transformation by which we pass from a distribution function $F(x)$ to the corresponding characteristic function $\varphi(t)$ is a continuous transformation, i.e. $\lim_{n\to\infty} F_n(x) = F(x)$ and $\lim_{n\to\infty} \varphi_n(t) = \varphi(t)$ are equivalent.

Lemma II.15: For any real a and h > 0, we have

$$\int_0^h [F(a+z) - F(a-z)]\,dz = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{1-\cos ht}{t^2}\, e^{-ita}\varphi(t)dt.$$

Proof: Let

$$J_1 = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{1-\cos ht}{t^2}\, e^{-ita}\, dt \int_{-\infty}^{\infty} e^{itx}\, dF(x);$$

hence

$$J_1 = \frac{1}{\pi}\int_{-\infty}^{\infty} dF(x) \int_{-\infty}^{\infty} \frac{1-\cos ht}{t^2}\, e^{it(x-a)}dt,$$

the interchange of order of integration being justified as in Theorem II.14.

We thus have

$$J_1 = \frac{1}{\pi}\int_{-\infty}^{\infty} dF(x) \int_{-\infty}^{\infty} \frac{1-\cos ht}{t^2}\, [\cos t(x-a) + i \sin t(x-a)]dt$$

$$= \frac{1}{\pi}\int_{-\infty}^{\infty} dF(x) \int_{-\infty}^{\infty} \frac{\cos t(x-a) - \cos ht \cos t(x-a)}{t^2}\, dt$$

$$+ i \int_{-\infty}^{\infty} [\frac{\sin t(x-a)}{t^2} - \frac{\sin t(x-a)\cos ht}{t^2}]dt.$$

The last integral is zero, since the integrand is an odd function and is integrable. Applying a standard trigonometric substitution, we obtain that

$$J_1 = \frac{2}{\pi}\int_{-\infty}^{\infty} dF(x)\int_0^{\infty} \frac{2\cos t(x-a) - \cos(x-a-h)t - \cos(x-a+h)t}{2t^2}\,dt$$

$$= \frac{2}{\pi}\int_{-\infty}^{\infty} dF(x)\Bigg[\frac{1}{2}\cdot\frac{2}{\pi}\int_0^{\infty}\frac{2\cos t(x-a)-2}{t^2}\,dt$$

$$+ \frac{2}{\pi}\int_0^{\infty}\frac{1-\cos(x-a+h)t}{t^2}\,dt + \frac{2}{\pi}\int_0^{\infty}\frac{1-\cos(x-a-h)t}{t^2}\,dt\Bigg].$$

But, by (II.11),

$$\frac{2}{\pi}\int_0^{\infty}\frac{1-\cos(x-a+h)t}{t^2}\,dt = c(x-a+h,T) = |x-a+h|\,,$$

$$\frac{2}{\pi}\int_0^{\infty}\frac{1-\cos(x-a-h)t}{t^2}\,dt = c(x-a-h,T) = |x-a-h|\,,$$

and

$$\frac{2}{\pi}\int_0^{\infty}\frac{2\cos t(x-a)-2}{t^2}\,dt = -2\cdot\frac{2}{\pi}\int_0^{\infty}\frac{1-\cos t(x-a)}{t^2}\,dt$$

$$= -2\cdot c(x-a,T) = -2|x-a|\,.$$

Thus

$$J_1 = \int_{-\infty}^{\infty}\left[\frac{-2|x-a| + |x-a+h| + |x-a-h|}{2}\right] dF(x).$$

When $x \leq a-h$ or $x \geq a+h$, $\dfrac{-2|x-a| + |x-a-h| + |x-a+h|}{2} = 0$, so that $J_1 = 0$. On the other hand, when $a-h < x < a+h$,

$$J_1 = \int_{a-h}^{a} (x-a+h)dF(x) + \int_{a}^{a+h} (h+a-x)dF(x).$$

Let $G_1 = x-(a-h)$, $G_2 = h+a-x$. By Theorem B,

$$J_1 = \int_{a-h}^{a} G_1 dF(x) + \int_{a}^{a+h} G_2 dF(x)$$

$$= F(a)\cdot h - \int_{a-h}^{a} F(x)\, dx + (-F(a)\cdot h) + \int_{a}^{a+h} F(x)dx$$

$$= - \int_{a-h}^{a} F(x)dx + \int_{a}^{a+h} F(x)dx.$$

We let $x = a-u$ in $\int_{a-h}^{a} F(x)dx$ and $x = a+u$ in $\int_{a}^{a+h} F(x)dx$ and find that

$$J_1 = \int_{h}^{0} F(a-u)du + \int_{0}^{h} F(a+u)du$$

$$= \int_{0}^{h} [F(a+u) - F(a-u)]\, du,$$

and the lemma is proved.

