MTP Presentation

Topic: Central Limit Theorem

Aniket Saha
MDS202207

Chennai Mathematical Institute

January 20, 2023

Outline

Introduction

In the world of Statistics, there are some theorems without which we can not put one step forward. The 'Central Limit Theorem' or popularly known as CLT is one such theorem, and probably the most important. There are many forms of this theorem, we are particularly interested in Lindeberg-Levy CLT for i.i.d. random variables. [Source: Wikipedia]
First of all, we will see what random variables are, what do we mean by sequence of random variables and their convergence, what is characteristic function and then we will go into the proof of the theorem.

Notations Used in the Presentation

We will use the following notations in our presentations:

- First of all, we will work under the probability space $(\Omega, \mathcal{A}, P)$
- Here, Ω is the sample space, $\mathcal{A}$ is the σ-field on Ω, and P is the probability measure.
- $X_1, X_2, ..., X_n$ and X are random variables defined on our probability space, $n \in \mathbb{N}$.
- $E(X_n) = \mu$ and $var(X_n) = \sigma^2$, $\forall n \in \mathbb{N}$
- $S_n = \sum_{k=1}^{n} X_k$, $\forall n \in \mathbb{N}$
- μ and μ_n's are measures on $(\Omega, \mathcal{A})$, $n \in \mathbb{N}$
- F and F_n's are Cumulative Distribution Functions of X and X_n's respectively.

What is a Random Variable?

Definition: Given a probability space $(\Omega, \mathcal{A}, P)$, a random variable is a measurable function from Ω to $\mathbb{R}$.
Mathematically, for all Borel set B in $\mathbb{R}$, X will be called a random variable if $X^{-1}(B) = \{\omega : X(\omega) \in B\} \in \mathcal{A}$.

Sequence of Random Variables

If we have $X_1, X_2, ...$ to be random variables, then $\{X_n\}_{n\geq 1}$ is called a sequence of random variables.
Now, like a sequence of real numbers, we can talk about the convergence of a sequence of random variables. However, unlike real sequences, this convergence can happen in a number of ways. Such as:

1. **Convergence in distribution**
2. **Convergence in probability**
3. **Almost sure convergence**

Convergence of Random Variables

In this presentation we will only encounter convergence in distribution. It is described below:

- **Convergence in Distribution:** Suppose we have a sequence of random variables $\{X_n\}_{n\geq 1}$. We say that X_n converges to X in distribution if:

  $$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

  for all x at which $F_X(x)$ is continuous.

  We write $X_n \xrightarrow{\text{d}} X$.

  Here F_{X_n} and F_X denote the cumulative distribution functions of X_n for $n \in \mathbb{N}$ and X respectively.

Tightness

Now we will see a very useful concept called tightness.

Definition

A sequence of probability measures μ_n on $(\mathbb{R}, \mathcal{R})$ is said to be tight if for each $\epsilon > 0$, there exists a finite interval $(a, b]$ such that $\mu_n(a, b] > 1 - \epsilon$ for all n.
In terms of corresponding distribution functions F_n, the condition is that for each $\epsilon > 0$, there exist x and y such that $F_n(x) < \epsilon$ and $F_n(y) > 1 - \epsilon$ for all n.

Change of Variables

Let $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$ be measurable spaces, and suppose that the mapping $T : \Omega \to \Omega'$ is measurable $\mathcal{F}/\mathcal{F}'$. For a measure μ on $\mathcal{F}$, define a measure μT^{-1} on $\mathcal{F}'$ by,

$$\mu T^{-1}(A') = \mu(T^{-1}A'), \quad A' \in \mathcal{F}' \ldots\ldots\ldots\ldots\ldots\ldots(c1)$$

Suppose f is a real function on Ω' that is measurable $\mathcal{F}'$, so that the composition fT is a real function on Ω that is measurable $\mathcal{F}$.

Theorem

If f is nonnegative, then

$$\int_{\Omega} f(T\omega)\mu(d\omega) = \int_{\Omega'} f(\omega')\mu T^{-1}(d\omega') \ldots\ldots\ldots\ldots(c2)$$

CHENNAI MATHEMATICAL INSTITUTE

Change of Variables

Theorem (Continued)

A function f (not necessarily nonnegative) is integrable with respect to μT^{-1} if and only if fT is integrable with respect to μ, in which case $(c2)$ and

$$\int_{T^{-1}A'} f(T\omega)\mu(d\omega) = \int_{A'} f(\omega')\mu T^{-1}(d\omega') .....(c3)$$

hold.

For nonnegative f, $(c3)$ always holds.

Complex Functions

A complex-valued function on Ω has the form $f(\omega) = g(\omega) + i.h(\omega)$, where g and h are ordinary finite-valued real functions on Ω.
Now, by definition, f is measurable if g and h are measurable. If g and h are integrable, then f is integrable and its integral is defined as:

$$\int (g + ih) d\mu = \int g d\mu + i \int h d\mu$$

. Note that $max\{|g|, |h|\} \leq |f| \leq |g| + |h|$.
Hence f is integrable iff $\int |f| d\mu < \infty$, similar to the real integration case.

Complex Functions

Now we list two results related to integration of complex functions:

1. $|\int f d\mu| \leq \int |f| d\mu$.
   Outline of proof: Suppose $f = g + i.h$, $\therefore |f| = \sqrt{g^2 + h^2}$.
   Then $|\int f d\mu| = |\int (g + ih) d\mu| = |\int g d\mu + i \int h d\mu|$
   $= \sqrt{(\int g d\mu)^2 + (\int h d\mu)^2} \leq \sqrt{\int g^2 d\mu + \int h^2 d\mu}$
   $= \sqrt{\int (g^2 + h^2) d\mu} \leq \int |\sqrt{g^2 + h^2}| d\mu = \int |f| d\mu$.
2. If $f_k = g_k + ih_k$ are complex functions satisfying $\sum_k \int |f_k| d\mu < \infty$. Then $\sum_k \int |g_k| d\mu < \infty$ and hence $\sum_k g_k$ is integrable. Similarly for the imaginary part also we can show that $\sum_k h_k$ is also integrable. Hence $\sum_k f_k$ is integrable and:

$$\int \sum_k f_k d\mu = \sum_k \int f_k d\mu$$

.

