ON THE CENTRAL LIMIT THEOREM

by

MARIETTA RETEK

M.Sc. Mathematics

ABSTRACT

This thesis deals with the limit behaviour of sums of independent random variables specifically investigating the conditions under which convergence to normal distribution occurs. These conditions are expressed in the central limit theorems.

In Chapter I we present direct proofs which means we give direct estimations for the distribution functions of the sums.

In Chapter II and III we turn to indirect methods in proving central limit theorems. The convergence of the distribution functions is proved by showing the convergence of operators in Chapter II and that of characteristic functions in Chapter III.

Although the main object of this thesis is to present central limit theorems for sums of independent random variables, in Chapter IV we give a brief discussion of certain sums of dependent random variables where the distribution of the sum converges to the normal distribution.

ON THE CENTRAL LIMIT THEOREMS

by

MARIETTA RETEK

ON THE CENTRAL LIMIT THEOREMS

by

MARIETTA RETEK

A thesis being submitted to the Faculty of Graduate Studies and Research in partial fulfilment of the requirements for the degree of Master of Science.

Department of Mathematics
McGill University
Montréal 110, Québec
Canada

March 1971

TABLE OF CONTENTS

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to Professor Anatole Joffe, my research director, for his guidance and encouragement in the preparation of this thesis.

I owe special thanks to Dr. János Komlós who made his manuscript available to me and whose helpful comments, as my acting research director throughout the summer months, I greatly appreciated.

0.1 Introduction

It was already known to Bernoulli that the standardized[1] binomial distribution with a fixed parameter p tended to the standard normal distribution as n approached infinity. This result can be restated the following way. If we form the sum of n independent Bernoullian trials with the same parameter p (we obtain a new random variable with binomial distribution and parameters n and p) then the distribution of the standardized sum tends to the standard normal distribution as n tends to infinity.

This is a special case of a widely known phenomenon, namely that in many cases the sum of a large number of independent effects is nearly normally distributed. The precise mathematical model can be described by the limit behaviour of sums of independent random variables and different conditions can be found under which convergence to normal distribution occurs. These conditions are expressed in the central limit theorems, which we are going to deal with in this thesis.

In Chapter I we present direct proofs for different C.L.T's beginning with the Bernoulli case and then going to more general situations. These proofs give direct estimations for the distribution functions of the sums.

In Chapters II and III we make use of indirect methods, where the convergence of distribution functions is proved by showing the convergence

---

(1) If X is a random variable with expectation EX and standard deviation $\sigma(X)$, then the r.v. $(X-EX)/\sigma(X)$ is called the standardized r.v. .

of some different mathematical objects (these will be operators in Chapter II and the Fourier transforms in Chapter III.) defined for the distribution functions as their characteristics.

There are some different types of proofs of the C.L.T. which we are not going to discuss. As an example we mention an Information theoretic proof due to Linnik [23]. He makes use of the extremum property of the normal distribution connected with entropy.

In Chapter IV, we mention theorems about the convergence in distribution of sums of non-independent (but not strongly dependent) random variables to the normal distribution.

Remark.

We will not deal with convergence to other than normal distributions but it should be emphasized, that even for independent variables, the appropriately standardized sum can tend to other distributions.

In the practical problems there is another frequent source of misunderstanding; the mathematical model for the 'sum of effects' is not always the sum of variables, e.g. in colloid-chemistry the usual model which fits the situation works rather for the product of variables, that is the reason why the usual limiting distribution occuring in this field is not the normal but the logarithmic normal distribution.

As far as the originality of the work done in this thesis is concerned, apart from the usual filling in the details there is one completely original theorem and that is Theorem 3 of Chapter I.

Chapter I

Direct Methods in Proving the Central Limit Theorem

1.1. The de Moivre - Laplace Theorem.

The first and simplest form of the Central Limit Theorem, namely the convergence of the distribution of the sum of certain variables to the normal distribution under certain conditions, is what we call today the de Moivre - Laplace theorem, which proves the convergence of the standardized binomial distribution with a fixed parameter p to the standard normal distribution as n approaches infinity.

It is quite easy to see why this theorem should fall under the category of C.L.T.'s. One should just consider a sequence of independent identically distributed Bernoullian trials - by that we mean discrete random variables taking value 1 with probability p and value 0 with probability $1-p$. Assuming that 1 denotes a success, let $X_k = 1$ if the k^{th} trial is a success, and 0 otherwise, then the sum of the first n random variables X_k is nothing, but the number of successes in n trials, which is, what we call, a binomially distributed random variable.

In other words, the binomial distribution can be obtained as the convolution of Bernoullian distributions with the same parameter p.

Since the de Moivre - Laplace theorem proves the convergence of the binomial distribution to the normal distribution, it proves the convergence of the distribution of the sum of independent r.v.'s to the normal distribution and that's what we call a C.L.T.

We are now ready to state and prove the theorem (see p. 131, Rényi [27]).

Theorem 1.

Let $X_1, X_2, \ldots, X_n$ be independent Bernoullian r.v.'s with common parameter p. Then the sum $Y_n = X_1 + X_2 + \ldots + X_n$ is binomially distributed with mean np and standard deviation $\sqrt{npq}$, namely $P(Y_n = k) = \binom{n}{k} p^k q^{n-k}$.

Let $Y_n^* = \frac{Y_n - np}{\sqrt{npq}}$. Then Y_n^* is a standardized r.v. with mean 0 and standard deviation 1. Then the de Moivre - Laplace theorem states that:

$$(1) \qquad \lim_{n \to \infty} P(Y_n^* < x) = \Phi(x),$$

where $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2}\, du$ - the standard normal distribution function.

We are going to prove an equivalent statement, i.e.

$$(2) \qquad \lim_{n \to \infty} P(a < Y_n^* < b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2}\, dx,$$

meaning that,

$$(3) \qquad \sum_{np + a\sqrt{npq} \leq k \leq np + b\sqrt{npq}} \binom{n}{k} p^k q^{n-k} \longrightarrow \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2}\, dx \quad \text{as } n \to \infty.$$

Proof:

We are going to use Stirling's formula:

$$(4) \qquad n! = \sqrt{2\pi n}\, \frac{n^n}{e^n}\, e^{\frac{\theta_n}{12n}} \qquad 0 < \theta_n < 1.$$

Let

$$w_{n,k} = \binom{n}{k} p^k q^{n-k} = \frac{n!}{k!\,(n-k)!} p^k q^{n-k}.$$

Then by (4):

$$w_{n,k} = \frac{1}{\sqrt{2\pi}} \sqrt{\frac{n}{k(n-k)}} \frac{n^n p^k q^{n-k}}{k^k (n-k)^{n-k}} e^{\frac{\theta_n}{12n} - \frac{\theta_k}{12k} - \frac{\theta_{n-k}}{12(n-k)}} .$$

Let

$$R = \frac{\theta_n}{12n} - \frac{\theta_k}{12k} - \frac{\theta_{n-k}}{12(n-k)} .$$

Then

$$w_{n,k} = \frac{1}{\sqrt{2\pi\, npq}} \left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k} \sqrt{\frac{n^2 pq}{k(n-k)}}\; e^R .$$

Define $x = x_k$ such that:

(5) $k = np + x\sqrt{npq}$ then obviously

(6) $n-k = nq - x\sqrt{npq}$.

Furthermore

$$w_{n,k} = \frac{1}{\sqrt{2\pi\, npq}} e^{-k \log \frac{k}{np} - (n-k)\log\frac{n-k}{nq} - \frac{1}{2}\log\frac{k}{np} - \frac{1}{2}\log\frac{n-k}{nq} + R} .$$

Let

$$\alpha = (k+\tfrac{1}{2})\log\frac{k}{np} + (n-k+\tfrac{1}{2})\log\frac{n-k}{nq} - R.$$

Then, using (5) and (6) we get:

(7) $$\alpha = (k+\tfrac{1}{2})\log\left(1+\frac{x\sqrt{q/p}}{\sqrt{n}}\right) + (n-k+\tfrac{1}{2})\log\left(1-x\frac{\sqrt{p/q}}{\sqrt{n}}\right) - R.$$

Obviously:

$$w_{n,k} = \frac{1}{\sqrt{2\pi npq}} e^{-\alpha} .$$

We will make use of the expansions:

$$(8) \qquad \log(1+y) = y - \frac{y^2}{2} + r_1(y)$$

$$(9) \qquad \log(1-y) = -y - \frac{y^2}{2} + r_2(y)$$

where $|r_1(y)| \leqq |y|^3$ for $|y| \leqq \frac{1}{2}$

$|r_2(y)| \leqq |y|^3$ for $|y| \leqq \frac{1}{2}$.

Then, using (5), (6), (8), (9) in (7) we have:

$$\alpha = (np + x\sqrt{npq} + \tfrac{1}{2})\left(x \frac{\sqrt{q/p}}{\sqrt{n}} - x^2 \frac{q/p}{2n} + r_1\left(x \frac{\sqrt{q/p}}{\sqrt{n}}\right)\right) +$$

$$+ (nq - x\sqrt{npq} + \tfrac{1}{2})\left(-x \frac{\sqrt{p/q}}{\sqrt{n}} - x^2 \frac{p/q}{2n} + r_2\left(x \frac{\sqrt{p/q}}{\sqrt{n}}\right)\right) - R .$$

We restrict the estimation for $|x| \leqq A$.

Then

$$\alpha = (x\sqrt{npq} - \frac{x^2}{2} q + x^2 q) + (-x\sqrt{npq} - \frac{x^2}{2} p + x^2 p) + \bar{R}$$

where

$$|\bar{R}| \leqq \frac{K(A,p)}{\sqrt{n}} \qquad \text{for } |x| \leqq A \text{ and } n > n_0(A\ p).$$

Since $p + q = 1$, it follows that

$$w_{n,k} = \frac{1}{\sqrt{2\pi npq}} e^{-x^2/2 - \bar{R}} .$$

Also since

$$|e^{-\bar{R}} - 1| \leq 2|\bar{R}| \qquad \text{for} \quad |\bar{R}| \leq 1,$$

we have

$$|w_{n,k} - \frac{e^{-x^2/2}}{\sqrt{2\pi npq}}| \leq \frac{2K(A,p)}{\sqrt{2\pi pq} \cdot n} .$$

So far we have proved a local theorem, namely we have obtained an estimate for one term $w_{n,k}$.

Now let $b > a$ be real numbers. Then:

$$(10) \quad |\sum_{a\sqrt{npq} \leq k-np \leq b\sqrt{npq}} \left(w_{n,k} - \frac{e^{-x_k^2/2}}{\sqrt{2\pi npq}}\right)| \leq$$

$$\leq \frac{2K(a,b,p)}{\sqrt{2\pi pq} \cdot n} (b-a)\sqrt{npq} = \frac{K'}{\sqrt{n}} \to 0 \quad \text{as} \quad n \to \infty.$$

But

$$\sum_{a \leq \frac{k-np}{\sqrt{npq}} \leq b} \frac{1}{\sqrt{npq}} \frac{e^{-\left(\frac{k-np}{\sqrt{npq}}\right)^2/2}}{\sqrt{2\pi}} \longrightarrow \int_a^b \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx \qquad \text{as} \quad n \to \infty,$$

since the expression on the left can be recognized as the Riemann sums in the interval $[a, b]$.

Combining this with (10) we obtain the conclusion of the theorem:

$$\sum_{np+a\sqrt{npq} \leq k \leq np+b\sqrt{npq}} w_{n,k} \longrightarrow \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx,$$

with a rate of $\frac{1}{n}$ as $n \to \infty$.

As we have seen this method is rather elaborate and involves a lot of calculations. We will be able to prove the de Moivre-Laplace theorem as a special case of a more general theorem in a very short way in Chapter III by using the method of characteristic functions.

In the following section of this chapter we are going to introduce another direct method which will enable us to prove more general C.L.T.'s than the de Moivre-Laplace theorem.

1.2. Proof of the Central Limit Theorem Using Heat Equations.

1.21. The case of independent identically distributed r.v.'s.

A direct method for proving the C.L.T. for independent identically distributed random variables was presented by Petrovsky and Kolmogorov [14]. In their proof they made use of the so called 'heat equations' (see (17)) realizing that the distribution function of the normal variable is a solution of this differential equation. In our presentation we will follow M.Rosenblatt's [29] version of the original proof.

Theorem 2.

Given a sequence of independent identically distributed random variables $X_1, \ldots, X_n$ with mean m and with finite variance σ^2, we have for any x,

$$(11) \quad \lim_{n\to\infty} P\left[\frac{S_n - n\cdot m}{\sigma\sqrt{n}} < x\right] = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2}\, du,$$

where $S_n = \sum_{j=1}^{n} X_j$.

Remark.

Without loss of generality instead of random variables with mean m and variance σ^2 we can work with the new variables:

$$(12) \quad X_j' = \frac{X_j - m}{\sigma}$$

which have mean 0 and variance 1.

In the following we will consider this type of variables.