Theorem II.16: Let $\{F_n(x)\}$ be a sequence of distribution functions and $\{\varphi_n(t)\}$ the corresponding sequence of characteristic functions. A necessary and sufficient condition for the convergence of the sequence $\{F_n(x)\}$ to a distribution function $F(x)$ is that, for every t, the sequence $\{\varphi_n(t)\}$ converges to a limit $\varphi(t)$ which is continuous at $t = 0$. Where this condition is satisfied, the limit $\varphi(t)$ is identical with the characteristic function of the limiting distribution function $F(x)$.

Proof: We first prove the condition is necessary. Suppose $F_n(x) \to F(x)$. We will show that

$$\lim_{n\to\infty} \left| \int_{-\infty}^{\infty} e^{itx}\, dF_n(x) - \int_{-\infty}^{\infty} e^{itx} dF(x) \right| = 0;$$

i.e., that $\varphi_n(t)$ converges to a characteristic function. It follows from Theorem C that, for any finite interval (a,b), we have

$$\lim_{n\to\infty} \int_a^b e^{itx}\, dF_n(x) = \int_a^b e^{itx}\, dF(x).$$

We choose a and b, continuity points of $F(x)$, such that $F(x) < \frac{\epsilon}{7}$ when $x \leq a$, and $F(x) > 1 - \frac{\epsilon}{7}$, when $x \geq b$. Now we choose n so large that $|F_n(b) - F(b)| < \frac{\epsilon}{7}$, $|F_n(a) - F(a)| < \frac{\epsilon}{7}$, and

$$\left| \int_a^b e^{itx}\, dF_n(x) - \int_a^b e^{itx}\, dF(x) \right| < \frac{\epsilon}{7}.$$

Then

$$\left|\int_{-\infty}^{\infty} e^{itx}\, dF_n(x) - \int_{-\infty}^{\infty} e^{itx}\, dF(x)\right| \leq \left|\int_{-\infty}^{a} e^{itx}\, dF_n(x)\right|$$

$$+ \left|\int_{-\infty}^{a} e^{itx}\, dF(x)\right| + \left|\int_{a}^{b} e^{itx}\, dF_n(x) - \int_{a}^{b} e^{itx}\, dF(x)\right|$$

$$+ \left|\int_{b}^{\infty} e^{itx}\, dF_n(x)\right| + \left|\int_{b}^{\infty} e^{itx}\, dF(x)\right| .$$

But,

$$\left|\int_{-\infty}^{a} e^{itx}\, dF_n(x)\right| \leq \int_{-\infty}^{a} \left|e^{itx}\right|\, dF_n(x) \leq \int_{-\infty}^{a} dF_n(x)$$

$$= F_n(a), \text{ and } \int_{-\infty}^{a} e^{itx}\, dF(x) \leq F(z) < \frac{\epsilon}{7} ;$$

further,

$$\left|\int_{b}^{\infty} e^{itx}\, dF_n(x)\right| \leq \int_{b}^{\infty} dF_n(x)$$

$$= 1-F_n(b), \text{ and } \left|\int_{b}^{\infty} e^{itx}\, dF(x)\right| \leq 1-F(b) < \frac{\epsilon}{7} .$$

By choice of n, $\left|F_n(a)\right| \leq \left|F_n(a) - F(a)\right| + \left|F(a)\right| < \frac{\epsilon}{7} + \frac{\epsilon}{7} = \frac{2\epsilon}{7}$. Similarly, we find that $F_n(b) \leq \frac{2\epsilon}{7}$. Thus,

$$\left| \int_{-\infty}^{\infty} e^{itx}\, dF_n(x) - \int_{-\infty}^{\infty} e^{itx}\, dF(x) \right| < \epsilon .$$

Since ϵ is arbitrary, necessity is proved.

To prove sufficiency we assume that for every t, $\varphi_n(t)$ tends to a limit $\varphi(t)$ which is continuous for t = 0, and we prove that $F_n(x)$ tends to a distribution function $F(x)$.