CHENNAI MATHEMATICAL INSTITUTE

For a distribution to be uniquely determined, we need to know the cumulative distribution function of the distribution. However, there are some other ways a distribution can be uniquely determined or 'characterised'. One such way is to use moment generating functions or MGF. But since MGF does not exist for many distributions (for example Cauchy distribution) we can not use MGF to prove any general result about any distribution. So we use the characteristic function of a distribution, which exists for all distributions

Definition

The characteristic function $\phi_X(t)$ of a random variable X is defined as:

$$\phi_X(t) = E[e^{itX}] = \int_{\mathbb{R}} e^{itx} dF_X(x), t \in \mathbb{R}, i = \sqrt{-1}$$

Now we look at some of the useful properties of characteristic functions.

- $\phi_X(t) = E[e^{itX}] = E[cos(tX) + i.sin(tX)]$
- $\phi_X(0) = 1$ and $|\phi_X(t)| \leq 1$.
- $|\phi_X(t+h) - \phi_X(t)| \leq E[|e^{ihX} - 1|]$, Uniform Continuity. Specifically, we will need $\phi_X(t)$ is continuous at 0.
- $\phi_{aX_1+bX_2}(t) = [\phi_{X_1}(at)].[\phi_{X_2}(bt)]$, if X_1 and X_2 are independent.

We will prove the most important properties next.

P1. $\phi_X(0) = 1$ and $|\phi_X(t)| \leq 1$
Proof:

$$\phi_X(0) = E[e^{it.0}] = E[e^0] = E[1] = 1$$

$$|\phi_X(t)| = |E[cos(tX) + i.sin(tX)]|$$

$$\leq E[|cos(tX) + i.sin(tX)|]$$

$$= \sqrt{[E[(cos(tX))^2 + (sin(tX))^2]}$$

$$= \sqrt{[E(1)]} = 1$$

Hence the proof.

P2. $\phi_{aX_1+bX_2}(t) = [\phi_{X_1}(at)].[\phi_{X_2}(bt)]$, if X_1 and X_2 are independent.
Proof:

$$\phi_{aX_1+bX_2}(t) = E[e^{it(aX_1+bX_2)}]$$

$$= E[e^{iatX_1}.e^{ibtX_2}] \;=\; E[e^{iatX_1}].E[e^{ibtX_2}] \; [\because \textit{applying independence}]$$

$$= [\phi_{X_1}(at)].[\phi_{X_2}(bt)]$$

Hence the proof.

P3. $\phi_X(t)$ is continuous at 0. **Proof:** If for $t \to 0$ we can show $\phi_X(t) \to \phi_X(0) = 1$, then we are done.
Now,

$$\lim_{t\to 0-} \phi_X(t) = \lim_{t\to 0-} E[e^{itX}]$$
$$= E[e^0] = E[1] = 1$$

Also,

$$\lim_{t\to 0+} \phi_X(t) = \lim_{t\to 0+} E[e^{itX}]$$
$$= E[e^0] = E[1] = 1$$

Hence the proof.

We now state and prove the Levy-Cramer Continuity Theorem.

Continuity Theorem

Let μ_n, μ be probability measures with characteristic functions ϕ_n, ϕ. A necessary and sufficient condition for $\mu_n \Rightarrow \mu$ is that $\phi_n(t) \to \phi(t)$ for each t.

We actually need some theorems and results in order to prove this theorem. We will discuss about those in the next few slides.

Theorem 1

Suppose that μ_n and μ are probability measures on $(R^1, \mathcal{R}^1)$ and $\mu_n \Rightarrow \mu$. There exist random variables Y_n and Y on a common probability space $(\Omega, \mathcal{A}, P)$ such that Y_n has distribution μ_n, Y has distribution μ, and $Y_n(\omega) \to Y(\omega)$ for each ω.

Proof: For the probability space $(\Omega, \mathcal{A}, P)$, take $\Omega = (0, 1)$, let $\mathcal{A}$ consists of the Borel subsets of $(0, 1)$, and for $P(A)$ take the Lebesgue measure of A, for $A \in \mathcal{A}$.
Consider the distribution functions F_n and F corresponding to μ_n and μ.
For $0 < \omega < 1$, put $Y_n(\omega) = inf[x : \omega \leq F_n(x)]$ and
$Y(\omega) = inf[x : \omega \leq F(x)]$. Since $\omega \leq F_n(x)$ if and only if $Y_n(x) \leq \omega$,
$P[\omega : Y_n(\omega) \leq x] = P[\omega : \omega \leq F_n(x)] = F_n(x)$. Thus Y_n has distribution function F_n; similarly, Y has distribution F.