Let's rewrite the C.L.T. in this modified new form.

$$(13) \quad \lim_{n\to\infty} P\left[\frac{S_n}{\sqrt{n}} < x\right] = \Phi(x).$$

Proof:

Consider the functions:

$$(14) \quad U_{nk}(x) = P\left(\sum_{j=1}^{k} X_j/\sqrt{n} < x\right).$$

By the independence of the r.v.'s X_j we have,

$$(15) \quad U_{nk}(x) = \int U_{n(k-1)}(x-\xi)dF_k(\xi) \qquad \text{for } 1 < k \leq n,$$

where F_k denotes the distribution function of $X_k/\sqrt{n}$.

By the notation (14) $U_{nn}(x)$ is the distribution function of $\sum_{j=1}^{n} X_j/\sqrt{n}$ which we will denote by $U_n(x)$ from now on, i.e. the C.L.T. restated with the new symbols is as follows:

$$(16) \quad \lim_{n\to\infty} U_n(x) = \Phi(x).$$

Now we will introduce the above mentioned heat equation:

$$(17) \quad \frac{\partial \Phi}{\partial t} = \frac{1}{2}\frac{\partial^2 \Phi}{\partial x^2}.$$

The crucial observation is that the function $\Phi(x/\sqrt{t})$ is a solution of (17) in the half-plane $t > 0$.

We will introduce 'upper' and 'lower' functions $V(x,t)$ and $T(x,t)$, which will be essential to our proof,

(18) $V(x,t) = \Phi(x/\sqrt{t}) + \epsilon t$

(19) $T(x,t) = \Phi(x/\sqrt{t}) - \epsilon t.$

Clearly, $V(x,t)$ is a solution of the partial differential equation

$$(20)\quad \frac{\partial V}{\partial t} = \frac{1}{2}\frac{\partial^2 V}{\partial x^2} + \epsilon$$

and $T(x,t)$ solves the equation

$$(21)\quad \frac{\partial T}{\partial t} = \frac{1}{2}\frac{\partial^2 T}{\partial x^2} - \epsilon .$$

Our aim is to replace each distribution function $F_j(x)$ by the distribution function $\Phi(\sqrt{n}\cdot x)$ and show that the replacement results in a negligible error only.

We need two lemmas to give an estimate on the error made in the replacement.

<u>Lemma 1.</u>

Given any $\delta > 0$, there exists an $n(\delta, \epsilon)$ such that

$$(22)\quad V(x, t+\frac{1}{n}) > \int V(x-\xi, t)\, dF_n(\xi)$$

in the half-space $t > \delta$.

Proof:

By the Taylor expansion we have:

$$(23) \quad V(x-\xi, t) = V(x,t) - \xi \frac{\partial V}{\partial x}(x,t) + \tfrac{1}{2}\xi^2 \frac{\partial^2 V}{\partial x^2}(x,t) + \rho(x, \xi, t),$$

where

$$(24) \quad \rho(x, \xi, t) = \tfrac{1}{2}\xi^2 \left[\frac{\partial^2 V}{\partial x^2}(x - \theta\xi, t) - \frac{\partial^2 V}{\partial x^2}(x,t) \right] \qquad 0 < \theta < 1.$$

Using that:

$$(25) \quad \int dF_n(\xi) = 1 \qquad \int \xi \, dF_n(\xi) = 0 \qquad \int \xi^2 dF_n(\xi) = \frac{1}{n}$$

we get

$$(26) \quad \int V(x-\xi, t)\, dF_n(\xi) = V(x,t) + \frac{1}{2n} \frac{\partial^2 V}{\partial x^2}(x,t) + J \quad ,$$

where

$$(27) \quad J = \int \rho(x, \xi, t)\, dF_n(\xi).$$

But

$$(28) \quad |\rho(x, \xi, t)| < \xi^2/\delta \quad \text{and}$$

$$(29) \quad |\rho(x, \xi, t)| < |\xi|^3/\delta^{3/2} \qquad \text{for } t \geqq \delta.$$

These above estimates (28) and (29) can be obtained the following way:

Consider the function

$$(30) \quad \omega(\alpha) = \frac{1}{\sqrt{2\pi}} e^{-\alpha^2/2},$$

then obviously

$$(31) \quad \omega(\alpha) \leqq \frac{1}{\sqrt{2\pi}} < 1.$$

By a simple investigation of the derivatives we can get estimates on the maximum value of the functions $|\alpha|\varphi(\alpha)$ and $\alpha^2\varphi(\alpha)$:

$$(32) \quad |\alpha|\varphi(\alpha) \leqq 1\cdot\varphi(1) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}} < \tfrac{1}{2} \qquad \text{for all } \alpha.$$

$$(33) \quad \alpha^2\varphi(\alpha) \leqq (\sqrt{2})^2\varphi(\sqrt{2}) = \frac{2}{\sqrt{2\pi}}\, e^{-1} < 1 \qquad \text{for all } \alpha.$$

Now using equations (17), (20), (32) and (33), we get

$$(34) \quad \left|\tfrac{1}{2}\frac{\partial^2 v}{\partial x^2}\right| = \left|-\frac{x}{t^{3/2}}\,\varphi\!\left(\frac{x}{\sqrt{t}}\right)\right| = \left(\frac{|x|}{\sqrt{t}}\varphi\!\left(\frac{x}{\sqrt{t}}\right)\right)\cdot\frac{1}{t} \leqq \frac{1}{2t}$$

for $t > 0$, and for all x.

$$(35) \quad \left|\tfrac{1}{2}\frac{\partial^3 v}{\partial x^3}\right| = \varphi\!\left(\frac{x}{\sqrt{t}}\right)\left|\frac{x^2}{2t^{5/2}} - \frac{1}{2t^{3/2}}\right| \leqq \left(\frac{x^2}{t}\varphi\!\left(\frac{x}{\sqrt{t}}\right)\right)\cdot\frac{1}{2t^{3/2}} + \varphi\!\left(\frac{x}{\sqrt{t}}\right)\frac{1}{2t^{3/2}} \leqq$$

$$\leqq \frac{1}{2t^{3/2}} + \frac{1}{2t^{3/2}} = \frac{1}{t^{3/2}}.$$

Now we recall equation (24) and make use of the mean value theorem getting:

$$(36) \quad \rho(x,\xi,t) = \tfrac{1}{2}\xi^2\left[\frac{\partial^2 v}{\partial x^2}(x-\theta\xi,t) - \frac{\partial^2 v}{\partial x^2}(x,t)\right] =$$

$$= \tfrac{1}{2}\xi^2\ \frac{\partial^3 v}{\partial x^3}(x-\theta'\xi,t)(-\theta\xi).$$

Substituting (34) and (35) into the above two equalities respectively, we obtain for $t \geqq \delta$:

$$|\rho(x,\xi,t)| \leqq \frac{\xi^2}{\delta}$$

and

$$|\rho(x,\xi,t)| \leqq \xi^2\cdot\frac{1}{\delta^{3/2}}\cdot|\xi| = \frac{|\xi|^3}{\delta^{3/2}}$$

which is what we stated in (28) and (29).

Making use of (29) we get:

$$(37)\quad |\rho(x,\xi,t)| < \frac{\epsilon}{3}\,\xi^2 \qquad \text{in the half-plane } t > \delta \text{ for } |\xi| \leqq \tau = \frac{\epsilon}{3}\cdot\delta^{3/2}.$$

Dividing the region of integration into two sets we obtain

$$|J| \leqq \int_{|\xi|\leqq\tau} |\rho(x,\xi,t)|\, dF_n(\xi) + \int_{|\xi|>\tau} |\rho(x,\xi,t)|\, dF_n(\xi)$$

$$(38)\qquad \leqq \frac{\epsilon}{3} \int_{|\xi|\leqq\tau} \xi^2 dF_n(\xi) + \frac{1}{\delta} \int_{|\xi|>\tau} \xi^2 dF_n(\xi)$$

$$(39)\qquad \leqq \frac{\epsilon}{3}\cdot\frac{1}{n} + \frac{1}{\delta n} \int_{|y|>\tau\sqrt{n}} y^2 dF(y)$$

where F(y) denotes the common distribution functions of the variables X_j.

Since $\int y^2 dF(y)$ is finite, we have:

$$(40)\quad |J| < \frac{2}{3}\cdot\frac{\epsilon}{n}\ , \text{ if } n \text{ is large enough.}$$

Using the relation (20) in the equation (26) and considering the estimate (40) for J it follows that:

$$(41)\quad \int V(x-\xi,t)\, dF_n(\xi) < V(x,t) + \frac{1}{n}\frac{\partial V}{\partial t} - \frac{\epsilon}{3}\cdot\frac{1}{n}\ .$$

Consider the Taylor expansion:

$$(42)\quad V(x,t+\frac{1}{n}) = V(x,t) + \frac{1}{n}\frac{\partial V}{\partial t} + \frac{1}{2n^2}\left[\frac{\partial^2 V}{\partial t^2}\right]_{x,\,t+\frac{\theta}{n}} \qquad 0 < \theta < 1$$

and the inequality

$$(43)\quad \left|\frac{\partial^2 V}{\partial t^2}\right| < \frac{1}{\delta^2} \qquad \text{when } t > \delta$$

to obtain:

$$(44)\quad V(x,t+\frac{1}{n}) > V(x,t) + \frac{1}{n}\frac{\partial V}{\partial t} - \frac{1}{2n^2\delta^2}\ .$$

If n is sufficiently large

(45) $\frac{\epsilon}{3} \cdot \frac{1}{n} > \frac{1}{2n^2 \delta^2}$ and therefore

(46) $V(x, t+\frac{1}{n}) > V(x,t) + \frac{1}{n} \frac{\partial V}{\partial t} - \frac{\epsilon}{3} \cdot \frac{1}{n}$.

Combining (41) and (46) we get the conclusion of Lemma 1.

Lemma 2.

If G_1 and G_2 are distribution functions with mean zero and variances less than β then:

(47) $G_1(x) - G_2(x+2\alpha) \leqq \beta/\alpha^2$ for all x and $\alpha > 0$.

Proof.

If $x \leqq -\alpha$

(48) $G_1(x) \leqq G_1(-\alpha) \leqq \beta/\alpha^2$

by Chebyshev's inequality.

Hence

(49) $G_1(x) - G_2(x+2\alpha) \leqq \beta/\alpha^2$.

Now consider the case when $x > -\alpha$.

(50) $G_2(x+2\alpha) \geqq G_2(\alpha) \geqq 1 - \beta/\alpha^2$

And therefore:

(51) $G_1(x) - G_2(x+2\alpha) \leqq 1 - G_2(x+2\alpha) \leqq \beta/\alpha^2$

which proves Lemma 2.

We are now ready to complete the proof of the C.L.T..

Take a fixed δ, $0 < \delta < 1$.

For some value of s $(s = 1, \ldots, n)$

$$(52) \quad \delta < \frac{s}{n} < 2\delta .$$

Now consider the distribution function $U_{ns}(x)$. It has mean zero and variance $\frac{s}{n} < 2\delta$.

Then $\Phi(x/\sqrt{\frac{s}{n}})$ has the same mean and variance as $U_{ns}(x)$, therefore we can apply Lemma 2 for all x and $\alpha > 0$.

$$(53) \quad U_{ns}(x) - \Phi\left(\frac{x+2\alpha}{\sqrt{s/n}}\right) < \frac{2\delta}{\alpha^2} .$$

Therefore:

$$(54) \quad U_{ns}(x) - V(x+2\alpha, \frac{s}{n}) < \frac{2\delta}{\alpha^2} .$$

By Lemma 1

$$(55) \quad V(x+2\alpha, \frac{k}{n}) > \int V(x+2\alpha-\xi, \frac{k-1}{n}) \, dF_n(\xi) \qquad \text{for } k > s.$$

Introduce notation

$$(56) \quad W_k(x) = U_{nk}(x) - V(x+2\alpha, \frac{k}{n}).$$

From (15) and (55) it follows that

$$(57) \quad W_k(x) < \int W_{k-1}(x-\xi) \, dF_n(\xi).$$

If we denote the least upper bound of $W_k(x)$ by μ_k then (57) implies that

$\mu_k \leqq \mu_{k-1}$ for $k > s$, consequently,

$\mu_n \leqq \mu_s$.

Furthermore

(58) $U_n(x) - V(x+2\alpha, 1) = U_n(x) - \Phi(x+2\alpha) - \epsilon \cdot 1 \leq \mu_s < \frac{2\delta}{\alpha^2}$.

Using (58) we obtain:

$$(59) \quad U_n(x) < \Phi(x) + \frac{1}{\sqrt{2\pi}} \int_x^{x+2\alpha} e^{-u^2/2} du + \epsilon + \frac{2\delta}{\alpha^2} <$$

$$< \Phi(x) + \frac{2\alpha}{\sqrt{2\pi}} + \epsilon + \frac{2\delta}{\alpha^2} .$$

Let

$$\alpha = \frac{\sqrt{2\pi}}{4} \epsilon, \quad \delta = \frac{\alpha^3}{\sqrt{2\pi}},$$

then we have:

(60) $U_n(x) < \Phi(x) + 2\epsilon$ for n large enough.

Using the 'lower function' $T(x, t)$, in the same manner we can obtain:

(61) $U_n(x) > \Phi(x) - 2\epsilon$ for n large enough.