By Theorem I.6, the sequence $\{F_n(x)\}$ contains a convergent subsequence whose limit is a non-decreasing function $F(x)$, where $F(x)$ may be determined so as to be continuous on the right. We show first that $F(x)$ is a distribution function. Since we have $0 \leq F(x) \leq 1$, it is sufficient to prove that $F(+\infty) - F(-\infty) = 1$. From Lemma II.15, we have, putting a = 0,

$$\int_0^h [F_{n_v}(z) - F_{n_v}(-z)]dz = \int_0^h F_{n_v}(z)dz - \int_{-0}^{0} F_{n_v}(z)dz$$

$$= \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{1-\cos ht}{t^2}\, \varphi_n(t)dt.$$

The integrands $F_{n_v}(z)$ and $F_{n_v}(-z)$ are uniformly bounded by 1, and $\lim_{v} F_{n_v}(x) = F(x)$ a.e., since $F(x)$ has only a countable number of discontinuities on these intervals. Thus Theorem A applies, and we have

$$\lim_{v \to \infty} \int_0^h [F_{n_v}(z) - F_{n_v}(-z)]dz = \int_0^h F(z)dz - \int_{-h}^{0} F(z)dz.$$

Since $|\varphi_n(t)| \leq 1$, $\frac{1-\cos ht}{t^2}\varphi_{n_v}(t)$ is dominated by the function $\frac{1-\cos ht}{t^2}$, which is integrable over $(-\infty,\infty)$; thus, by Theorem A,

$$\lim_{v\to\infty}\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{1-\cos ht}{t^2}\varphi_{n_v}(t)dt = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{1-\cos ht}{t^2}\varphi(t)dt.$$

Dividing by h, we obtain

$$\frac{1}{h}\int_0^h F(z)dz - \frac{1}{h}\int_{-h}^0 F(z)dz = \frac{1}{\pi h}\int_{-\infty}^{\infty}\frac{1-\cos ht}{t^2}\varphi(t)dt.$$

We let $z = uh$, $r = th$, and obtain

$$\frac{1}{h}\int_0^h F(z)dz - \frac{1}{h}\int_{-h}^0 F(z)dz = \int_0^1 [F(uh) - F(-uh)]du$$

$$= \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{1-\cos r}{r^2}\varphi\left(\frac{r}{h}\right) dr.$$

We now allow h to tend to infinity. Then,

$$\lim_{h\to\infty}\int_0^1 [F(uh) - F(-uh)]du = \int_0^1 [F(\infty) - F(-\infty)]du$$

$$= F(\infty) - F(-\infty).$$

On the other hand, by hypothesis $\varphi(\frac{r}{h})$ is continuous for $r = 0$, so that $\varphi(\frac{r}{h})$ tends, for every r, to the limit $\varphi(0)$. But by hypothesis $\varphi(0) = \lim_{n\to\infty}\varphi_n(0)$, and $\varphi_n(0) = 1$ for every n. Again applying Theorem A, we have

$$\lim_{h\to\infty} \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{1-\cos r}{r^2} \varphi\left(\frac{r}{h}\right) dr = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{1-\cos r}{r^2} \varphi(0) dr = 1.$$

Thus,

$$F(+\infty) - F(-\infty) = 1.$$

Since $0 \leq F(x) \leq 1$, we must have $F(+\infty) = 1$, $F(-\infty) = 0$, and the limit function $F(x)$ of the sequence $\{F_n(x)\}$ is a distribution function. By the first part of the proof, it follows that $\lim_v \varphi_{n_v}(t) = \varphi(t)$ is necessarily the characteristic function of $F(x)$.

We now consider another convergent subsequence of $\{F_n(x)\}$, and denote the limit of the new subsequence by $F'(x)$, assuming this function to be determined so as to be continuous to the right. In the same way as above, we may show that $F'(x)$ is a distribution function. But by hypothesis the characteristic functions of the new subsequence have, for all values of t, the same limit $\varphi(t)$ as before. $\varphi(t)$ is thus the characteristic function of both $F(x)$ and $F'(x)$; hence, according to the uniqueness theorem (II.14), we have $F(x) = F'(x)$ for all x.