It remains to show that $Y_n(\omega) \to Y(\omega)$. Suppose that $0 < \omega < 1$. Given ϵ, choose x so that $Y(\omega) - \epsilon < x < Y(\omega)$ and $\mu(x) = 0$. Then $F(x) < \omega$; $F_n(x) \to F(x)$ now implies that , for n large enough, $F_n(x) < \omega$ and hence

$$Y(\omega) - \epsilon < x < Y_n(\omega)$$

Thus,

$$liminf_n \; Y_n(\omega) \geq Y(\omega)$$

If $\omega < \omega'$ and ϵ is positive, choose a y for which $Y(\omega') < y < Y(\omega') + \epsilon$ and $\mu\{y\} = 0$. Now $\omega < \omega' \leq F(Y(\omega')) \leq F(y)$, and so, for n large enough, $\omega \leq F_n(y)$ and hence

$$Y_n(\omega) \leq y < Y(\omega') + \epsilon$$

Thus if $\omega < \omega'$ then,

$$limsup_n Y_n(\omega) \leq Y(\omega')$$

and so ,

$$Y(\omega) \leq liminf_n Y_n(\omega) \leq limsup_n Y_n(w) \leq Y(\omega')$$

$i.e$, Y is nondecreasing. If Y is continuous at ω, then $\lim_{\omega' \to \omega} Y(\omega') = Y(\omega)$.
and so,

$$limsup_n Y_n(\omega) \leq Y(\omega') \Rightarrow limsup_n Y_n(\omega) \leq Y(\omega)$$

$$\therefore \; Y(\omega) \geq limsup_n Y_n(\omega) \geq liminf_n Y_n(\omega) \geq Y(\omega)$$

$$\implies Y(\omega) = limsup_n Y_n(\omega) = liminf_n Y_n(\omega)$$

$$\implies \lim_n Y_n(\omega) = Y(\omega)$$

$i.e$, $Y_n(\omega) \to Y(\omega)$ if Y is continuous at ω.

Since Y is nondecreasing on $(0,1)$, it has at most countably many discontinuities. At discontinuity points ω of Y, redefine $Y_n(\omega) = Y(\omega) = 0$. With this change, $Y_n(\omega) \to Y(\omega)$ for every ω. Since Y and the Y_n have been altered only on a set of Lebesgue measure 0, their distribution are still μ_n and μ.
Hence the proof.

Theorem 2

$\mu_n \to \mu \implies \int f d\mu_n \to \int f d\mu$ for every bounded continuous real valued function f.

Proof: Suppose that $\mu_n \to \mu$ and consider the random variable Y_n and Y of theorem 1. Suppose that f is bounded continuous real function.

$$Y_n(\omega) \to Y(\omega) \forall \omega \in \Omega$$

$i.e. f(Y_n) \to f(Y)$ almost everywhere.
f is bounded $\implies \exists M \in \mathbb{R} s.t |f(Y_n)(\omega)| \leq M \ \forall n$ and ω. which means $|f(Y_n)|$ are uniformly bounded.
Also by definition $P(\Omega) = 1 < \infty$.
$\therefore$ By Bounded Convergence Theorem,

$$\int_\Omega f(Y_n) dP \to \int_\Omega f(Y) dP \tag{1}$$

Continuity Theorem

Theorem 2

By change of variables,

$$\int_\Omega f(Y_n)dP = \int_\Omega f(Y_n(\omega))P(d\omega) = \int_\mathbb{R} f(x)\mu_n(dx) = \int_{-\infty}^{\infty} fd\mu_n$$

and,

$$\int_\Omega f(Y)dP = \int_\Omega f(Y(\omega))P(d\omega) = \int_\mathbb{R} f(x)\mu(dx) = \int_{-\infty}^{\infty} fd\mu$$

$$\therefore (1) \implies \int_{-\infty}^{\infty} fd\mu_n \to \int_{-\infty}^{\infty} fd\mu$$

Hence the proof.

Theorem 3

For every sequence $\{F_n\}$ of distribution functions there exists a subsequence $\{F_{n_k}\}$ and a nondecreasing, right-continuous function F such that $\lim_k F_{n_k}(x) = F(x)$ at continuity points x of F.

Proof: Here we are using the following result
"Suppose that each row of the array

$$\begin{array}{cccc} x_{1,1} & x_{1,2} & x_{1,3} & \cdots \\ x_{2,1} & x_{2,2} & x_{2,3} & \cdots \\ \ldots & \ldots & \ldots & \ldots \end{array}$$

is a bounded sequence of real numbers. Then there exists an increasing sequence $n_1, n_2, \ldots$ of integers such that the limit $\lim_{k\to\infty} x_{r,n_k}$ exists for $r = 1, 2, \ldots$"

Since $\mathbb{Q}$ is countable, we can write it as $\mathbb{Q} = \{r_1, r_2, ...\}$. Consider the array,

$$\begin{array}{cccc} F_1(r_1) & F_2(r_1) & F_3(r_1) & ... \\ F_1(r_2) & F_2(r_2) & F_3(r_2) & ... \\ ... & ... & ... & ... \end{array}$$

By the above result, we gets a sequence $\{n_k\}$ of integers along which the limit $G(r) = \lim_k F_{n_k}(r)$ exists for every rational r. Define
$F(x) = inf[G(r) : x < r]$.
if $x_1 < x_2$, then $\{G(r) : x_1 < r\} \supseteq \{G(r) : x_2 < r\} \Rightarrow F(x_1) \leq F(x_2)$. So F is nondecreasing.