(60) and (61) together imply the conclusion of the C.L.T.

1.22. The case of non-identically distributed random variables.

In 1922 Lindeberg [22] found a sufficient condition for the C.L.T. to hold in the case of independent variables of arbitrary distribution with finite mean and variance:

Let $\{X_n\}$ be a sequence of independent r.v.'s with finite expectation EX_j and finite variance $\sigma^2(X_j) = \sigma^2_j$. We will introduce the following notations.

$$S_n = X_1 + X_2 + \ldots + X_n, \qquad s_n^2 = \sigma^2(S_n)$$

then obviously

$$s_n^2 = \sum_{j=1}^{n} \sigma_j^2, \quad \text{also} \quad ES_n = \sum_{j=1}^{n} EX_j.$$

Then in order that

$$\lim_{n\to\infty} P\left(\frac{S_n - ES_n}{s_n^2} < x\right) = \Phi(x)$$

the following is a sufficient condition:

$$(61) \quad \text{for all } \eta > 0 \quad \sum_{j=1}^{n} \frac{1}{s_n^2} \int_{|x|>\eta s_n} x^2\, dF_j(x) \to 0 \quad \text{as } n\to\infty,$$

where F_j denotes the distribution function of X_j. Condition (61) is called the 'Lindeberg condition'.

We are going to refer to the following assumptions as Assumptions (1):

We are given a double array of random variables $\{X_{nj}\}$ ($n=1,2,\ldots$; $j=1,2,\ldots N_n$) with independence between the r.v.'s in each row, without loss of generality, we can assume that $EX_{nj} = 0$ for all n and j and $\sum_{j=1}^{N_n} \sigma_{nj}^2 = 1$ for all n. The Lindeberg condition for such a double array can be formulated as follows:

$$(62) \quad \text{for all } \eta > 0 \quad \sum_{j=1}^{N_n} \int_{|x|>\eta} x^2\, dF_{nj}(x) \to 0 \quad \text{as } n\to\infty,$$

where F_{nj} denotes the distribution function of the variable X_{nj}.

Going through the proof of Petrovsky and Kolmogorov our idea was that the same techniques could be used for not identically distributed r.v.'s for which the Lindeberg condition (61) holds. Following the steps of the proof it turned out that this condition was not quite enough and one needed a stronger Lindeberg type condition when trying to deduce equation (40) from equation (39). In the case of a double array

$$\{X_{nj}\} \quad (n = 1, 2, \ldots; \; j = 1, 2, \ldots N_n)$$

with independence in each row, all the variables having zero expectation and finite variances such that $\sum_{j=1}^{N_n} \sigma_{nj}^2 = 1$, one needed the following inequality to hold for $\eta > 0$, $\lambda > 0$ and $N_0(\lambda, \eta)$:

$$(63) \qquad \int_{|X_{nj}|>\eta} X_{nj}^2 \, dP < \sigma_{nj}^2 \cdot \lambda \qquad \text{for } n > N_0(\lambda, \eta).$$

At this point one could argue the following way: Assuming that the Lindeberg condition (62) holds for the double array $\{X_{nj}\}$, the measure of the set of j's for which $\int_{|X_{nj}|>\eta} X_{nj}^2 \, dP > \sigma_{nj}^2 \cdot \lambda$, tends to zero as n tends to infinity. Therefore we can forget about such variables X_{nj} and concentrate our attention to those r.v.'s for which (63) is true.

Still we feel it is more elegant if one doesn't simply disregard these 'bad' variables but finds a way to construct an equivalent double array $\{Y_{nj}\}$ having the property that all its variables are 'good' ones. In other words all its variables satisfy what we are going to call the 'strong Lindeberg condition':

Given $\eta, \lambda > 0$, there exists $N_0(\eta, \lambda)$ such that for all j:

$$(64) \qquad \frac{\int_{|Y_{nj}|>\eta} Y_{nj}^2 \, dP}{\sigma^2(Y_{nj})} < \lambda \qquad \text{for } n \geq N_0(\eta, \lambda).$$

The fact that the construction of such an equivalent double array $\{Y_{nj}\}$ is possible is shown in Theorem 3.

Before going into the proof of Theorem 3 we would like to prove Lemma 3 which shows a property of double arrays satisfying the Lindeberg condition (62).

Lemma 3.

Let Assumptions (1) be assumed for the double array $\{X_{nj}\}$. Also let $\{X_{nj}\}$ satisfy condition (62) (Lindeberg). Then given $\epsilon > 0$, there exists $N_0(\epsilon)$ such that

(65) $$\max_{j=1}^{N_n} \sigma_{nj}^2 < \epsilon \quad \text{for } n > N_0(\epsilon).$$

Proof:

$$\sigma_{nj}^2 = \int_{|X_{nj}| \leq \sqrt{\epsilon/2}} X_{nj}^2 \, dP + \int_{|X_{nj}| > \sqrt{\epsilon/2}} X_{nj}^2 \, dP$$

$$\leq \frac{\epsilon}{2} + \sum_{j=1}^{N_n} \int_{|X_{nj}| > \sqrt{\epsilon/2}} X_{nj}^2$$

$$< \frac{\epsilon}{2} + \frac{\epsilon}{2} \qquad \text{for } n > N_0(\epsilon).$$

This latter inequality follows from the Lindeberg condition. Thus we have:

$$\sigma_{nj}^2 < \epsilon \quad \text{for all } j, \text{ for } n > N_0(\epsilon),$$

implying (65).

Theorem 3.

Given a double array of r.v.'s $\{X_{nj}\}$ ($n=1,2,\ldots$; $j=1,2,\ldots N_n$) for which the Lindeberg condition (62) holds, then one can obtain a new double array of r.v.'s $\{Y_{nj}\}$ ($n=1,2,\ldots$; $j=1,2,\ldots M_n$) from the original one for which the following conditions hold:

(66) For all n: $$\sum_{j=1}^{N_n} X_{nj} = \sum_{j=1}^{M_n} Y_{nj}.$$

For any given $\eta, \lambda > 0$ there exists $N_0 = N_o(\eta,\lambda)$ such that

(67) $$\frac{\int_{|Y_{nj}|>\eta} Y_{nj}^2\, dP}{\int Y_{nj}^2\, dP} < \lambda \quad \text{for } n \geqq N_0 \text{ and } j=1,2,\ldots,M_n.$$

Actually we will construct the Y_{nj}'s to be the same as the X_{nj}'s with the exception of one (we will call it Y_{n1}) which will be the sum of some X_{nj}'s. Therefore if the variables X_{nj} are independent, so will be the variables Y_{nj}.

The idea of the construction is that the variables X_{nj} for which the fraction (67) is greater than λ don't have much weight (the sum of their variances is very small if n is large). We will add to these 'bad' variables as many 'good' ones as to make the sum, regarded as one variable, a 'good' variable.

Clearly it is enough to prove the following:

For all s, there exists n_s ($1 = n_1 < n_2 < \ldots$) such that, for all $n \geqq n_s$ there exists a set of integers $C_{n,s} \subset \{1,2,\ldots,N_n\}$ satisfying:

(68) $$\frac{\int_{|X_{nj}|>1/s} X_{nj}^2\, dP}{\int X_{nj}^2\, dP} \leqq \frac{1}{s} \quad \text{for } j \notin C_{ns}$$

$$(69)\qquad \frac{\displaystyle\int_{\left|\sum_{j\in C_{ns}} X_{nj}\right|>1/s}\Big(\sum_{j\in C_{ns}} X_{nj}\Big)^2 dP}{\displaystyle\int\Big(\sum_{j\in C_{ns}} X_{nj}\Big)^2 dP} \leq \frac{1}{s}.$$

In fact once this is proved we can do the following construction:

for $n_s \leqq n < n_{s+1}$

$$Y_{n1} = \sum_{j\in C_{ns}} X_{nj} \qquad \text{and} \qquad Y_{n2},\ Y_{n3},\ \text{etc.}$$

are equal to the variables X_{nj} $(j \notin C_{ns})$ in some order.

Now, given $\eta, \lambda > 0$, choose s so large that $\frac{1}{s} < \eta$ and $\frac{1}{s} < \lambda$. Then, for $n \geqq n_s$ $(=N_0(\eta, \lambda))$ and all j we have:

$$\frac{\displaystyle\int_{|Y_{nj}|>\eta} Y_{nj}^2\, dP}{\displaystyle\int Y_{nj}^2\, dP} \leqq \frac{\displaystyle\int_{|Y_{nj}|> 1/s} Y_{nj}^2\, dP}{\displaystyle\int Y_{nj}^2\, dP} \leqq \frac{1}{s} < \lambda$$

in accordance with Theorem 3.

We will introduce the following notations:

$$\sigma_{nj\eta}^2 = \int_{|X_{nj}|>\eta} X_{nj}^2\, dP \qquad \text{and}$$

$$\sigma_{nj}^2 = \int X_{nj}^2\, dP.$$

From the Lindeberg condition we have

$$(70)\qquad \delta(\eta, n) = \sum_{j=1}^{N_n} \int_{|X_{nj}|>\eta} X_{nj}^2\, dP \to 0 \qquad \text{as } n\to\infty.$$

Then by Lemma 3 we have $\lim_{n\to\infty} \max_{j=1}^{N_n} \sigma_{nj}^2 = 0$.

In the proof of Theorem 3 we are going to make use of the following inequality (p.90, Billingsley [5]):

If $X_1, X_2, \ldots, X_m$ are independent r.v.'s with zero expectation,

$Y = X_1 + X_2 + \ldots + X_m$

$\sigma^2 = EY^2$

then for any $\alpha > 0$:

$$(71) \quad \int_{\left\{\frac{Y^2}{\sigma^2} \geq \alpha\right\}} \frac{Y^2}{\sigma^2}\, dP \leq K\left[\frac{1}{\alpha} + \sum_{j=1}^{m} \int_{\left\{\left|\frac{X_j}{\sigma}\right| \geq \frac{1}{4}\alpha\right\}} \frac{X_j^2}{\sigma^2}\, dP\right]$$

where $K > 1$ is a universal constant.

Now we are about to construct the set C_{ns} and the number n_s we were referring to:

Choose $n_1 = 1$, $C_{n1} = \emptyset$ for all n (obviously satisfying (68) and (69)). Suppose the numbers $n_1, \ldots, n_{s-1}$ have already been chosen.

Define

$$A_{ns} = \left\{ j : \frac{\sigma^2_{nj\frac{1}{s}}}{\sigma^2_{nj}} \geq \frac{1}{4Ks} \right\} .$$

Clearly

$$(72) \quad \sum_{j \in A_{ns}} \sigma^2_{nj} \leq 4Ks \sum_{j \in A_{ns}} \sigma^2_{nj\frac{1}{s}} \leq 4Ks\, \delta\left(\frac{1}{s}, n\right).$$

Choose a positive number μ so small that it satisfies:

$$(73) \quad 2s^2K\mu^2 < \frac{1}{4s} \; ; \quad 2\mu^2 \leq 1 \quad \text{and} \quad \frac{1}{4\sqrt{2}\, s^2 \mu} \geq \frac{1}{s} .$$

By (62) and (65) we can choose n_s so large that it satisfies:

(74) $n_s > n_{s-1}$

(75) $K \frac{4Ks\,\delta(1/s,\, n)}{\mu^2} < \frac{1}{4s}$ for $n \geqq n_s$

(76) $\max_{j=1}^{N_n} \sigma_{nj}^2 < \mu^2$ for $n \geqq n_s$

From (72) and (75) we have $\sum_{j \in A_{ns}} \sigma_{nj}^2 < \mu^2$ for $n \geqq n_s$, this together with (76) implies that we can choose a set B_{ns} of integers disjoint to A_{ns} such that

(77) $\mu^2 \leqq \sum_{j \in A_{ns}} \sigma_{nj}^2 + \sum_{j \in B_{ns}} \sigma_{nj}^2 \leqq 2\mu^2$.

Put $C_{ns} = A_{ns} \cup B_{ns}$.