Therefore every convergent subsequence of $F_n(x)$ has the same limit $F(x)$. Further, $\{F_n(x)\}$ itself converges to $F(x)$ in all continuity points of $F(x)$. For suppose not. Then there must be a continuity point x_0 of $F(x)$ and an $\epsilon > 0$ such that for every N there exists an $n > N$ such that $|F_n(x_0) - F(x_0)| \geq \epsilon$. Thus the set

$\{F_{n_v}(x_o)\} = \{F_n(x_o) \mid |F_n(x_o) - F(x_o)| \geq \epsilon\}$ is an infinite set of real numbers. The corresponding sequence of functions has, by Theorem I.6, a convergent subsequence $\{F_{n_{v_k}}(x)\}$. We have already established that every convergent subsequence of $\{F_n(x)\}$ converges to F(x) in continuity points of F(x). We must therefore have $\lim_{k\to\infty} F_{n_{v_k}}(x_o) = F(x_o)$, which contradicts the assumption that $|F_{n_v}(x_o) - F(x_o)| \geq \epsilon$ for all v. We conclude that $\{F_n(x)\}$ converges to F(x) on the set of continuity points of F(x).

CHAPTER III

The main results of this chapter are Theorem III.6, dealing with the convergence in distribution of a random variable with a binomial distribution to a random variable with a Poisson distribution, and Theorems III.15 and III.17, both variations of the Central Limit Theorem. Brief discussions of the pertinent properties of the binomial, Poisson and normal probability distributions have been included for ease of reference.

Let us first consider a random experiment in which we will denote by E an event with fixed probability p of occurring on any given trial. Let a series of n independent trials of the experiment be performed, and define a random variable Y_k, attached to the kth repetition, as follows:

$$(III.1) \qquad Y_k = \begin{cases} 1 \text{ if E occurs at the kth repetition (probability p),} \\ 0 \text{ otherwise (probability } q = 1-p). \end{cases}$$

Then the random variable

$$(III.2) \qquad X_n = \sum_{k=1}^{n} Y_k$$

denotes the total number of occurrences of E in the n repetitions of the experiment.

Theorem III.3 [2]: Consider n repetitions of a random experiment, and let Y_k be defined by (III.1), and X_n by (III.2). The the probability distribution of X_n is given by

$$P[X_n=r] = \binom{n}{r} p^r q^{n-r}.$$

Proof: The c.f. for Y_k is

$$\varphi_{Y_k}(t) = \sum_{j=1}^{2} e^{ity_k} p_j = e^{it}\cdot p + q.$$

Therefore, $\varphi_{X_n}(t) = (e^{it}\cdot p+q)^n$. We thus obtain, by the binomial expansion, that

$$\varphi_{X_n}(t) = (e^{it}\cdot p+q)^n = \sum_{r=1}^{n} \binom{n}{r} p^r q^{n-r}\cdot e^{itr}.$$

But this is the characteristic function of a variable which takes on values r = 0,1,2,...with the probabilities

$$P_r = \binom{n}{r} p^r q^{n-r}$$

By the one to one correspondence between distribution functions and characteristic functions, we thus conclude that

$$P[X_n=r] = \binom{n}{r} p^r q^{n-r},$$

as asserted. The distribution just derived is called the binomial distribution.

Before obtaining an important result concerning the binomial distribution, we discuss briefly the Poisson distribution.

Definition III.4: If the distribution function of a random variable X is specified by

$$P[X=k] = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0,1,2,\ldots,$$

X is said to have a Poisson distribution.

A consideration of the Maclaurin expansion for $e^{\lambda e^{itr}}$ shows at once that the c.f. for the Poisson distribution is

$$\text{(III.5)} \qquad \varphi(t) = \sum_{r=0}^{\infty} \frac{\lambda^r}{r!} e^{-\lambda} e^{itr} = e^{\lambda(e^{it}-1)}.$$

Theorem III.6 [2]: Let $\{X_n\}$ be the sequence of binomially distributed random variables defined in (III.2). Then if $n \to \infty$ in such a way that $np_n \to \lambda$(constant), the distribution of X_n approaches the Poisson distribution with parameter λ.