To each x and ϵ there is an r for which $x < r$ and $G(r) < F(x) + \epsilon$. If $x \leq y < r$, then $F(x) \leq F(y) \leq G(r) < F(x) + \epsilon$. *i.e*, if x is fixed, then for given $\epsilon > 0$, we can find $\delta = r - x$, such that
$y \in [x, x + \delta) \Rightarrow F(y) \in [F(x), F(x) + \epsilon)$. It is true for all $x \in \mathbb{R}$. Hence F is continuous from the right.
Now we show that $\lim_k F_{n_k}(x) = F(x)$ at continuity points x of F.
If F is continuous at x, choose $y < x$ so that $F(x) - \epsilon < F(y) \leq F(x)$.
Now choose rational r and s so that $y < r < x < s$ and $G(s) < F(x) + \epsilon$.
So,

$$F(x) - \epsilon < G(r) \leq G(s) < F(x) + \epsilon \tag{2}$$

and,

$$F_{n_k}(r) \leq F_{n_k}(x) \leq F_{n_k}(s) \tag{3}$$

$$\implies liminf_k F_{n_k}(r) \le liminf_k F_{n_k}(x) \le liminf_k F_{n_k}(s)$$

and $$limsup_k F_{n_k}(r) \le limsup_k F_{n_k}(x) \le limsup_k F_{n_k}(s)$$

$$\implies G(r) \le liminf_k F_{n_k}(x) \le G(s)$$

and $$G(r) \le limsup_k F_{n_k}(x) \le G(s) \therefore (2) \implies \text{ for all } \epsilon > 0,$$

$$F(x) - \epsilon < liminf_k F_{n_k}(x) < F(x) + \epsilon$$

and $$F(x) - \epsilon < limsup_k F_{n_k}(x) < F(x) + \epsilon$$

$$\implies liminf_k F_{n_k}(x) = limsup_k F_{n_k}(x) = F(x)$$

$\therefore \lim_k F_{n_k}(x) = F(x)$ at continuity points x of F.
Hence the proof.

Theorem 4

Let the sequence of probability measures $\{\mu_n\}$ is tight. Then for every subsequence $\{\mu_{n_k}\}$ there exist a further subsequence $\{\mu_{n_{k_{(j)}}}\}$ and a probability measure μ such that $\mu_{n_{k_{(j)}}} \Rightarrow \mu$ as $j \to \infty$.

Proof: Apply theorem 3 to the subsequence $\{F_{n_k}\}$ of corresponding distribution functions. There exists a further subsequence $\{F_{n_{k_{(j)}}}\}$ such that $\lim_j F_{n_{k_{(j)}}}(x) = F(x)$ at continuity points of F, where F is nondecreasing and right-continuous. So, there exists a measure μ on $(R^1, \mathcal{R}^1)$ such that $\mu(a, b] = F(b) - F(a)$. Given ϵ choose a and b so that $\mu_n(a, b] > 1 - \epsilon$ for all n, which is possible by tightness. By decreasing a and increasing b one can ensure that they are continuity points of F.

Thus,

$$\mu_n(a,b] > 1-\epsilon \quad \forall n \tag{4}$$

and, $\mu(a,b] = F(b) - F(a) = \lim_j F_{n_{k_{(j)}}}(b) - \lim_j F_{n_{k_{(j)}}}(a)$

$$= \lim_j (F_{n_{k_{(j)}}}(b) - F_{n_{k_{(j)}}}(a)) = \lim_j \mu_{n_{k_{(j)}}}(a,b]$$

$\therefore (4) \implies \mu(a,b] \geq 1-\epsilon$

$i.e$, for given $\epsilon > 0$, $\exists$continuity points of F, a and b such that $\mu(a,b] \geq 1-\epsilon$. Therefore, μ is a probability measure, and of course $\mu_{n_{k_{(j)}}} \Rightarrow \mu$.

Hence the proof.

Theorem 5: Corollary of Theorem 4

If $\{\mu_n\}$ is a tight sequence of probability measures, and if each subsequence that converges weakly at all converges weakly to the probability measure μ, then $\mu_n \Rightarrow \mu$.

Proof: By theorem 4, each subsequence $\{\mu_{n_k}\}$ contains a further subsequence $\{\mu_{n_{k(j)}}\}$ converging weakly ($j \to \infty$) to some limit, and that limit must by hypothesis be μ. Thus every subsequence $\{\mu_{n_k}\}$ contains a further subsequence $\{\mu_{n_{k(j)}}\}$ converging weakly to μ.

Suppose that $\mu_n \Rightarrow \mu$ is false. Then there exists some x such that $\mu(x) = 0$ but $\mu_n(-\infty, x]$ does not converge to $\mu(-\infty, x]$. But then there exists a positive ϵ such that $|\mu_{n_k}(-\infty, x] - \mu(-\infty, x]| \geq \epsilon$ for an infinite sequence $\{n_k\}$ of integers, and no subsequence of $\{\mu_{n_k}\}$ can converge weakly to μ. This contradiction shows that $\mu_n \Rightarrow \mu$.
Hence the proof.

Now we are equipped to prove the Continuity Theorem.
Necessity Part: $\mu_n \Rightarrow \mu$. For each t, both $'cos(tx)'$ and $'sin(tx)'$ are bounded and continuous in x. Therefore, by theorem 2, $\int cos(tx)\ d\mu_n \to \int cos(tx)\ d\mu$ and $\int sin(tx)\ d\mu_n \to \int sin(tx)\ d\mu$. So,

$$\begin{aligned}
\lim_{n\to\infty} \phi_n(t) &= \lim_{n\to\infty} \int (cos(tx) + isin(tx))\ d\mu_n \\
&= \lim_{n\to\infty} (\int cos(tx)\ d\mu_n + i \int sin(tx)\ d\mu_n) \\
&= \lim_{n\to\infty} \int cos(tx)\ d\mu_n + i \lim_{n\to\infty} \int sin(tx)\ d\mu_n \\
&= \int cos(tx)\ d\mu + i \int sin(tx)\ d\mu
\end{aligned}$$

$$= \int (cos(tx) + isin(tx))\ d\mu$$
$$= \phi(t)$$

So, $\phi_n(t) \to \phi(t)$ for each t.
Hence the proof of the Necessity Part.
Sufficiency Part: For $u > 0$,

$$\frac{1}{u}\int_{-u}^{u}(1-\phi_n(t))dt = \frac{1}{u}\int_{-u}^{u}(\int_{-\infty}^{\infty} 1\mu_n(dx) - \int_{-\infty}^{\infty} e^{itx}\mu_n(dx))dt$$
$$= \int_{-u}^{u}\int_{-\infty}^{\infty}\frac{1}{u}(1-e^{itx})\mu_n(dx)dt$$