Since $A_{ns} \subset C_{ns}$ we have for $n \geqq n_s$ and all j·

$$\frac{\sigma_{nj\frac{1}{s}}^2}{\sigma_{nj}^2} < \frac{1}{4Ks} < \frac{1}{s} .$$

It remains to check (69), i.e. the validity of the inequality:

$$\frac{\int_{|\sum_{j \in C_{ns}} X_{nj}| > \frac{1}{s}} (\sum_{j \in C_{ns}} X_{nj})^2 \, dP}{\int (\sum_{j \in C_{ns}} X_{nj})^2 \, dP} \leqq \frac{1}{s} .$$

Applying inequality (71) with

$$\alpha = \frac{1}{s^2\sigma^2} , \quad Y = \sum_{j \in C_{ns}} X_{nj}$$

we obtain:

$$(78)\qquad \int_{|Y|>\frac{1}{s}} \frac{Y^2}{\sigma^2} \leqq Ks^2\sigma^2 + \frac{K}{\sigma^2} \sum_{j\in C_{ns}} \sigma_{nj}^2 \frac{1}{4s^2\sigma}$$

and using the inequalities (77), (73)

$$(\mu^2 \leqq \sigma^2 \leqq 2\mu^2 \quad \text{and} \quad \frac{1}{4s^2\sigma} \geqq \frac{1}{4\sqrt{2}\, s^2\mu} \geqq \frac{1}{s})$$

and then (72) and (75) we get that the expression on the right of (78) is at most:

$$2Ks^2\mu^2 + \frac{K}{\mu^2} \sum_{j\in A_{ns}} \sigma_{nj}^2 + \frac{K}{\mu^2} \sum_{j\in B_{ns}} \sigma_{nj}^2 \frac{1}{s} \leqq$$

$$\leqq \frac{1}{4s} + \frac{K}{\mu^2} 4Ks\,\delta(\frac{1}{s}, n) + \frac{K}{\mu^2} \frac{1}{4Ks} \sum_{j\in B_{ns}} \sigma_{nj}^2 \leqq$$

$$\leqq \frac{1}{4s} + \frac{1}{4s} + \frac{1}{4s} \frac{\sigma^2}{\mu^2} \leqq \frac{1}{s} \;.$$

This completes the proof, i.e. we have found sets C_{ns} so that (68) and (69) hold and therefore all we need to do is to construct the double array Y_{nj} in the previously described manner.

Because of this theorem, in the proof of the C.L.T. with Lindeberg condition we can assume, that the double array of random variables $\{X_{nj}\}$ satisfies the strong Lindeberg condition (64) rather than just condition (62).

After these introductory remarks we are ready to prove the C.L.T. for a double array of r.v.'s with independence in each row and for which the Lindeberg condition holds again following Petrovsky's and Kolmogorov's idea.

Theorem 4.

Given a double array of r.v.'s $\{X_{nj}\}$ $(n=1,2,\ldots; j=1,2,\ldots,N_n)$ with independence between the variables in each row, $E X_{nj} = 0$ for all n and j, $\sum_{j=1}^{N_n} \sigma_{nj}^2 = 1$ for all n, the variables satisfying the Lindeberg condition (62), then

$$(79) \quad \lim_{n\to\infty} P(S_n = \sum_{j=1}^{N_n} X_{nj} < x) = \Phi(x).$$

Proof:

We are going to make use of the same techniques as we have seen in the proof for identically distributed random variables.

Let $F_{nj}(x)$ be the distribution function of X_{nj} $(j=1,2,\ldots N_n; n=1,2,\ldots)$ and $U_{nk}(x)$ be the distribution function of $\sum_{j=1}^{k} X_{nj}$ $(k=1,2,\ldots,N_n)$.

As the r.v.'s are mutually independent,

$$(80) \quad U_{nk}(x) = P(\sum_{j=1}^{k} X_{nj} < x)$$

$$= \int U_{n(k-1)}(x-\xi)\, dF_{nk}(\xi) \qquad 1 < k \leqq N_n .$$

$U_n(x) = U_{nN_n}(x)$ is the distribution function of $S_n = \sum_{j=1}^{N_n} X_{nj}$, i.e. restating the theorem in the new notation, we want to show that

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

$\Phi(x)$ denoting the distribution function of the standard normal variable. We are going to use the heat equation (17) and the 'upper' function $V(x,t)$ introduced in (18) and (20).

The idea of the proof this time is to replace each distribution function $F_{nj}(x)$ by the distribution function $\Phi\left(\frac{x}{\sigma_{nj}}\right)$, and we are going to show that we are justified to do so, since the overall error made is negligible. The lemmas which give estimates on the error will be modified as follows:

Lemma 4.

Given any $\delta > 0$, there is an n (depending on δ, ϵ) sufficiently large such that

(81) $$V(x, t+\sigma_{nj}^2) > \int V(x-\xi, t)\, dF_{nj}(\xi)$$

in the half plane $t > \delta$.

The proof starts out the same way as in Theorem 2.

(82) $$V(x-\xi, t) = V(x, t) - \xi\frac{\partial V}{\partial x}(x, t) + \tfrac{1}{2}\xi^2\frac{\partial^2 V}{\partial x^2}(x, t) + \rho(x, \xi, t)$$

(83) $$\rho(x, \xi, t) = \tfrac{1}{2}\xi^2\left[\frac{\partial^2 V}{\partial x^2}(x-\theta\xi, t) - \frac{\partial^2 V}{\partial x^2}(x, t)\right] \qquad 0 < \theta < 1.$$

Recalling that

$$\int dF_{nj}(\xi) = 1, \quad E(\xi_{nj}) = \int \xi\, dF_{nj}(\xi) = 0, \quad \int \xi^2\, dF_{nj}(\xi) = \sigma_{nj}^2$$

we get from (82):

(84) $$\int V(x-\xi, t)\, dF_{nj}(\xi) = V(x, t) + \frac{\sigma_{nj}^2}{2}\cdot\frac{\partial^2 V}{\partial x^2}(x, t) + J_{nj}$$

where

(85) $$J_{nj} = \int \rho(x, \xi, t)\, dF_{nj}(\xi),$$

(86) $$|\rho(x, \xi, t)| < \frac{\xi^2}{\delta} \qquad \text{for } t > \delta$$

since $\frac{\partial^2 V}{\partial x^2}$ is bounded by $\frac{1}{2\delta}$ in the half plane $t > \delta$. Also:

(87) $|\rho(x, \xi, t)| < |\xi|^3/\delta^{3/2}$ for $t > \delta$.

The inequalities (86) and (87) were derived in Theorem 2 so we are going to omit their proofs.

When $|\xi| \leqq \eta = \frac{\epsilon}{3}\delta^{3/2}$ using (87) we get:

(88) $|\rho(x, \xi, t)| < \frac{\epsilon}{3}\xi^2$.

Then:

(89) $$|J_{nj}| \leqq \int_{|\xi| \leqq \eta} |\rho(x, \xi, t)| \, dF_{nj}(\xi) + \int_{|\xi| > \eta} |\rho(x, \xi, t)| \, dF_{nj}(\xi) \leqq$$

$$\leqq \frac{\epsilon}{3} \int_{|\xi| \leqq \eta} \xi^2 dF_{nj}(\xi) + \frac{1}{\delta} \int_{|\xi| > \eta} \xi^2 dF_{nj}(\xi) \leqq$$

$$\leqq \frac{\epsilon}{3} \sigma_{nj}^2 + \frac{1}{\delta} \cdot \frac{\epsilon \cdot \delta}{3} \sigma_{nj}^2 .$$

In the above chain of inequalities first we've used (88) and (86) then we have made use of the strong Lindeberg condition i.e.

$$\int_{|\xi| > \eta} \xi^2 \, dF_{nj}(\xi) < \sigma_{nj}^2 \cdot \lambda \quad \text{for } n \geqq N_0(\eta, \lambda),$$

here $\lambda = \epsilon \cdot \delta$. Thus we obtain:

(90) $|J_{nj}| \leqq \frac{2}{3}\epsilon \cdot \sigma_{nj}^2$ for sufficiently large n.

It follows from (84), (90) and the heat equation that

(91) $$\int V(x-\xi, t) \, dF_{nj}(\xi) < V(x, t) + \sigma_{nj}^2 \frac{\partial V}{\partial t} - \sigma_{nj}^2 \epsilon + \frac{2}{3}\epsilon \cdot \sigma_{nj}^2 .$$

The relation

$$(92)\quad V(x, t+\sigma_{nj}^2) = V(x,t) + \sigma_{nj}^2 \frac{\partial V}{\partial t} + \frac{\sigma_{nj}^4}{2}\left[\frac{\partial^2 V}{\partial t^2}\right]_{x,\, t+\theta\sigma_{nj}^2} \qquad 0 < \theta < 1$$

implies that,

$$(93)\quad V(x, t+\sigma_{nj}^2) > V(x,t) + \sigma_{nj}^2 \cdot \frac{\partial V}{\partial t} - \frac{\sigma_{nj}^4}{\delta^2},$$

since $\left|\frac{\partial^2 V}{\partial t^2}\right| < \frac{1}{\delta^2}$ when $t > \delta$.

Recalling that $\sigma_{nj}^2 \to 0$ as $n \to \infty$ by Lemma 3, for sufficiently large n we have:

$$\frac{\sigma_{nj}^4}{\delta^2} < \frac{\epsilon}{3}\sigma_{nj}^2 \quad \text{implying that:}$$

$$(94)\quad V(x, t+\sigma_{nj}^2) > V(x,t) + \sigma_{nj}^2 \frac{\partial V}{\partial t} - \frac{\epsilon}{3}\sigma_{nj}^2.$$

From this last inequality (94) and from (91) Lemma 4 follows, i.e.:

$$V(x, t+\sigma_{nj}^2) > \int V(x-\xi, t)\, dF_{nj}(\xi).$$

Lemma 2 of Theorem 2 remains unaltered for our purposes. We can now complete the proof of the C.L.T.

Fix $\delta > 0$.

For $n \geq N_0$ and some s $(1 < s < N_n)$

$$(95)\quad \delta < \sum_{j=1}^{s} \sigma_{nj}^2 < 2\delta.$$

The distribution function U_{ns} will have variance $\sigma_n^2(s) = \sum_{j=1}^{s} \sigma_{nj}^2 < 2\delta$, and mean 0, but $\Phi\left(\frac{x}{\sigma_n(s)}\right)$ has the same property, namely, has mean 0 and variance less than 2δ.

By Lemma 2 for all x and all $\alpha > 0$:

$$(96) \quad U_{ns}(x) - \Phi\left(\frac{x+2\alpha}{\sigma_n(s)}\right) < \frac{2\delta}{\alpha^2},$$

and by (18) it follows that

$$(97) \quad U_{ns}(x) - V(x+2\alpha, \sigma_n^2(s)) < \frac{2\delta}{\alpha^2}.$$

Since $\sigma_n^2(s) > \delta$, for $k > s$, we have $\sigma_n^2(k-1) > \delta$, therefore we can apply Lemma 4 with $t = \sigma_n^2(k-1)$ and get:

$$(98) \quad V(x+2\alpha, \sigma_n^2(k)) > \int V(x+2\alpha - \xi, \sigma_n^2(k-1))\, dF_{nk}(\xi) \quad \text{for } k > s.$$

Let

$$W_{nk}(x) = U_{nk}(x) - V(x+2\alpha, \sigma_n^2(k)).$$

Using (80) and (98) we obtain:

$$W_{n(k+1)}(x) = \int U_{nk}(x)\, dF_{n(k+1)}(x) - V(x+2\alpha, \sigma_n^2(k+1) <$$

$$< \int W_{nk}(x)\, dF_{n(k+1)}(\xi).$$

Let μ_k be the least upper bound of $W_{nk}(x)$.

Since $\int dF_{nj}(x) = 1$, $\mu_{k+1} \leqq \mu_k$ $(k > s)$ and hence $\mu_{N_n} \leqq \mu_s$.

From this and (97) it follows that

$$U_n(x) - V(x+2\alpha, 1) = U_n(x) - \Phi(x+2\alpha) - \epsilon \leqq \mu_s < \frac{2\delta}{\alpha^2}.$$

Therefore

$$U_n(x) < \Phi(x) + \int_x^{x+2\alpha} e^{-u^2/2} + \epsilon + \frac{2\delta}{\alpha^2}$$

$$< \Phi(x) + \frac{2\alpha}{\overline{2\pi}} + \epsilon + \frac{2\delta}{\alpha^2}.$$

Let $\alpha = \frac{\sqrt{2\pi}}{4}\epsilon$, $\delta = \frac{\alpha^3}{\sqrt{2\pi}}$ then we get

(99) $U_n(x) < \Phi(x) + 2\epsilon$.

A lower bound can be obtained in a completely analogous manner by using the lower function $\Phi(x/\sqrt{t}) - \epsilon t$ which leads us to the inequality:

(100) $U_n(x) > \Phi(x) - 2\epsilon$.

Combining (99) and (100) we arrive at the conclusion of the C.L.T. (79).

Chapter II

The Operator's Method

2.1. Proof of the C.L.T. using operators

What we call today the operator's method was basically first introduced by Lindeberg in 1922 [22], although the name 'operator's method' has been given only recently to a modern version of the proof by H.F.Trotter [30], which utilizes Lindeberg's idea. Lindeberg's method appeared to be quite cumbersome and complicated, not so with Trotter's proof which is very clear and fairly simple. In the introduction of [30] the author emphasizes the fact that his approach is basically 'elementary', since it doesn't require the use of characteristic functions or any other such tools, i.e. it is a much more direct way of proving the problem.

What seems 'mystical' about the proof is that although he is proving the C.L.T., namely, the convergence of the distribution of the sum of certain r.v.'s to the normal distribution, this latter distribution doesn't appear explicitly. The only facts he uses are, that it has a finite variance and if ξ and η are independent normally distributed r.v.'s, then so are $\xi+\eta$, $a\xi+b$ (a, b arbitrary constants, $a \neq 0$). An important remark should be made at this point: the fact that the C.L.T. holds true implies that the above properties characterize the normal distribution (for which it is quite complicated to give a direct proof). More precisely it means the following. If a distribution function $\Phi(x)$ has the following properties:

1.) $\int_{-\infty}^{\infty} x\, d\Phi(x) = 0$

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

3.) For arbitrary positive numbers σ_1, σ_2

$$\Phi\left(\frac{x}{\sigma_1}\right) * \Phi\left(\frac{x}{\sigma_2}\right) = \Phi\left(\frac{x}{\sigma}\right)$$

where $\sigma = \sigma_1^2 + \sigma_2^2$,

then $\Phi(x) = \frac{1}{2\pi} \int_{-\infty}^{x} e^{-t^2/2}\, dt$.