Proof: The c.f. for X_n was found to be

$$\varphi_n(t) = (e^{it}p_n + q)^n = [1 + p_n(e^{it}-1)]^n.$$

Thus

$$\varphi_n(t) = [1 + \frac{np_n(e^{it}-1)}{n}]^n.$$

But, $np_n \to \lambda$ as $n \to \infty$, thus we may write $np_n = \lambda + \eta(n)$, where $\eta(n) \to 0$ as $n \to \infty$. Hence

$$\varphi_n(t) = [1 + \frac{\lambda + \eta(n)(e^{it}-1)}{n}]^n.$$

Letting $\eta(n)(e^{it}-1) = \alpha(n)$ and $\lambda(e^{it}-1) = \gamma$, we have, by the binomial expansion,

$$\varphi_n(t) = (1+\tfrac{\gamma}{n})^n + [\tbinom{n}{1}(1+\tfrac{\gamma}{n})^{n-1}(\tfrac{\alpha(n)}{n}) + \tbinom{n}{2}(1+\tfrac{\gamma}{n})^{n-2}(\tfrac{\alpha(n)}{n})^2 + \ldots + (\tfrac{\alpha(n)}{n})^n].$$

We show that the quantity in brackets converges absolutely to 0. Let

$$z(n) = [\tbinom{n}{1}(1+\tfrac{|\gamma|}{n})^{n-1}\left|\tfrac{\alpha(n)}{n}\right| + \tbinom{n}{2}(1+\tfrac{|\gamma|}{n})\left|\tfrac{\alpha(n)}{n}\right|^2 + \ldots + \left|\tfrac{\alpha(n)}{n}\right|^n].$$

Since $(1+\frac{|\gamma|}{n})^k \leq e^{|\gamma|}$ for $k \leq n$, and since $\binom{n}{r}\frac{1}{n^r} \leq \frac{1}{r!}$ when r is a positive integer not exceeding n,

$$z(n) \leq |\alpha(n)| e^{|\gamma|}[1 + \left|\frac{\alpha(n)}{2!}\right| + \left|\frac{\alpha^2(n)}{3!}\right| + \ldots + \left|\frac{\alpha^{n-1}(n)}{n!}\right|].$$

But, for sufficiently large n, $|\alpha^k(n)| < 1$, so that

$$z(n) \leq |\alpha(n)| e^{|\gamma|}[1 + \frac{1}{2!} + \frac{1}{3!} + \ldots + \frac{1}{n!}] \leq |\alpha(n)| e^{|\gamma|+1}.$$

Thus,

$$\lim_{n\to\infty} z(n) = 0,$$

and

$$\lim_{n\to\infty} \varphi_n(t) = e^{\gamma} = e^{\lambda(e^{it}-1)}.$$

Expression (III.5) shows this to be the c.f. of a random variable with the Poisson distribution, and the theorem is proved.

Suppose a random variable X is defined such that X almost always assumes a constant value c. This distribution function of this variable is the function F(x-c) defined by

(III.7) $$F(x-c) = \begin{cases} 0 \text{ if } x < c, \\ 1 \text{ if } x \geq c. \end{cases}$$

A consideration of the definition of the characteristic function of X shows that $\varphi(t) = e^{cit}$. The following theorem shows a relation between convergence in distribution and convergence in probability to a constant.

Theorem III.8 [5]: Let $\{X_n\}$ be a sequence of random variables, with corresponding distribution functions $\{F_n(x)\}$. Then X_n converges in probability to a constant c if $\lim_{n\to\infty} F_n(x) = F(x-c)$, where F(x-c) is defined by (III.7).

Proof: $P[|X_n - c| \geq \epsilon] = P[X_n \geq c + \epsilon] + P[X_n \leq c - \epsilon]$

$$= 1 - F_n(c+\epsilon) + F_n(c-\epsilon+0).$$

But by hypothesis, $\lim_{n\to\infty} F_n(x) = F(x-c)$. Hence

$$\lim_{n\to\infty} P[|X_n-c| \geq \epsilon] = 1 - \lim_{n\to\infty} F_n(c+\epsilon) + \lim_{n\to\infty} F_n(c-\epsilon+0) = 0,$$

as asserted by the theorem.

A stronger theorem than the converse is true:

Theorem III.9 [5]: Let $\{X_n\}$ and $\{F_n(x)\}$ be as in the previous theorem. If X_n converges in probability to a random variable X, with d.f. F(x), then $F_n(x)$ converges to F(x).

Theorem III.10 (Khintchine) [2]: Let $\{X_n\}$ be a sequence of identically distributed independent random variables, and suppose $E(X_n) = m$ exists. Then the random variable

$\bar{X} = \frac{1}{n}\sum_{k=1}^{n} X_k$ converges in probability to m.