$$= \int_{-\infty}^{\infty} [\frac{1}{u} \int_{-u}^{u} (1 - e^{itx}) dt] \mu_n(dx) \quad (By\ Fubini's\ theorem)$$

$$\int_{-u}^{u} 1 dt = 2u,$$

$$\int_{-u}^{u} e^{itx} dt = \int_{-u}^{u} (cos(tx) + isin(tx)) dt = \int_{-u}^{u} cos(tx) dt + i \int_{-u}^{u} sin(tx) dt$$

$$= [\frac{sin(tx)}{x}]_{-u}^{u} + i[\frac{-cos(tx)}{x}]_{-u}^{u} = \frac{2sin(ux)}{x}$$

$$\therefore \frac{1}{u} \int_{-u}^{u} (1 - \phi_n(t)) dt = 2 \int_{-\infty}^{\infty} (1 - \frac{sin(ux)}{ux}) \mu_n(dx)$$

$$(1 - \tfrac{sinx}{x}) \geq 0 \ \ \forall x \in \mathbb{R} \implies \tfrac{sin(ux)}{ux} \leq |\tfrac{sin(ux)}{ux}| \leq \tfrac{1}{|ux|} \implies$$

$$1 - \frac{sin(ux)}{ux} \geq 1 - \frac{1}{|ux|} \quad \forall x \in \mathbb{R} \tag{5}$$

$$(1 - \tfrac{sin(ux)}{ux}) \geq 0 \ \ \forall x \in \mathbb{R} \implies$$

$$\int_{-2/u}^{2/u} (1 - \frac{sin(ux)}{ux})\mu(dx) \geq \int_{-2/u}^{2/u} 0\mu(dx) = 0 \tag{6}$$

$$
\begin{aligned}
\therefore \frac{1}{u}\int_{-u}^{u}(1-\phi_n(t))dt &= 2\int_{-\infty}^{\infty}(1-\frac{sin(ux)}{ux})\mu_n(dx) \\
&= 2[\int_{|x|\geq 2/u}(1-\frac{sin(ux)}{ux})\mu_n(dx) \\
&\qquad + \int_{|x|<2/u}(1-\frac{sin(ux)}{ux})\mu_n(dx)] \\
&\geq 2\int_{|x|\geq 2/u}(1-\frac{sin(ux)}{ux})\mu_n(dx) && (\because (6)) \\
&\geq 2\int_{|x|\geq 2/u}(1-\frac{1}{|ux|})\mu_n(dx) && (\because (5))
\end{aligned}
$$

$|x| \geq \frac{2}{u} \Rightarrow \frac{1}{|ux|} \leq \frac{1}{2} \Rightarrow (1 - \frac{1}{|ux|}) \geq \frac{1}{2}$

$$\therefore \frac{1}{u} \int_{-u}^{u} (1 - \phi_n(t))dt \geq 2 \int_{|x| \geq 2/u} (1 - \frac{1}{|ux|}) \mu_n(dx)$$

$$\geq 2 \int_{|x| \geq 2/u} \frac{1}{2} \mu_n(dx) = \int_{|x| \geq 2/u} \mu_n(dx)$$

$$\therefore \frac{1}{u} \int_{-u}^{u} (1 - \phi_n(t))dt \geq \mu_n[x : |x| \geq 2/u] \tag{7}$$

Since ϕ is continuous at the origin and $\phi(0) = 1$, there for positive ϵ a u for which $|0 - t| < u \Rightarrow |\phi(0) - \phi(t)| < \frac{\epsilon}{4}$. thus,

$$\frac{1}{u} \int_{-u}^{u} (1 - \phi_n(t))dt \leq \frac{1}{u} \int_{-u}^{u} \frac{\epsilon}{4} dt = \frac{\epsilon}{2}$$

Note from (7) that $\frac{1}{u}\int_{-u}^{u}(1-\phi_n(t))dt$ is real and non-negative.

$$\therefore \frac{1}{u}\int_{-u}^{u}(1-\phi_n(t))dt = \left|\frac{1}{u}\int_{-u}^{u}(1-\phi_n(t))dt\right| \leq \frac{1}{u}\int_{-u}^{u}|1-\phi_n(t)|dt \leq \frac{\epsilon}{2}$$

$$\therefore \frac{1}{u}\int_{-u}^{u}(1-\phi_n(t))dt < \epsilon \tag{8}$$

since $\phi_n \to \phi$, $(1-\phi_n) \to (1-\phi)$.
$|\phi_n(t)| \leq 1 \; \forall t \in \mathbb{R} \implies |1-\phi_n(t)| \leq |1| + |\phi_n(t)| \leq 2 \;\; \forall n, t$
$\therefore \{(1-\phi_n(t))\}_n$ is uniformly bounded, also clearly μ is a bounded measure.