(Here $F(x) * G(x)$ stands for the convolution $H(x) = \int_{\infty}^{\infty} F(x-t)\, dG(t)$).

Going through Trotter's proof we have found that actually a stronger theorem was proved than the one stated, namely that the convergence to the standard normal distribution function was not just a pointwise convergence (as the author stated) , but it was uniform on the line. This fact makes the operator's method a more efficient way of proving the C.L.T., since it proves a stronger result, still this method has its limitations and for more general limit theorems we cannot do away with important tools, like the method of characteristic functions.

The idea of the operator's method is, that we investigate the following convolution type integrals:

$$\int u(t-x)\, dF(x)$$

for bounded measurable functions u and distribution functions F.

It would be sufficient to consider the above convolution for the function:

$$u_0(x) = 0 \qquad \text{if } x \leqq 0$$

$$u_0(x) = 1 \qquad \text{if } x > 0.$$

Since $(u_0)F$ equals F in every continuity point t of the distribution function F, so this single convolution determines F. However we will work with a family of functions u, which can be handled much better analytically and then - roughly speaking - approximate u_0 with such functions, namely this is going to be the family of functions (called C_2 from now on) for which u, u' and u'' are uniformly continuous and bounded.

Definition.

For a distribution function F we define an operator T mapping C_2 to C_2 as follows:

(1) $(u)\, T = \int u(t-x)\, dF(x)$.

The operator $T: u \rightarrow (u)T$ is obviously a linear operator on C_2.

(2) $T_1 T_2$ denotes the product of the operators T_1 and T_2, i.e. the operator mapping the function u to the function $((u)T_1)T_2$.

We can extend this notation to several factors $T_1\, T_2 \ldots T_n$ in an obvious way.

Note that if T_1 and T_2 are the operators corresponding to the distribution functions F_1 and F_2, respectively, then $T_1 T_2$ is the operator corresponding to the convolution of F_1 and F_2 and hence:

$$T_1 T_2 = T_2 T_1$$

Note also that:

(3) $\|u\| = \sup_x |u(x)|$ defines a norm in C_2.

In the following we would like to show that the family of functions C_2 is sufficiently large for our purpose as it will be seen from Lemma 1:

Lemma 1.

Let F_n and F be distribution functions, T_n and T the corresponding operators and suppose that for each $u \in C_2$:

(4) $$\lim_{n\to\infty} \|(u)T_n - (u)T\| = 0$$

then

(5) $$\lim_{n\to\infty} F_n(t) - F(t) \quad \text{in every continuity point } t \text{ of } F.$$

If in addition the distribution $F(t)$ is uniformly continuous then:

(6) $$\lim_{n\to\infty} F_n(t) = F(t) \quad \text{uniformly for } -\infty < t < \infty.$$

Proof:

Let $\epsilon > 0$ arbitrary. Choose a function $u_1(y)$ for which

(7) $$\begin{aligned} u_1(y) &= 0 && \text{for } y \leqq -\epsilon \\ 0 \leqq u_1(y) &\leqq 1 && \text{for } -\epsilon < y \leqq 0 \\ u_1(y) &= 1 && \text{for } y > 0 \end{aligned}$$

and $u_1(y)$ belongs to C_2. (This can be achieved by polynomial approximation). By assumption:

$$\lim_{n\to\infty} \|(u_1)T_n - (u_1)T\| = 0,$$

i.e. for arbitrary $\epsilon > 0$, there exists $N_0(\epsilon)$ such that

(8) $$\sup_{-\infty<t<\infty} \left|\int u_1(t-x)\, dF_n(x) - \int u_1(t-x)\, dF(x)\right| < \epsilon \qquad \text{for } n > N_0(\epsilon).$$

By the definition (7) of u_1:

(9) $$\int u_1(t-x)\, dF_n(x) \geqq \int_{-\infty}^{t} dF_n(x) = F_n(t)$$

and by (8) for $n > N_0(\epsilon)$:

(10) $\int u_1(t-x)\, dF_n(x) \leqq \int u_1(t-x)\, dF(x) + \epsilon \leqq \int_{-\infty}^{t+\epsilon} dF(x) + \epsilon = F(t+\epsilon) + \epsilon.$

Combining (9) and (10) we get:

(11) $F_n(t) \leqq F(t+\epsilon) + \epsilon$ for $n > N_0(\epsilon)$ and for all t.

By a similar argument applied to $u_2(y) = u_1(y-\epsilon)$ we obtain:

(12) $F\ (t-\epsilon) - \epsilon \leqq F_n(t)$ for $n > N_0(\epsilon)$ and for all t.

(11) and (12) together imply that

(13) $F(t-\epsilon) - \epsilon \leqq F_n(t) \leqq F(t+\epsilon) + \epsilon$ for $n > N_0(\epsilon)$ and for all t.

If t is a continuity point of F(t) then letting ϵ tend to zero we get:

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

Suppose now that F(t) is uniformly continuous. Then for any $\eta > 0$, there exists $\epsilon > 0$ such that:

$$0 \leqq \epsilon + (F(t+\epsilon) - F(t)) < \eta \quad \text{and}$$

$$0 \leqq \epsilon + (F(t) - F(t-\epsilon)) < \eta \quad \text{for all t.}$$

Define $N_1(\eta) = N_0(\epsilon)$, then for $n > N_1(\eta)$ we can apply (13) (Note that $N_0(\epsilon)$ doesn't depend on t). But then:

$$F_n(t) - F(t) = (F_n(t) - F(t+\epsilon)) + (F(t+\epsilon) - F(t)) \leqq \epsilon + (F(t+\epsilon) - F(t)) < \eta$$

and

$$F_n(t) - F(t) = (F_n(t) - F(t-\epsilon)) + (F(t-\epsilon) - F(t)) \geqq -\epsilon - (F(t) - F(t-\epsilon)) > -\eta.$$

The last two chains of inequalities imply

$|F_n(t) - F(t)| < \eta$ for $n > N_1(\eta)$ and for all t, i.e. $F_n(t)$ converges to $F(t)$ uniformly in t.

This completes the proof of Lemma 1.

From the way we have defined the operator T corresponding to a distribution function, it is clear that T is a contraction operator, i.e.:

$$\|(u)T\| \leqq \|u\| \quad \text{for all } u \in C_2.$$

Hence we can prove the following:

Lemma 2.

Let $T_1, \ldots, T_n$ and $S_1, \ldots, S_n$ be arbitrary operators (corresponding to distribution functions) and let u belong to C_2. Then

$$(14) \quad \|(u)T_1 T_2 \ldots T_n - (u)S_1 S_2 \ldots S_n\| \leqq \sum_{i=1}^{n} \|(u)T_i - (u)S_i\|.$$

In particular

$$(15) \quad \|(u)T^n - (u)S^n\| \leqq n\|(u)T - (u)S\|.$$

The proof follows from the identity

$$(u)T_1 \ldots T_n - (u)S_1 \ldots S_n = \sum_{i=1}^{n} (u)T_1 T_2 \ldots T_{i-1}(T_i - S_i)S_{i+1} \ldots S_n,$$

the triangular inequality for norms and the inequality:

$$(16) \quad \|(u)T_1 \ldots T_{i-1}(T_i - S_i)S_{i+1} \ldots S_n\| \leqq \|(u)\ (T_i - S_i)\|$$

(where $((u)\ (T-S))$ means $(u)T - (u)S$). This latter inequality is a consequence of the fact that T_k, S_k are contraction operators.

After this introduction we are ready to prove the C.L.T.. Since there is no essential difference between the identically distributed case and the general (Lindeberg) case while proving by operator's method, we will present the proof of the Lindeberg theorem. As it was pointed out before, we will prove a stronger theorem which can be stated as follows:

Theorem 1.

Let $\{X_{nj}\}$ a double array of r.v.'s $(n=1,2,\ldots; j=1,2,\ldots,N_n)$ with independence in each row, the variables satisfying:

(17) $E X_{nj} = 0$ for all n and j

(18) $\sum_{j=1}^{N_n} X_{nj}^2 = 1$ for all n

(19) for $\eta > 0$ $\sum_{j=1}^{N_n} \int_{|x|>\eta} x^2 \, dF_{nj}(x) \to 0$ as $n\to\infty$,

then the distribution function of $\sum_{j=1}^{N_n} X_{nj}$ tends to the standard normal distribution function uniformly on the line.

Proof:

Let σ_{nj}^2 denote the variance of X_{nj}.

Let T_{nj} and T_n denote the operators associated with the distribution functions F_{nj} and F_n of the variables X_{nj} and $\sum_{j=1}^{N_n} X_{nj}$, respectively.

Let S_{nj} and S be the operators corresponding to the normal distributions $N(0,\sigma_{nj}) = \Phi_{nj}$ and $N(0,1) = \Phi$ respectively.

Obviously $T_n = T_{n1} T_{n2} \ldots T_{nN_n}$ and $S = S_{n1} S_{n2} \ldots S_{nN_n}$ for $n=1,2,\ldots$.

By Lemma 1 it is enough to prove that

(20) $\lim_{n\to\infty} \|(u)T_n - (u)S\| = 0$ for all u in C_2,

the same lemma also ensures that the convergence of the distribution functions is uniform on the line since the standard normal distribution function is uniformly continuous.

Fix $\epsilon > 0$. By Lemma 2:

$$\|(u)T_n - (u)S\| = \|(u)T_{n1}T_{n2}\cdots T_{nN_n} - (u)S_{n1}S_{n2}\cdots S_{nN_n}\| \leq$$

$$\leq \sum_{j=1}^{N_n} \|(u)\, T_{nj} - (u)S_{nj}\|.$$

For estimating the norm:

$$\|(u)T_{nj} - (u)S_{nj}\|,$$

consider arbitrary two distribution functions $F(x), G(x)$ with common expectation 0 and variance σ^2 and corresponding operators V and W respectively. Then:

$$(u)V = \int u(t-x)\, dF(x) =$$

$$= \int [u(t) - xu'(t) + \frac{x^2}{2}u''(t) + \frac{x^2}{2}(u''(t-\theta x) - u''(t)]\, dF(x) =$$

$$= u(t) + \frac{\sigma^2}{2}\, u''(t) + \int \frac{x^2}{2}\, (u''(t-\theta x) - u''(t))\, dF(x),$$

where $\theta = \theta(x) \in (0, 1)$.

Similarly

$$(u)W = u(t) + \frac{\sigma^2}{2}\, u''(t) + \int \frac{x^2}{2} (u''(t-\theta x) - u''(t))\, dG(x).$$

By subtraction we obtain:

$$(u)V - (u)W = \int \frac{x^2}{2}\, (u''(t-\theta x) - u''(t))\, dF(x) -$$

$$- \int \frac{x^2}{2}\, (u''(t-\theta x) - u''(t))\, dG(x).$$

Since $u''(t)$ is uniformly continuous there exists a $\delta > 0$ such that

$|u''(t-\theta x) - u''(t)| < \frac{\epsilon}{2}$ for $|x| \leq \delta$, whatever t and $0 < \theta < 1$ are.

Also using the boundedness of u"(t) we set

$$|(u)V - (u)W| \leqq \frac{\epsilon}{2} \int_{|x| \leqq \delta} \frac{x^2}{2} dF(x) + \frac{\epsilon}{2} \int_{|x| \leqq \delta} \frac{x^2}{2} dG(x) +$$

$$+ 2K \int_{|x| > \delta} \frac{x^2}{2} dF(x) + 2K \int_{|x| > \delta} \frac{x^2}{2} dG(x) \leqq$$

$$\leqq \frac{\epsilon}{4} \sigma^2 + \frac{\epsilon}{4} \sigma^2 + K \int_{|x| > \delta} x^2 dF(x) + K \int_{|x| > \delta} x^2 dG(x).$$

Since the right hand side does not depend on t:

$$(21) \quad \|(u)V - (u)W\| \leqq \frac{\epsilon}{2} \sigma^2 + K \int_{|x| > \delta} x^2 dF(x) + K \int_{|x| > \delta} x^2 dG(x).$$

Applying inequality (21) to the operators T_{nj} and S_{nj} we obtain:

$$(22) \quad \|(u)T_{nj} - (u)S_{nj}\| \leqq \frac{\epsilon}{2} \sigma_{nj}^2 + K \int_{|x| > \delta} x^2 dF_{nj}(x) + K \int_{|x| > \delta} x^2 d\Phi_{nj}(x)$$

and hence by Lemma 2 and (22)

$$\|(u)T_n - (u)S\| \leqq \frac{\epsilon}{2} \sum_{j=1}^{N_n} \sigma_{nj}^2 + K \sum_{j=1}^{N_n} \int_{|x| > \delta} x^2 dF_{nj}(x) + K \sum_{j=1}^{N_n} \int_{|x| > \delta} x^2 d\Phi_{nj}(x) \leqq$$

$$\leqq \frac{\epsilon}{2} + \frac{\epsilon}{4} + K \sum_{j=1}^{N_n} \int_{|x| > \delta} x^2 d\Phi_{nj}(x)$$

for n large enough by (18) and (19).