Proof: Let $\varphi(t)$ be the c.f. corresponding to the common distribution of the X_k. Then, according to (II.9) and (II.7) the c.f. of $\bar{X}$ is $[\varphi(\frac{t}{n})]^n$. By property 6) of the c.f., we have

$$\varphi(t) = 1 + mit + o(t);$$

Thus, for any fixed t,

$$[\varphi(\tfrac{t}{n})]^n = [1 + \frac{mit}{n} + o(\tfrac{t}{n})]^n$$
$$= [1 + \frac{mit}{n} + \frac{o(\frac{t}{n})\cdot n}{n}]^n .$$

Hence,

$$\lim_{n\to\infty} [\varphi(\tfrac{t}{n})]^n = \lim_{n\to\infty} [1 + \frac{mit+o(\frac{t}{n})}{n}] = e^{mit}.$$

By Theorem II.16, e^{mit} is the c.f. for some random variable, since e^{mit} is continuous at t = 0. The assertion of the theorem thus follows from Theorem III.8, Theorem II.14, and (III.7).

The important result in each of the Theorems III.15, III.17 in the following discussion is that, subject to very general conditions regarding their distributions, certain commonly encountered sequences of random variables converge in distribution to the normal distribution.

The standard normal distribution function is defined by

$$\text{(III.11)} \qquad \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt.$$

The mean value of the distribution is zero, the standard deviation is 1, as seen by

$$\int_{-\infty}^{\infty} x d\,\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-\frac{x^2}{2}}\, dx = 0,$$

and

$$\int_{-\infty}^{\infty} x^2 d\,\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2}}\, dx = 1.$$

Theorem III.12: The characteristic function of $\Phi(x)$ is

$$\varphi(t) = \int_{-\infty}^{\infty} e^{itx} d\,\Phi(x) = e^{-\frac{t^2}{2}}.$$

Proof: We have

$$\varphi(t) = \int_{-\infty}^{\infty} e^{itx} d\,\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx - x^2/2} dx.$$

Completing the square on the exponent, we obtain

$$\text{(III.13)} \quad \varphi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-it)^2}{2}} \cdot e^{-\frac{t^2}{2}} \, dx$$

$$= e^{-\frac{t^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-it)^2}{2}} \, dx \, .$$

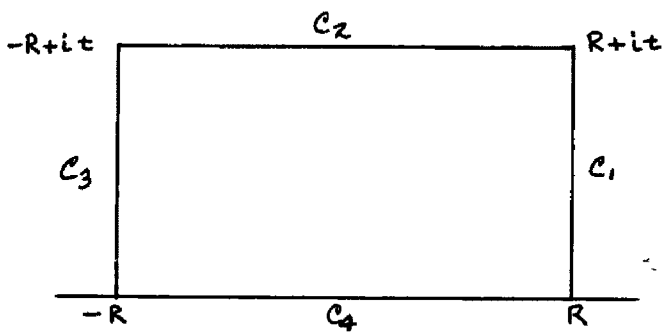

Figure 2.

A consideration of the integral

$$\frac{1}{\sqrt{2\pi}} \int_C e^{-\frac{z^2}{2}} \, dz,$$

where $C = \bigcup_{i=1}^{4} C_i$ (Figure 2), shows that the value of the last integral of (III.13) is 1, and the theorem is proved.

We extend the notions of the previous paragraphs as follows.

Definition III.14. The random variable X is said to be normally distributed with mean m and standard deviation σ if the distribution function of X is

$$\frac{1}{\sigma} \Phi\left(\frac{x-m}{\sigma}\right).$$

It follows that the variable $\frac{X-m}{\sigma}$ has the distribution function $\Phi(x)$, hence the characteristic function $e^{-\frac{t^2}{2}}$.

We finally consider the Central Limit Theorem.

Theorem III.15(Lindberg-Lévy) [2]: Let $\{X_n\}$ be a sequence of identically distributed independent random variables with mean m_1 and standard deviation σ_1. Let random variables X and $\overline{X}$ be defined by

$$\text{(III.16)} \qquad X = \sum_{k=1}^{n} X_k, \text{ and } \overline{X} = \frac{1}{n}\sum_{k=1}^{n} X_k .$$

Then the d.f. of the random variable

$$\frac{X-nm_1}{\sigma_1\sqrt{n}}$$

converges to $\Phi(x)$, defined by III.11.

(It follows that the d.f. of the random variable

$$\frac{\overline{X}-m_1}{\sigma_1/\sqrt{n}}$$

converges to $\Phi(x)$.)