$\therefore$ by Bounded Convergence Theorem,

$$\lim_{n\to\infty}\int_{-u}^{u}(1-\phi_n(t))dt = \int_{-u}^{u}(1-\phi(t))dt$$

$$\therefore \lim_{n\to\infty}\frac{1}{u}\int_{-u}^{u}(1-\phi_n(t))dt = \frac{1}{u}\int_{-u}^{u}(1-\phi(t))dt$$

So, for given $\epsilon > 0$, $\exists n_0 \in \mathbb{N}$ such that,

$$\left|\frac{1}{u}\int_{-u}^{u}(1-\phi(t))dt - \frac{1}{u}\int_{-u}^{u}(1-\phi_n(t))dt\right| < \epsilon \quad \forall n \geq n_0$$

$$\implies \frac{1}{u}\int_{-u}^{u}(1-\phi(t))dt - \epsilon < \frac{1}{u}\int_{-u}^{u}(1-\phi_n(t))dt$$

Continuity Theorem

Proof

$$< \frac{1}{u} \int_{-u}^{u} (1 - \phi(t))dt + \epsilon < 2\epsilon \ \ \forall n \geq n_0 \quad (\because (8))$$

if $a = 2/u$ in (3), then

$$\mu_n[x : |x| \geq a] < 2\epsilon \quad \forall n \geq n_0$$

So,

$$\mu_n[x : x \in (-a, a)] > 1 - 2\epsilon \quad \forall n \geq n_0$$

$$\therefore \mu_n[x : x \in (-a, a]] \geq \mu_n[x : x \in (-a, a)] > 1 - 2\epsilon \ \ \forall n \geq n_0$$

$$\Rightarrow \mu_n[x : x \in (-a', a']] \geq \mu_n[x : x \in (-a, a]] > 1 - 2\epsilon \ \ \forall a' \geq a, n \geq n_0$$

For each $i < n_0$, $\exists b_i$ such that $\mu_i[x : x \in (-b_i, b_i]] > 1 - 2\epsilon$. Now take $max\{a, b_1, b_2, .., b_{n_0 - 1}\} = A$, then,

$$\mu_n[x : x \in (-A, A]] > 1 - 2\epsilon \quad \forall n \in \mathbb{N}$$

$\therefore \{\mu_n\}$ is tight.

By Theorem 5, $\mu_n \Rightarrow \mu$ will follow if it is shown that each subsequence $\{\mu_{n_k}\}$ that converges weakly at all converges weakly to μ.
But if $\mu_n \Rightarrow \nu$ as $k \to \infty$, then by the necessity half of the theorem, already proved, ν has characteristic function $\lim_k \phi_{n_k}(t) = \phi(t)$. By uniqueness of characteristic functions, ν and μ must coincide.
Hence the proof of the Sufficiency Part.

Central Limit Theorem
Statement

Now we are going to state and prove the Central Limit Theorem for iid random variables.

Central Limit Theorem: Lindeberg-Levy Version

Suppose $\{X_n\}_{n\geq 1}$ is a sequence of independent and identically distributed (i.i.d.) random variables with mean μ and finite positive variance σ^2. If $S_n = X_1 + X_2 + ... + X_n$, for $n \in \mathbb{N}$, then-

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \Rightarrow N(0,1)$$

To prove the CLT, we need the following.

1. First of all we need the 'Continuity theorem' proved previously.
2. We will need Lemma C1, proved next.
3. We need to know about the Lindeberg condition, which is also discussed in the following slides.
4. For complex z, we have:

$$|e^z - 1 - z| \leq |z|^2 \sum_{k=2}^{\infty} \frac{|z|^{k-2}}{k!} \leq |z|^2 e^{|z|}$$

5. If X has a moment of order n, it follows that:

$$|\phi(t) - \sum_{k=0}^{n} \frac{(it)^k}{k!} E(X^k)| \geq E[min\{\frac{|tX|^{n+1}}{(n+1)!}, \frac{2|tX|^n}{n!}\}]$$

We need the following lemma in the proof of the CLT. We have named the lemma C1 for reference.

Lemma C1

Let $z_1, z_2, ..., z_m$ and $w_1, w_2, ..., w_m$ be complex numbers of modulus at most 1, then:

$$\left|z_1.z_2...z_m - w_1.w_2...w_m\right| \leq \sum_{k=1}^{m} \left|z_k - w_k\right|$$

Proof: Now,

$$\left|z_1.z_2...z_m - w_1.w_2...w_m\right| = \left|z_1...z_m \;-\; z_1...z_{m-1}w_m \;+\; z_1...z_{m-1}w_m\right.$$

$$- \; z_1...z_{m-2}w_{m-1}.w_m \;+\; z_1...z_{m-2}w_{m-1}.w_m$$

$$\left. - \; ... \;+\; z_1.w_2...w_m \;-\; w_1.w_2...w_m\right|$$

$$= \left|z_1 \ldots z_{m-1}[z_m \ - \ w_m] + z_1 \ldots z_{m-2}.w_m.[z_{m-1} \ - \ w_{m-1}] \right.$$
$$\left. + \ \ldots \ + \ w_2.w_3 \ldots w_m[z_1 \ - \ w_1]\right|$$

$$\leq |z_m \ - \ w_m|.|z_1 \ldots z_{m-1}| \ + \ |z_{m-1} \ - \ w_{m-1}|.|z_1 \ldots z_{m-2}.w_m| \ + \ \ldots \ + \ |z_1 \ - \ w_1|.|w_2.w_3 \ldots w_m|$$

$$\leq \sum_{k=1}^{m} |(z_k \ - \ w_k).1| \qquad [\because \text{ modulus of } z_k \text{ and } w_k \text{ are at most } 1, \ \forall \text{ k}]$$

$$= \sum_{k=1}^{m} |z_k \ - \ w_k|$$

Hence, $|z_1.z_2 \ldots z_m - w_1.w_2 \ldots w_m| \leq \sum_{k=1}^{m} |z_k \ - \ w_k|$ (Proved)

The Lindeberg condition is essential in our proof of CLT, we will discuss about it now.

Lindeberg Condition

Let $\{X_{nj} : 1 \leq j \leq r_n\}_{n \geq 1}$ be a triangular array of independent random variables such that $E(X_{nj}) = 0$ and $0 < E(X_{nj}^2) = \sigma_{nj}^2 < \infty$ for all $1 \leq j \leq r_n$, $n \geq 1$.
Then, $\{X_{nj} : 1 \leq j \leq r_n\}_{n \geq 1}$ is said to satisfy the Lindeberg condition if for every $\epsilon > 0$,

$$\lim_{n \to \infty} s_n^{-2} \sum_{j=1}^{r_n} E(X_{nj}^2 . I(|X_{nj}| > \epsilon . s_n)) = 0$$

where $s_n^2 = \sum_{j=1}^{r_n} \sigma_{nj}^2$, $n \geq 1$.