We can complete the proof by showing that the last term tends to 0.

$$\sum_{j=1}^{N_n} \int_{|x|>\delta} x^2 d\,\Phi_{nj}(x) = \sum_{j=1}^{N_n} \int_{|x|>\delta} x^2 d\,\Phi\left(\frac{x}{\sigma_{nj}}\right) =$$

$$= \sum_{j=1}^{N_n} \int_{|y|>\frac{\delta}{\sigma_{nj}}} y^2 \cdot \sigma_{nj}^2 \, d\,\Phi(y) \leqq \sum_{j=1}^{n} \sigma_{nj}^2 \int_{|y|>\frac{\delta}{\max_j \sigma_{nj}}} y^2 d\,\Phi(y).$$

Since by Lemma 3 of Chapter I $\max_j \sigma_{nj} \to 0$ as $n \to \infty$ and since $\int y^2 d\,\Phi(y)$ is finite, the integral $\int_{|y|>\frac{\delta}{\max_j \sigma_{nj}}} y^2 d\,\Phi(y)$ tends to 0 as $n \to \infty$ by the absolute continuity of the integral.

Thus we have proved that the norm

$\|(u)T_n - (u)S\|$ tends to 0 as $n \to \infty$ for every u in C_2,

which by Lemma 1 implies the conclusion of Theorem 1.

Chapter III

The Method of Characteristic Functions in Proving the Central Limit Theorem

3.0. Historical Background in the Usage of Characteristic Functions for C.L.T.'s.

The method of characteristic functions (or Fourier transforms) in proving the C.L.T. was first used by Lyapunov [25] in 1901. This method is based on a very important result, namely the continuity theorem of characteristic functions (see p.171 Breiman [6].):

Theorem 1.

Given a sequence of distribution functions $F_n(x)$ with characteristic functions $\varphi_n(t)$ (i.e. $\varphi_n(t) = \int_{-\infty}^{\infty} e^{itx}\, dF_n(x)$). The characteristic functions converge to a characteristic function if and only if the corresponding distribution functions tend to the corresponding distribution function in every continuity point of the latter.

In our case the limiting distribution function is the standard normal distribution function $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2}\, du$ (which is everywhere continuous) whose characteristic function is $e^{-t^2/2}$.

In other words, instead of proving convergence of distributions, it is enough to prove pointwise convergence of the characteristic functions to the function $e^{-t^2/2}$.

By using this tool Lyapunov was able to prove a much more general form of the C.L.T. than anyone before. Let's state Lyapunov's version of the theorem (without proof, since it is a corollary of Theorem 5 of Chapter III):

Theorem 2

Given a sequence of independent random variables $\{X_n\}$ with the first three central moments $EX_j = m_j$, $\sigma^2(X_j) = \sigma_j^2$, and $E|X_j - m_j|^3 = \gamma_j^3$ existing (j=1, 2, . . .), then setting

$$s_n = \sqrt{\sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2},$$

$$h_n = \sqrt[3]{\gamma_1^3 + \gamma_2^3 + \ldots + \gamma_n^3},$$

$$Y_n = \sum_{j=1}^{n} X_j \quad \text{and} \quad Z_n = \frac{Y_n - EY_n}{\sigma(Y_n)},$$

if the Lyapunov condition

$$(1) \qquad \lim_{n\to\infty} \frac{h_n}{s_n} = 0$$

is satisfied then

$$\lim_{n\to\infty} F_n(x) = \Phi(x) \qquad (-\infty < x < \infty),$$

where $F_n(x)$ denotes the distribution function of Z_n.

Lyapunov proved the C.L.T. under an even more general condition, namely, that instead of the existence of the third moments γ_j^3 it was enough to require the existence of the $(2+\epsilon)^{th}$ moment ($\epsilon > 0$) and the condition

(2) $$\lim_{n\to\infty} \frac{h_n(2+\epsilon)}{s_n} = 0 \quad \text{where} \quad h_n(p) = \Big(\sum_{j=1}^{n} E|X_j - m_j|^p\Big)^{\frac{1}{p}} .$$

The method of characteristic functions was the tool first used by Kolmogorov [15] in 1932 and P. Lévy [21] in 1935, for general limit theorems (the limit distribution was not necessarily the normal one). A good survey of limit theorems for sums of independent r.v.'s can be found in the book of Gnedenko and Kolmogorov [11].

In this chapter we are going to present proofs for the C.L.T. under different conditions using the method of characteristic functions. We start with the simplest case:

3.1. The Case of Independent, Identically Distributed Random Variables.

Theorem 3

Given a sequence of independent identically distributed r.v.'s $\{X_n\}$ with mean 0 and variance 1. Then

$$\lim_{n\to\infty} P\Big(\frac{S_n}{\sqrt{n}} = \frac{\sum_{j=1}^{n} X_j}{\sqrt{n}} < x\Big) = \Phi(x).$$

Proof.

By Theorem 1 of this chapter it is enough to prove that

$$E\, e^{it\frac{S_n}{\sqrt{n}}} \quad \text{tends to} \quad e^{-t^2/2}.$$

Let's get an estimate for the characteristic function of $S_n/\sqrt{n}$.

$$E\, e^{it\frac{S_n}{\sqrt{n}}} = \left(E\, e^{it\frac{X_1}{\sqrt{n}}}\right)^n = \left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right)^n$$

using the independence and the assumptions, $EX_1=0$, $EX_1^2=1$ and the wellknown expansion

$$\varphi(y) = 1 + iyEX_1 + \frac{(iy)^2}{2} \cdot EX_1^2 + o(y^2)$$

of characteristic functions for $t \to 0$ (see Loève [24], p. 199).

Therefore $E\, e^{it\frac{S_n}{\sqrt{n}}}$ tends to $e^{-t^2/2}$ as $n \to \infty$.

Obviously the variables of the de Moivre-Laplace theorem satisfy the conditions of Theorem 3 (the common distribution being Bernoullian), i.e. we've shown a very short proof of the de Moivre-Laplace theorem (as we indicated in Chapter I).

3.2. The Non-Identically Distributed Case.

In this section we would like to concentrate our attention to some recent results by P.Révész [28] 1965 and J.Komlós [17] 1970 who were using the method of characteristic functions in proving C.L.T.'s.

Actually they replaced the independence of the random variables by a weaker condition, the so called strongly multiplicative systems, but in this chapter we present these proofs for independent variables only and we are going to show the original proofs for the dependent case in Chapter IV.

The idea of both proofs is that the C.L.T. is reduced to the law of large numbers, but while the Révész proof requires the condition of uniform boundedness, Komlós replaces this rather strong requirement by the usual Lindeberg condition (62) of Chapter I.

We feel it is worthwhile to recall the Révész proof since it is quite short and it serves as a basis for the proof of the Komlós theorem.

Theorem 4

Given a sequence of independent uniformly bounded r.v.'s $\{X_n\}$ with $EX_j=0$ and $EX_j^2 = 1$ for all j, then the distribution function of the normalized sum

$$\frac{S_n}{\sqrt{n}} = \frac{X_1 + X_2 + \ldots + X_n}{\sqrt{n}}$$

tends to the standard normal distribution function.

Proof.

By virtue of Theorem 1 all we need to show is that

$$(3) \qquad E\, e^{it\frac{S_n}{\sqrt{n}}} \to e^{-t^2/2} \quad \text{for every given t.}$$

We use the following simple expansion:

$$(4) \qquad e^{is} = (1 + is)\, e^{-s^2/2} + r(s),$$

where $|r(s)| \leq |s|^3$ for all real s.

In view of expansion (4) we can write:

$$(5) \qquad e^{it\frac{S_n}{\sqrt{n}}} = \prod_{j=1}^{n} e^{it\frac{X_j}{\sqrt{n}}} = \prod_{j=1}^{n} (1 + it\frac{X_j}{\sqrt{n}})\, e^{-t^2/2 \sum_j \frac{X_j^2}{n}} + \sum_j r(t\frac{X_j}{\sqrt{n}}) =$$

$$= e^{-t^2/2} \prod_{j=1}^{n} (1+ it\frac{X_j}{\sqrt{n}}) + \prod_{j=1}^{n} (1+ it\frac{X_j}{\sqrt{n}})[e^{\frac{-t^2}{2} \sum_{j=1}^{n} \frac{X_j^2}{n} + \sum_{j=1}^{n} r(t\frac{X_j}{\sqrt{n}})} - e^{\frac{-t^2}{2}}].$$

Denote

$$A_n = \prod_{j=1}^{n} (1+ it \frac{X_j}{\sqrt{n}})$$

$$B_n = e^{-\frac{t^2}{2} \sum_{j=1}^{n} \frac{X_j^2}{n} + \sum_{j=1}^{n} r(t \frac{X_j}{\sqrt{n}})} - e^{-\frac{t^2}{2}}$$

$$C_n = A_n B_n,$$

then

$$(6) \qquad E e^{it \frac{S_n}{\sqrt{n}}} = e^{-\frac{t^2}{2}} E \prod_{j=1}^{n} (1+ it \frac{X_j}{\sqrt{n}}) + EC_n.$$

But by the independence of the r.v.'s X_j

$$(7) \qquad E \prod_{j=1}^{n} (1+ it\frac{X_j}{\sqrt{n}}) = \prod_{j=1}^{n} E(1+ it \frac{X_j}{\sqrt{n}}) = 1$$

since $EX_j=0$ for all j.

Therefore

$$E e^{it \frac{S_n}{\sqrt{n}}} = e^{-\frac{t^2}{2}} + EC_n.$$

We want to show that $EC_n \to 0$.

We will do it in 3 steps.

1) Show that $B_n \to 0$ a.e.

2) Show that A_n is uniformly bounded, this together with 1) implies that $C_n \to 0$ a.e.

3) Show that C_n is uniformly bounded, therefore $C_n \to 0$ a.e. implies that $EC_n \to 0$ by the Dominated Convergence Theorem.

1) To prove that $B_n \to 0$ a.e., it is enough to show that

$$(8) \quad -\frac{t^2}{2} \sum_{j=1}^{n} \frac{X_j^2}{n} + \sum_{j=1}^{n} r(t \frac{X_j}{n}) \longrightarrow -\frac{t^2}{2} \text{ a.e.}$$

Since the variables X_j^2 are independent, bounded with $EX_j^2=1$, by the strong law of large numbers (see Chung [7], p.97):

$$(9) \quad \sum_{j=1}^{n} \frac{X_j^2}{n} \longrightarrow 1 \text{ a.e.} \quad \text{as } n\to\infty.$$

Furthermore

$$(10) \quad \left|\sum_{j=1}^{n} r(t \frac{X_j}{\sqrt{n}})\right| \leq \sum_{j=1}^{n} |t|^3 \frac{|X_j|^3}{\sqrt{n}^3} \leq \cancel{n} \cdot |t|^3 \frac{K^3}{\cancel{n}\sqrt{n}} \to 0 \text{ a.e.},$$

where K is the common bound for the variables $|X_j|$.

(9) and (10) together imply (8).

2) To show the boundedness of A_n consider

$$|A_n|^2 = \prod_{j=1}^{n} (1+t^2 \frac{X_j^2}{n}) \leq (1+t^2 \frac{K^2}{n})^n \leq e^{t^2 K^2},$$

indicating clearly that A_n is bounded.

3) Since $e^{it \frac{S_n}{\sqrt{n}}}$ is uniformly bounded and so is $e^{-t^2/2}$. A_n therefore their difference C_n is uniformly bounded proving 3).

Theorem 5

(For the statement of the theorem see Theorem 3 of Chapter I).

The structure of the proof basically agrees with the previous one, but the strong law of large numbers will be replaced by a weak law (Theorem 5A) and the Dominated Convergence Theorem by a more general theorem of the same type (Lemma 1). We will start by first proving these lemma and theorem.

Lemma 1

If $\{\xi_n\}$ is a uniformly integrable sequence of r.v.'s tending to zero in measure then $E|\xi_n| \to 0$, also $E\xi_n \to 0$.

Definition

A sequence $\{f_n\}$ of real or complex-valued functions is said to be uniformly integrable if for any $\epsilon > 0$, there exists an $A > 0$ such that for all n $\int_{|f_n|>A} |f_n|\, dP < \epsilon$.