Proof: Since the X_k are identically distributed, it follows that the random variable X, defined by (III.16), has mean $m = nm_1$ and standard deviation $\sigma = \sigma_1\sqrt{n}$. We consider the variable

$$\frac{X-m}{\sigma} = \frac{X-nm_1}{\sigma_1\sqrt{n}} = \frac{1}{\sigma_1\sqrt{n}}\sum_{k=1}^{n}(X_k-m_1).$$

Denote by $\varphi_1(t)$ the characteristic function of the deviations $X_k - m_1$, and by $F(x)$ and $\varphi_x(t)$ the distribution function and characteristic function of the variable $\frac{X-m}{\sigma}$. Then by (II.9),

$$\varphi_x(t) = [\varphi_1(\frac{t}{\sigma_1\sqrt{n}})]^n.$$

Now, the first two moments of the variable $X_k - m_1$ are 0 and σ_1^2 respectively. Thus by property 6) of the characteristic function,

$$\varphi_1(t) = 1 - \frac{1}{2}\sigma_1^2 t^2 + o(t^2).$$

Substituting $\frac{t}{\sigma_1\sqrt{n}}$ for t we obtain

$$\varphi_x(t) = [1 - \frac{t^2}{2n} + \frac{\eta(n,t)}{n}]^n,$$

Where $\eta(n,t)$ tends to 0 for each fixed t as $n \to \infty$. Thus, for each t, as $n \to \infty$,

$$\varphi_x(t) \to e^{-\frac{t^2}{2}},$$

which, by Theorem III.12, is the characteristic function of $\Phi(x)$. We thus conclude that for each x, $F(x) \to \Phi(x)$, and the theorem is proved.

Theorem III.17 (Liapounoff) [2]: Let $\{X_n\}$ be a sequence of independent random variables, and denote by m_v, σ_v, and p_v^3, respectively the mean, standard deviation and third absolute moment about the mean of X_v. Let

$$m(n) = \sum_{v=1}^{n} m_v,$$

$$\sigma_{(n)}^2 = \sum_{v=1}^{n} \sigma_v^2,$$

and

$$p^3_{(n)} = \sum_{v=1}^{n} p^3_v \, .$$

Suppose that p^3_v is finite for every v. Then, if the condition

$$\lim_{n\to\infty} \frac{p(n)}{\sigma(n)} = 0$$

is satisfied, the distribution function of the random variable $\frac{X-m(n)}{\sigma(n)}$, where

$$X = \sum_{v=1}^{n} X_v,$$

converges to $\Phi(x)$.

In the proof that follows, the quantities θ_i will always be such that $|\theta_i| \le 1$.

Proof: We let $\varphi_v(t)$ be the c.f. of the v-th deviation, $X_v - m_v$, and $\varphi(t)$ be the c.f. of the standardized random variable

$$\frac{X-m(n)}{\sigma(n)} = \frac{1}{\sigma(n)} \sum_{v=1}^{n} (X_v - m_v) .$$

Then

$$\varphi(t) = \prod_{v=1}^{n} \varphi_v\left(\frac{t}{\sigma(n)}\right)$$

If we can show that for every t, $\lim_{n\to\infty} \varphi(t) = e^{-\frac{t^2}{2}}$, then the theorem follows immediately from Theorem II.14.

From the Maclaurin expansion

$$e^{iz} = \sum_{r=0}^{k-1} \frac{(iz)^r}{r!} + \theta_0 \frac{z^k}{k!} ,$$

where k is a positive integer, z real and $|\theta_0| \leq 1$, we obtain, for k = 3,

$$\varphi_v(t) = E(e^{it(X_v - m_v)})$$

$$= E(1 - \frac{t^2(X_v-m_v)^2}{2} + \frac{\theta_1 t^3 (X_v-m_v)^3}{6})$$

$$= 1 - \frac{t^2\sigma_v^2}{2} + \theta_2 \cdot \frac{p_v^3 t^3}{6} .$$

Further, $\varphi_v(\frac{t}{\sigma(n)}) = (1 - \frac{t^2\sigma_v^2}{2\sigma_{(n)}^2} + \theta_2 \cdot \frac{p_v^3 t^3}{6\sigma_{(n)}^3})$; so,

letting $z = -\frac{t^2\sigma_v^2}{2\sigma_{(n)}^2} + \theta_2 \cdot \frac{p_v^3 t^3}{6\sigma_{(n)}^3}$, we obtain

$$\log \varphi_v(\frac{t}{\sigma(n)}) = \log (1+z).$$

Since, by hypothesis, $\frac{p(n)}{\sigma(n)} < 1$, for sufficiently large values of n, we have

$$\frac{p_v}{\sigma(n)} \leq \frac{p(n)}{\sigma(n)} < 1.$$

It can be shown ([2] pages 175-6) that $\sigma_v \leq p_v$ for every v.