Now since we have all the necessary tools required for the proof of CLT, we will start the proof here.
First of all, let us define a traingular array as defined while mentioning Lindeberg condition using $X_1, X_2, ...$
We will take $r_n = n$, $\forall n \in \mathbb{N}$ and $X_{nj} = X_j$, $1 \le j \le r_n = n$, $\forall n \in \mathbb{N}$.
Therefore our triangular array will look like:

$$\begin{aligned} n = 1: &\qquad X_1 \\ n = 2: &\qquad X_1,\ X_2 \\ n = 3: &\qquad X_1,\ X_2,\ X_3 \\ &...and\ so\ on... \end{aligned}$$

Note that X_i's are independent, and their mean and variance exist and are finite. Hence the conditions for 'Lindeberg Condition' to hold are satisfied by this array.

Now we will change the X_{nj}'s to Y_{nj}'s as $Y_{nj} = (X_j - \mu)/s_n$, where $s_n = [\sum_{j=1}^{r_n} \sigma_{nj}^2]^{\frac{1}{2}} = [\sum_{j=1}^{n} \sigma^2]^{\frac{1}{2}} = [n\sigma^2]^{\frac{1}{2}} = \sigma\sqrt{n}$, for $1 \leq j \leq n$, $\forall n \in \mathbb{N}$.

By doing this transformation, we now have a triangular array of random variables $\{Y_{nj}\}, 1 \leq j \leq n, n \in \mathbb{N}$ with $E(Y_{nj}) = 0$ and,
$E(Y_{nj}^2) = E((X_j - \mu)/s_n)^2 = \frac{\sigma^2}{n.\sigma^2} = \frac{1}{n}$ for $1 \leq j \leq n$ and $n \in \mathbb{N}$.
Note that since X_i's are iid, for a fixed $n \in \mathbb{N}$, Y_nj's are iid,
$1 \leq j \leq n, \forall n \in \mathbb{N}$

For Y_{nj}'s, we define $S'_n = \sum_{k=1}^{n} Y_{nj} = \sum_{k=1}^{n} \frac{X_j - \mu}{s_n} = \frac{S_n - n\mu}{s_n}$, for $1 \leq j \leq n$ and $n \in \mathbb{N}$ and $s'_n = [\sum_{j=1}^{r_n} E(Y_{nj})^2]^{\frac{1}{2}} = [\sum_{j=1}^{n} \frac{1}{n}]^{\frac{1}{2}} = [n.\frac{1}{n}]^{\frac{1}{2}} = 1$ for $n \in \mathbb{N}$.
To show, $\lim_{n\to\infty} \frac{S_n - n\mu}{\sigma\sqrt{n}} = \lim_{n\to\infty} \frac{S_n - n\mu}{s_n} = \lim_{n\to\infty} \frac{S'_n}{s'_n} \Rightarrow N(0, 1)$.

Firstly we want to show that our array of random variables $Y_{nj}, 1 \leq j \leq n, n \in \mathbb{N}$ actually satisfies the Lindeberg condition. Now,

$$\lim_{n\to\infty} s_n'^{-2} \sum_{j=1}^{r_n} E(Y_{nj}^2 . I(|Y_{nj}| > \epsilon . s_n'))$$

$$= \lim_{n\to\infty} 1. \sum_{j=1}^{n} E(Y_{nj}^2 . I(|Y_{nj}| > \epsilon))$$

$$= \lim_{n\to\infty} \sum_{j=1}^{n} E((\frac{X_j - \mu}{s_n})^2 . I(|\frac{X_j - \mu}{s_n}| > \epsilon))$$

$$= \lim_{n\to\infty} \frac{1}{s_n^2} \sum_{j=1}^{n} E((X_j - \mu)^2 . I(|X_j - \mu| > s_n . \epsilon))$$

$$
\begin{aligned}
&= \lim_{n\to\infty} \frac{1}{n\sigma^2} n.E((X_1-\mu)^2.I(|X_1-\mu| > \epsilon\sigma\sqrt{n})) \quad [\because X_j's \text{ are } iid] \\
&= \frac{1}{\sigma^2} \lim_{n\to\infty} \int_{|X_1-\mu|>\epsilon\sigma\sqrt{n}} (X_j-\mu)^2 dP \\
&= \frac{1}{\sigma^2} \int_{\phi} (X_j-\mu)^2 dP \quad = \quad 0
\end{aligned}
$$

We can write the last line because of:

1. $E(X_1-\mu)^2 < \infty$ and hence we can apply DCT on the integrand.
2. $[|X_1-\mu| > \epsilon\sigma\sqrt{n}] \downarrow \phi$ as $n \to \infty$

Hence we can conclude that our triangular array of random variables Y_{nj} satisfy the Lindeberg Condition.

Since Y_{nj}'s satisfy the Lindeberg Condition, by Continuity Theorem, now it is enough to show that $\lim_{n\to\infty} E(e^{itS'_n}) = e^{-\frac{t^2}{2}}$.
Now observe that for $n \in \mathbb{N}$:

$$\begin{aligned}
\Delta_n &= max\{E(Y_{nj}^2) : 1 \leq j \leq n\} \\
&= max\{\frac{1}{n} : 1 \leq j \leq n\} \qquad [\because E(Y_{nj}^2) = \frac{1}{n}, 1 \leq j \leq n] \\
&= \frac{1}{n} \\
&\to 0 \; as \; n \to \infty
\end{aligned}$$

We will need this later in the proof. Call it R1.