Proof of Lemma 1

Let $\epsilon > 0$. Choose A so large that for all n $\int_{|\xi_n|>A} |\xi_n|\, dP \leq \frac{\epsilon}{3}$ (uniform integrability). Whence

$$\int |\xi_n| dP = \int_{|\xi_n|>A} |\xi_n|\, dP + \int_{\frac{\epsilon}{3}<|\xi_n|\leq A} |\xi_n|\, dP + \int_{|\xi_n|\leq \frac{\epsilon}{3}} |\xi_n|\, dP$$

$$\leq \frac{\epsilon}{3} + A \cdot P(\frac{\epsilon}{3} < |\xi_n|) + \frac{\epsilon}{3} .$$

Since $\xi_n \to 0$ in measure,

$$P(|\xi_n| > \frac{\epsilon}{3}) < \frac{\epsilon}{3A} \quad \text{for} \quad n \geqq N_0,$$

which implies that

$$\int |\xi_n| dP < \epsilon \qquad \text{for} \quad n \geqq N_0. \qquad \text{Q.E.D.}$$

The following lemma gives a sufficient condition for uniform integrability.

Lemma 2

If there exists an $\alpha > 0$ such that

$$(11) \quad \int |\xi_n|^{1+\alpha} dP < K \qquad n=1,2,\ldots$$

then $\{\xi_n\}$ is uniformly integrable.

Proof

Fix $\epsilon > 0$. Let A be such that

$$\frac{K}{A^\alpha} < \epsilon,$$

then

$$K > \int |\xi_n|^{1+\alpha} dP \geqq \int_{|\xi_n| > A} |\xi_n|^{1+\alpha} \geqq A^\alpha \int_{|\xi_n| > A} |\xi_n| dP.$$

Therefore,

$$\int_{|\xi_n| > A} |\xi_n| \, dP < \frac{K}{A^\alpha} < \epsilon \qquad n=1,2,\ldots \qquad \text{Q.E.D.}$$

Remark

In particular $\int |\xi_n|^2 dP < K$ implies that $|\xi_n|$ is uniformly integrable.

In the proof of Theorem 5 we will need (as has been pointed out before) a weak law of large numbers for double arrays.

Theorem 5A

Given a double array of random variables $\{\xi_{nj}\}$ $(n=1,2,\ldots;$ $j=1,2,\ldots N_n)$ pairwise independent for any given n.

If the following conditions hold:

$$(12)\quad \sum_{j=1}^{N_n} \int_{|\xi_{nj}|>1} |\xi_{nj}|\,dP \to 0$$

$$(13)\quad \sum_{j=1}^{N_n} \int_{|\xi_{nj}|\leq 1} |\xi_{nj}|^2\,dP \to 0,$$

then

$$(14)\quad \xi_n - E\xi_n \to 0 \text{ in probability, where } \xi_n = \sum_{j=1}^{N_n} \xi_{nj}.$$

Proof

Define

$$\xi_{nj}^{*} = \begin{cases} \xi_{nj} & \text{if } |\xi_{nj}| \leq 1 \\ 0 & \text{otherwise,} \end{cases}$$

$$\xi_n^{*} = \sum_{j=1}^{N_n} \xi_{nj}^{*} .$$

Then $P(\xi_n \neq \xi_n^{*}) \to 0$ as $n \to \infty$ by (12).

Also

$$(15)\quad E(\xi_n^* - E\xi_n^*)^2 = \sum_{j=1}^{N_n} \sigma^2(\xi_{nj}^*) \leqq \sum_{j=1}^{N_n} E(\xi_{nj}^*)^2 =$$

$$= \sum_{j=1}^{N_n} \int_{|\xi_{nj}| \leqq 1} \xi_{nj}^2 \, dP \to 0 \quad \text{by (13)},$$

i.e.

$$\xi_n^* - E\xi_n^* \to 0 \quad \text{in } L_2 \text{ implies that}$$

$$\xi_n^* - E\xi_n^* \to 0 \quad \text{in probability.}$$

We want to show that this implies

$$(16)\quad \xi_n - E\xi_n^* \to 0 \quad \text{in probability.}$$

This is true since

$$(17)\quad P(|\xi_n - E\xi_n^*| > \epsilon) = P(|\xi_n - \xi_n^* + \xi_n^* - E\xi_n^*| > \epsilon) \leqq$$

$$\leqq P(|\xi - \xi_n^*| > \tfrac{\epsilon}{2}) + P(|\xi_n^* - E\xi_n^*| > \tfrac{\epsilon}{2}) \leqq$$

$$\leqq P(\xi_n \neq \xi_n^*) + P(|\xi_n^* - E\xi_n^*| > \tfrac{\epsilon}{2}) \to 0.$$

Remains to show that

$$(18)\quad \xi_n - E\xi_n \to 0 \text{ in probability.}$$

It is sufficient to prove that $E\xi_n - E\xi_n^* \to 0$, since

$$(19)\quad \xi_n - E\xi_n = (\xi - E\xi_n^*) + (E\xi_n^* - E\xi_n).$$

$$(20)\quad |E\xi_n - E\xi_n^*| \leq \sum_{j=1}^{N_n} \left|\int \xi_{nj}\,dP - \int_{|\xi_{nj}|\leq 1} \xi_{nj}\,dP\right| =$$

$$= \sum_{j=1}^{N_n} \int_{|\xi_{nj}|>1} |\xi_{nj}|\,dP \to 0 \quad \text{by (12)},$$

which by (17) and (19) implies the conclusion of Theorem 5A.

Corollary to Theorem 5A.

If for the double array of r.v.'s $\{X_{nj}\}$ $(n=1,2,\ldots;\ j=1,2,\ldots N_n)$ where $\sum_{j=1}^{N_n} EX_{nj}^2 = 1$ the Lindeberg condition (62) of Chapter I holds then:

$$\sum_{j=1}^{N_n} X_{nj}^2 - \sum_{j=1}^{N_n} EX_{nj}^2 \to 0 \quad \text{in probability or}$$

$$\sum_{j=1}^{N_n} X_{nj}^2 \to 1 \quad \text{in probability.}$$

Proof

It is enough to show that the two conditions of Theorem 5A hold for the r.v.'s X_{nj}^2.

But $\sum_{j=1}^{N_n} \int_{|X_{nj}^2|>1} X_{nj}^2\,dP \to 0$ by the Lindeberg condition, i.e. (12) holds.

Also, let $\epsilon > 0$.

$$\sum_{j=1}^{N_n} \int_{|X_{nj}^2|\leq 1} X_{nj}^4\,dP = \sum_{j=1}^{N_n} \int_{|X_{nj}|\leq \eta} X_{nj}^4\,dP + \sum_{j=1}^{N_n} \int_{1\geq |X_{nj}|>\eta} X_{nj}^4\,dP \leq$$

$$(21) \qquad \leq \eta^2 \sum_{j=1}^{N_n} \int_{|X_{nj}|\leq\eta} X_{nj}^2 \, dP + \sum_{j=1}^{N_n} \int_{|X_{nj}|>\eta} X_{nj}^2 \, dP.$$

Since $\sum_{j=1}^{N_n} EX_{nj}^2 = 1$, the left term in (21) is less than $\frac{\epsilon}{2}$ if η is chosen so that $\eta^2 < \frac{\epsilon}{2}$, the right term is also less than $\frac{\epsilon}{2}$ if n is large enough by the Lindeberg condition, i.e. condition (13) is satisfied as well, therefore

$\sum_{j=1}^{N_n} X_{nj}^2 \to 1$ in probability. Q.E.D.

Now we have all the necessary tools for proving Theorem 5.

Proof of Theorem 5

The proof will be quite similar to that of Theorem 4. We will use the same expansion (4) as in Theorem 4.

Therefore

$$(22) \quad e^{itS_n} = \prod_{j=1}^{N_n} (1+itX_{nj})\, e^{-\frac{t^2}{2}\sum_{j=1}^{N_n} X_{nj}^2 + \sum_{j=1}^{N_n} r(tX_{nj})} =$$

$$= e^{-\frac{t^2}{2}} \prod_{j=1}^{N_n} (1+itX_{nj}) +$$

$$(23) \qquad + \prod_{j=1}^{N_n} (1+itX_{nj}) [e^{-\frac{t^2}{2}\sum_{j=1}^{N_n} X_{nj}^2 + \sum_{j=1}^{N_n} r(tX_{nj})} - e^{-\frac{t^2}{2}}]$$

Let's call the first factor of this very last expression (23) A_n, the second factor B_n and the product $A_n \cdot B_n$ we call C_n.

Here

(24) $$Ee^{itS_n} = e^{-\frac{t^2}{2}} + EC_n$$

for the same reasons as in steps (6) and (7) in Theorem 4.

By virtue of the corollary to Theorem 5A we know that

(25) $$\sum_{j=1}^{N_n} X_{nj}^2 \to 1 \quad \text{in probability.}$$

To establish 1) of Theorem 4 ($B_n \to 0$ in probability) it remains to show that $\Sigma r(tX_{nj}) \to 0$ in probability.

But

(26) $$|\Sigma r(tX_{nj})| \leq \Sigma |t|^3 |X_{nj}|^3$$

$$\leq |t|^3 \max_{j=1}^{N_n} |X_{nj}| \cdot \sum_{j=1}^{N_n} X_{nj}^2 .$$

Knowing (25) it is enough to show that

(27) $$\max_{j=1}^{N_n} |X_{nj}| \to 0 \quad \text{in probability.}$$

(28) $$P(\max_j |X_{nj}| > \epsilon) = P(\text{one of } |X_{nj}| > \epsilon) \leq \sum_{j=1}^{N_n} P(|X_{nj}| > \epsilon) =$$

$$= \sum_{j=1}^{N_n} \int_{|X_{nj}| > \epsilon} dP \leq$$

$$\leq \frac{1}{\epsilon^2} \sum_{j=1}^{N_n} \int_{|X_{nj}| > \epsilon} X_{nj}^2 \, dP.$$

Since ϵ is fixed this latter expression tends to 0 again by the Lindeberg condition, proving (27).

Finally, we want to show that C_n is uniformly integrable first by showing this for A_n.

$$(29) \quad E|A_n|^2 = E\,\Pi(1+t^2X_{nj}^2) \leqq e^{\sum_j t^2 EX_{nj}^2} = e^{t^2},$$

indicating that A_n is uniformly integrable (see remark to Lemma 2), obviously $e^{-t^2/2}\cdot A_n$ is uniformly integrable as well and since e^{itS_n} is bounded, consequently uniformly integrable therefore C_n (the difference of these latter two expressions) must be uniformly integrable too.

Before we can use Lemma 1 to (24) we need to show that $C_n \to 0$ in probability.

But

$$|A_n|^2 = \prod_{j=1}^{N_n} (1+t^2X_{nj}^2) \leqq e^{t^2 \Sigma X_{nj}^2} \to e^{t^2} \text{ in probability,}$$

implying that A_n is bounded in probability (by which we mean that $[P(|A_n| > e^{t^2}+1)] < \epsilon$ for n large enough). Since $B_n \to 0$ in probability we obtain $C_n \to 0$ in probability.

This together with uniform integrability implies that $EC_n \to 0$, applying Lemma 1 to C_n. From (24) it then follows that:

$$E\, e^{itS_n} \to e^{\frac{-t^2}{2}} \text{ as } n \to \infty. \qquad \text{Q.E.D.}$$

At this stage it should be pointed out that the Lindeberg condition is not only a sufficient condition, but under the assumption that the double array is 'infinitesimal' namely,

$$(30) \quad \lim_{n\to\infty} \max_{j=1}^{N_n} P(|X_{nj}| > \epsilon) = 0$$

it is also a necessary condition for the C.L.T. as it was first shown by Feller in 1937 through a novel application of characteristic functions.

This condition (30) clearly follows from the Lindeberg condition since,

$$\max_{j=1}^{N_n} P(|X_{nj}| > \epsilon) \leq P(|X_{nj}| > \epsilon) \text{ for at least one } j \leq$$

$$\leq \sum_{j=1}^{N_n} P(|X_{nj}| > \epsilon) =$$

$$= \sum_{j=1}^{N_n} \int_{|X_{nj}| > \epsilon} dP \leq$$

$$\leq \frac{1}{\epsilon^2} \sum_{j=1}^{N_n} \int_{|X_{nj}| > \epsilon} X_{nj}^2 \, dP \to 0$$

by the Lindeberg condition.

We will conclude this chapter by a corollary to Theorem 5. As we mentioned in section 3.0, Lyapunov's theorem (Theorem 2) follows from Theorem 5.

Corollary to Theorem 5

If the double array of Theorem 5 satisfies condition (2) instead of the Lindeberg condition, the C.L.T. holds true.

Proof

To prove this corollary it is enough to show that the Lindeberg condition (62) is a consequence of Lyapunov's condition.

Let $\epsilon > 0$. Fix $\delta > 0$. By Lyapunov's condition

$$\sum_{j=1}^{N_n} E(X_{nj})^{2+\epsilon} < \delta \quad \text{for} \quad n > N_0(\delta).$$

Therefore

$$\delta > \sum_{j=1}^{N_n} E(X_{nj})^{2+\epsilon} = \sum_{j=1}^{N_n} \int X_{nj}^{2+\epsilon} \, dP \geqq$$

$$\geqq \sum_{j=1}^{N_n} \eta^{\epsilon} \int_{|X_{nj}|>\eta} X_{nj}^{2} \, dP \quad \text{for} \quad n > N_0(\delta),$$

implying that the double array satisfies the Lindeberg condition and therefore by Theorem 5 the C.L.T. holds.