Thus,

$$z = \frac{\theta_3 p_v^2 t^2}{2\sigma_{(n)}^2} + \theta_2 \cdot \frac{p_v^3 t^3}{6\sigma_{(n)}^3} .$$

$$= \frac{p_v^2}{\sigma_{(n)}^2} (\frac{\theta_3 t^2}{2} + \theta_2 \cdot \frac{p_v}{\sigma(n)} \cdot \frac{t^3}{6}).$$

Since $\frac{p_v}{\sigma(n)} < 1$, $z = \frac{p_v^2}{\sigma_{(n)}^2} \left(\frac{\theta_3 t^2}{2} + \theta_4 \cdot \frac{|t|^3}{6}\right)$.

Thus,

$$z = \frac{p^2}{\sigma_{(n)}^2} \theta_5 \cdot \left(\frac{t^2}{2} + \frac{|t|^3}{6}\right).$$

Hence $z \to 0$ as $n \to \infty$ for every fixed t. In particular, we may choose n so large that $|z| < \frac{1}{2}$, in which case

$$\begin{aligned} \log(1+z) &= z - \frac{z^2}{2}\left(1 - \frac{2}{3}z + \frac{2}{4}z^2 - \frac{2}{5}z^3 + \ldots\right) \\ &= z + \theta_6 \frac{z^2}{2}\left(1 + \frac{1}{2} + \frac{1}{4} + \ldots\right) \\ &= z + \theta_6 z^2 . \end{aligned}$$

Hence,

$$\begin{aligned} \log \varphi_v\left(\frac{t}{\sigma}\right) &= \log(1+z) \\ &= -\frac{\sigma_v^2 t^2}{2\sigma_{(n)}^2} + \frac{\theta_2 p_v^3 t^3}{6\sigma_{(n)}^3} + \theta_7 \frac{p^4}{\sigma_{(n)}^4}\left(\frac{t^2}{2} + \frac{|t|^3}{6}\right)^2 \\ &= -\frac{\sigma_v^2 t^2}{2\sigma_{(n)}^2} + \frac{\theta_8 p_v^3}{\sigma_{(n)}^3}\left(\frac{|t|^3}{6} + \theta_9\left(\frac{t^2}{2} + \frac{|t|^3}{6}\right)^2\right). \end{aligned}$$

Summing for v=1,2,...,n, we have

$$\log \varphi(t) = -\frac{t^2}{2} + \theta_8 \frac{p_{(n)}^3}{\sigma_{(n)}^3}\left(\frac{|t|^3}{6} + \theta_9\left(\frac{t^2}{2} + \frac{|t|^3}{6}\right)^2\right).$$

As $n \to \infty$, $\frac{p_{(n)}^3}{\sigma_{(n)}^3} \to 0$, thus $\log \varphi(t) \to \frac{-t^2}{2}$ for every fixed t.

Hence, as $n \to \infty$, $\varphi(t) \to e^{-\frac{t^2}{2}}$, and the theorem is proved.

REFERENCES CITED

[1] Harald Cramér, The Elements of Probability Theory and Some of Its Applications, John Wiley and Sons, Inc., New York, 1955.

[2] Harald Cramér, Mathematical Methods of Statistics, Princeton University Press, Princeton, 1946.

[3] William Feller, An Introduction to Probability Theory and Its Applications, John Wiley and Sons, Inc., New York, 1957.

[4] Einar Hille, Analytic Function Theory, Vol. 1, Ginn and Co., Boston, 1959.

[5] Michel Loève, Probability Theory, D. van Nostrand, Co., New York, 1955.

[6] M. Evans Munroe, Introduction to Measure and Integration, Addison-Wesley Publishing Co., Inc., Cambridge, 1953.

[7] Emanuel Parzen, Modern Probability Theory and Its Applications, John Wiley and Sons, Inc., New York, 1960.

[8] Walter Rudin, Principles of Mathematical Analysis, McGraw-Hill Book Company Inc., New York, 1953.