Chapter IV

Central Limit Theorems for Non-Independent Random Variables

4.1. Multiplicative Systems.

First we are going to prove the C.L.T. for so called strongly multiplicative systems by modifying the proofs of Theorem 4 and Theorem 5 of Chapter III.

Definitions.

A sequence of r.v.'s $\{X_n\}$ is called multiplicative if

(1) $E X_1 X_2 \dots X_n = E X_1 E X_2 \dots E X_n$.

If in addition to (1) the variables satisfy the condition

(2) $E X_1^2 X_2^2 \dots X_n^2 = E X_1^2 E X_2^2 \dots E X_n^2$

then it is called strongly multiplicative (see Alexits [1]).

Remark.

Obviously independent r.v.'s with finite expectation (variance) form a multiplicative (strongly multiplicative) system. We are going to give an example for multiplicative and strongly multiplicative systems which are not independent, in Example 1 of this chapter.

Condition (1) and (2) are not only weaker than independence, but since they are analytic conditions it is easier to check them.

The result of Révèsz mentioned in Chapter III can be stated in the following way.

Theorem 1.

The statement of Theorem 4 of Chapter III remains true if we replace independence by strong multiplicativeness.

Proof:

In the proof of Theorem 4 we used independence at two points, namely, when we showed that:

$$(3) \qquad E \prod_{j=1}^{n} \left(1+it\frac{X_j}{\sqrt{n}}\right) = \prod_{j=1}^{n} E\left(1+it\frac{X_j}{\sqrt{n}}\right)$$

and when we applied the strong law of large numbers for the variables X_j^2. To establish (3) it is sufficient to assume (1) instead of independence as it can be easily seen by extending the product.

For the strong law of large numbers we don't need independence - the uncorrelatedness of the variables X_j^2 ($EX_j^2X_k^2 = EX_j^2EX_k^2$) is sufficient, and this is just a special case of (2) for n=2.

Remark.

The above argument shows that in the uniform bounded case we do not need strongly multiplicativeness, it suffices to assume multiplicativeness and (2) for only n=2.

The result of Komlós mentioned in Chapter III can be stated the following way.

Theorem 2.

The statement of Theorem 5 in Chapter III remains true if we replace the independence by strongly multiplicativeness.

Proof

We will only point out where the proof of Theorem 5 needs to be changed under this new assumption.

First we show that in proving Theorem 5A the condition of pairwise independence can be replaced by uncorrelatedness under the additional assumptions: the variables ξ_{nj} are non-negative and the sequence $E\xi_n$ is bounded.

We used indepedence only in (15) of Chapter III. Under the new conditions we can write

$$E(\xi_n^* - E\xi_n^*)^2 = E(\xi_n^*)^2 - (E\xi_n^*)^2 =$$

$$= E(\sum_{j=1}^{N_n} \xi_{nj}^*)^2 - (E\,\xi_n^*)^2 =$$

$$= \sum_{j=1}^{N_n} E(\xi_{nj}^*)^2 + \sum_{\substack{k \neq \ell \\ 1 \leqq k, \ell \leqq N_n}} E\,\xi_{nk}^* \xi_{n\ell}^* - (E\xi_n^*)^2 \leqq$$

$$\leqq \sum_{j=1}^{N_n} E(\xi_{nj}^*)^2 + \sum_{k,\ell=1}^{N_n} E\,\xi_{nk}\xi_{n\ell} - (E\,\xi_n^*)^2 =$$

$$= \sum_{j=1}^{N_n} E(\xi_{nj}^*)^2 + (E\,\xi_n + E\,\xi_n^*)(E\,\xi_n - E\,\xi_n^*)$$

which tends to zero as we showed in the proof of Theorem 5A.

Clearly the sequence X_{nj}^2 of Theorem 2 satisfies the three new conditions (uncorrelatedness, non-negativeness and boundedness for the $E\sum_{j=1}^{N_n} X_{nj}^2$ -s.).

We have already mentioned that under condition (1) we have

$$E \prod_{j=1}^{N_n} (1+it X_{nj}) = \prod_{j=1}^{N_n} E(1+it X_{nj}),$$

and the same way we get the equality

$$E \prod (1+t^2 X_{nj}^2) = \prod E(1+t^2 X_{nj}^2)$$

which we used when estimating $E|A_n|^2$ in step (29).

The rest of the proof is exactly the same as in the independent case.

We will conclude our argument on strongly multiplicative systems by an example.

Example 1

An example for strongly multiplicative systems is a lacunary trigonometric sequence which explains to some extent, why strongly lacunary trigonometric series behave the same way as independent functions. (They obviously are not independent functions).

Consider the probability space $(\Omega, \mathcal{B}, \mu)$ where

$\Omega = [0, 2\pi]$

$\mathcal{B}$ = Borel sets of $[0, 2\pi]$

and if A is any set in $\mathcal{B}$ then

$\mu(A) = \frac{1}{2\pi} \lambda(A)$ (λ denoting the Lebesgue measure),

and a sequence $\cos n_k x$ on this measure space ($n_1 < n_2 < \ldots$ integers).

It is easy to see that

$$\cos n_1 x \cos n_2 x \ldots \cos n_k x = \frac{1}{2^k} \sum_{\pm} \cos(\pm n_1 \pm n_2 \ldots \pm n_k)x,$$

where the sum is extended over all the 2^k possible choices of the signs +, - therefore if $n_k > n_{k-1} + n_{k-2} + \ldots + n_1$ (which is certainly true if $\frac{n_i+1}{n_i} \geqq 2$) then the integral of this product is zero.

Under the same conditions we have

$$\int_0^{2\pi} \cos^2 n_1 x \cos^2 n_2 x \ldots \cos^2 n_k x \; d\mu =$$

$$= \int_0^{2\pi} \cos^2 n_1 x d\mu \int_0^{2\pi} \cos^2 n_2 x \, d\mu \ldots \int_0^{2\pi} \cos^2 n_k x \, d\mu = \frac{1}{2^k}$$

as it can be seen using the same argument.

Thus we have seen that the above lacunary series $\cos n_k x$ satisfies condition (1) and (2).

4.2. C.L.T.'s for m-Dependent Random Variables.

Finally we will mention without proof a few interesting results in proving C.L.T.'s for different types of dependent variables.

Definition

(see p.196, Chung [7])

If m is a non-negative integer, a sequence $\{X_n\}$ of random variables is called m-dependent if $X_1, X_2, \ldots, X_s$ is independent of $X_t, X_{t+1}, \ldots$ provided $t-s > m$.

C.L.T.'s for m-dependent variables were proved by Hoeffding and Robbins [12] in 1948, Dianada [8] in 1955 and by Orey [26] in 1958. In this latter paper a fairly general theorem is proved, we state only a particular case of this theorem.

Theorem 3

The Lindeberg theorem for double arrays (Theorem 3 of Chapter I) remains valid if the condition of independence is replaced by m-dependence, for the variables X_{nj}.

4.3. C.L.T.'s for Martingales.

Before stating our theorem we will need a few definitions.

Definition

(See p.118, Breiman [6]).

An event A is invariant if there exists $B \in \mathcal{B}_\infty$ such that for every $n \geq 1$,

$$A = \{(X_n, X_{n+1}, \ldots) \in B\}.$$

Definition

(See p.119, Breiman [6]).

If a process $X_1, X_2, \ldots$ has the following properties:

i) for every k, the distribution of $X_k, X_{k+1}, \ldots$ is the same as the distribution of $X_1, X_2, \ldots,$

ii) every invariant event has probability zero or one,

then this process is called a stationary, ergodic stochastic process.

Definition

(See Dobrushin [9]).

Consider a sequence of random variables $\{X_n\}$. If:

i) $E|X_j| < \infty$ for all j,

ii) $E(X_n | X_1, X_2, \ldots, X_{n-1}) = 0$ a.e.

then such a sequence of r.v.'s is called a sequence of martingale differences.

Now we are ready to state our theorem due to Ibragimov [13] 1963 and Billingsley [4] 1961. The two authors published their results independently and while Billingsley was following P. Lévy's [19], [20, Chapter 4] idea, Ibragimov based his work on the papers [2], [3] of Bernshtein as well as Lévy's.

Theorem 4

If the stationary, ergodic stochastic process $X_1, X_2, \ldots$ is a sequence of martingale differences with EX_1^2 finite, then the distribution of $\sum_{j=1}^{n} X_j/\sqrt{n}$ approaches the normal distribution with mean 0 and variance EX_1^2.

Remark

Obviously an independent sequence $\{X_n\}$ with $EX_n = 0$ is a sequence of martingale differences.

Bibliography

[1] Alexits, G. (1961), Convergence problems of orthogonal series, Publishing House of the Hung.Acad.Sci., Budapest.

[2] Bernshtein, S.N. (1941), On sums of dependent quantities having a joint regression which is almost zero, Dokl.Akad.Nauk SSSR, 32, 303-307 (in Russian).

[3] Bernshtein, S.N. (1940), New applications of almost independent quantities,Izv.Akad.Nauk SSSR., Ser.Mat., 4, 137-150 (in Russian).

[4] Billingsley, P. (1961), The Lindeberg-Lévy theorem for martingales, Proc.Amer.Math.Soc., 12, 788-792.

[5] Billingsley, P. (1967), Convergence of probability measures, John Wiley and Sons Inc., New York.

[6] Breiman, L. (1968), Probability, Addison-Wesley Publishing Co.Inc., Reading, Mass. .

[7] Chung, K.L. (1968), A course in probability theory, Harcourt, Brace & World, Inc., Stanford.

[8] Dianada, P.H. (1955), The central limit theorem for m-dependent random variables, Duke Math.J., 25, 543-546.

[9] Dobrushin, R.L. (1956), Central limit theorem for nonstationary Markov chains, I, Theor.Prob.Appl., 1, 65-80 (English translation).

[10] Feller, W. (1966), An introduction to probability theory and its applications, Vol. II, Wiley, New York.

[11] Gnedenko, B.V. and Kolmogorov, A.N. (1954), Limit distributions of sums of independent random variables, Addison-Wesley Publishing Co., Reading, Mass. .

[12] Hoeffding, W. and Robbins, H. (1948), The central limit theorem for dependent random variables, Duke Math. J., 15, 773-780.

[13] Ibragimov, I.A. (1963), A central limit theorem for a class of dependent random variables, Theor., Prob., Appl., 8, 83-89.

[14] Khinchin, A.I. (1948), Asymptotische Gesetze der Wahrscheinlichkeitsrechnung, New York.

[15] Kolmogorov, A.N. (1932), Sulla forma generale di an processo stocastico omogeneo, Atti, Acad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Nat. (6), 15, 805-808, 866-869.

[16] Kolmogorov, A.N. (1933), Grundbegriffe der Wahrscheinlichkeitsrechnung, Springer-Verlag, Berlin.

[17] Komlós, J. (1970), A central limit theorem for multiplicative systems, (in manuscript).

[18] Lévy, P. (1925), Calcul des probabilités, Paris, Gauthiers-Villars.

[19] Lévy, P. (1935), Propriétés asymptotiques des sommes de variables enchaînées, Bull. Sci. Math., 59, 84-96, 109-128.

[20] Lévy, P. (1935), Propriétés asymptotiques des sommes de variables aléatoires indépendantes ou enchaînées, J. Math. Pures Appl. Ser. 8, 14, 347-402.

[21] Lévy, P. (1935), Sur les intégrans dont les termes sont des variables aléatoires indépendantes, Annali R.Scuola.Sup.Pisa, 3, 336-337 and 4, 217-218.

[22] Lindeberg, J.W. (1922), Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung, Math.Z., 15, 211-225.

[23] Linnik, Yu.V. (1959), An information theoretic proof of the central limit theorem with the Lindeberg condition, Theor.Prob.Appl., 4, 288-298.

[24] Loève, M. (1963), Probability theory, 3rd edition, van Nostrand, Princeton.

[25] Lyapunov, A.M. (1901), Nouvelle forme du théorème sur la limite de probabilité, Mem.Acad.Sci., St Petersbourg, 12, 1-24.

[26] Orey, S. (1958), A central limit theorem for m-dependent random variables, Duke Math.J., 25, 543-546.

[27] Rényi, A. (1966), Wahrscheinlichkeitsrechnung, 2nd edition, VEB Deutscher Verlag der Wissenschaften, Berlin.

[28] Révész, P. (1965), Some remarks on strongly multiplicative systems, Acta Math.Acad.Sci.Hung., 16, 441.

[29] Rosenblatt, M. (1962), Random processes, Oxford Univ.Press, New York.

[30] Trotter, H.F., (1959), An elementary proof of the central limit theorem, Arch.Math., 10, 226-234.

Index of Symbols

C.L.T.	Central Limit Theorem
r.v.	random variable
$\sigma^2(X)$	$E(X-EX)^2$ = variance of X
a.e.	almost everywhere
p.	page
$a \in I$	a is an element of the set I
B_∞	the smallest σ-field containing all sets of the form $\{(x_1, x_2, \ldots),\ x_1 \in I_1, \ldots, x_n \in I_n\}$ for any n where $I_1, \ldots, I_n$ are any intervals.
Ω	abstract set