Central Limit Theorem

Proof

Suppose $\phi_{nj}(t)$ is the characteristic function of Y_{nj},
$1 \leq j \leq n,\ \ n \in \mathbb{N}, t \in \mathbb{R}$.
Now, we know that the second order moments of Y_{nj}'s exist.
Hence we use the following result:

Result

If X has a moment of order n, and $\phi(t)$ is its characteristic function, then it follows that:

$$|\phi(t)\ -\ \sum_{k=0}^{n} \frac{(it)^k}{k!} E(X^k)|\ \leq\ E[min\{\frac{|tX|^{n+1}}{(n+1)!}, \frac{2|tX|^n}{n!}\}]$$

Hence we have:

$$|\phi_{nj}(t) - (1 - \frac{1}{2} t^2 \sigma_{nj}^2)| \leq E[min\{|tY_{nj}|^2, |tY_{nj}|^3\}]\ , 1 \leq j \leq n, n \in \mathbb{N}$$

Now, for $\epsilon > 0$, we have:

$$E[min\{|tY_{nj}|^2, |tY_{nj}|^3\}] \leq E(|tY_{nj}|^2.I(|Y_{nj}| > \epsilon)) + E(|tY_{nj}|^3.I(|Y_{nj}| \leq \epsilon))$$

$$\leq t^2 E(|Y_{nj}|^2.I(|Y_{nj}| > \epsilon)) + |t|^3 \epsilon \sigma'^2_{nj}$$

$$\rightarrow 0 \ as \ n \rightarrow \infty$$

$\because$ the first term goes to 0 due to Lindeberg Condition and the second term goes to 0 due to ϵ being arbitrary and σ'_{nj}'s are bounded by 0 and 1.
Hence we have:

$$\sum_{j=1}^{n} |\phi_{nj}(t) - (1 - \frac{t^2 \sigma'^2_{nj}}{2})| \rightarrow 0, \ as \ n \rightarrow \infty, \ for \ t \in \mathbb{R}$$

Now it is enough to show that:

$$\prod_{j=1}^{n} \phi_{nj}(t) \to \prod_{j=1}^{n}(1 - \frac{t^2\sigma_{nj}'^2}{2})$$

$$\to \prod_{j=1}^{n} e^{-\frac{t^2\sigma_{nj}'^2}{2}} \; = \; e^{-\frac{t^2}{2}}$$

Note that from R1, $\exists n_0 \in \mathbb{N}$ such that for all $n \geq n_0$:
$l_{1n} = max\{|1 - \frac{t^2\sigma_{nj}'^2}{2}| : 1 \leq j \leq n\} \leq 1$, where, $\sigma_{nj}'^2 = E(Y_{nj}^2)$.
Now we can apply Lemma 1 on $|\prod_{j=1}^{n} \phi_{nj}(t) - \prod_{j=1}^{n}(1 - \frac{t^2\sigma_{nj}'^2}{2})|$.
Hence we get as $n \to \infty$:

$$|\prod_{j=1}^{n} \phi_{nj}(t) - \prod_{j=1}^{n}(1 - \frac{t^2\sigma_{nj}'^2}{2})| \leq \sum_{j=1}^{n} |\phi_{nj}(t) - (1 - \frac{t^2\sigma_{nj}'^2}{2})| \to 0$$

Hence we have:

$$\prod_{j=1}^{n} \phi_{nj}(t) \xrightarrow{n\to\infty} \prod_{j=1}^{n} (1 - \frac{t^2\sigma'^2_{nj}}{2})$$

Now, to show: $\prod_{j=1}^{n}(1 - \frac{t^2\sigma'^2_{nj}}{2}) \to \prod_{j=1}^{n} e^{-\frac{t^2\sigma'^2_{nj}}{2}}$

Observe that, $|e^{-\frac{t^2\sigma'^2_{nj}}{2}}|, |1 - \frac{t^2\sigma'^2_{nj}}{2}| \leq 1$.

Hence we can use Lemma 1 again on $|\prod_{j=1}^{n} e^{-\frac{t^2\sigma'^2_{nj}}{2}} - \prod_{j=1}^{n}(1 - \frac{t^2\sigma'^2_{nj}}{2})|$
We get:

$$|\prod_{j=1}^{n} e^{-\frac{t^2\sigma'^2_{nj}}{2}} - \prod_{j=1}^{n}(1 - \frac{t^2\sigma'^2_{nj}}{2})| \leq \sum_{j=1}^{n} |e^{-\frac{t^2\sigma'^2_{nj}}{2}} - 1 + \frac{t^2\sigma'^2_{nj}}{2}|$$

Central Limit Theorem

Proof

Now, we use the following result:

Result

For complex z, we have:

$$|e^z - 1 - z| \leq |z|^2 \sum_{k=2}^{\infty} \frac{|z|^{k-2}}{k!} \leq |z|^2 e^{|z|}$$

Hence

$$\sum_{j=1}^{n} |e^{-\frac{t^2\sigma'^2_{nj}}{2}} - 1 + \frac{t^2\sigma'^2_{nj}}{2}| \leq \sum_{j=1}^{n} |-\frac{t^2\sigma'^2_{nj}}{2}|^2 e^{|-\frac{t^2\sigma'^2_{nj}}{2}|}$$

$$= \sum_{j=1}^{n} \frac{t^4.\sigma'^4_{nj} e^{\frac{t^2\sigma'^2_{nj}}{2}}}{4} \leq \frac{t^4 e^{t^2}}{4} \sum_{j=1}^{n} \sigma'^4_{nj} \to 0$$

The last line is true since, $e^{|-\frac{t^2\sigma'^2_{nj}}{2}|} \leq e^{t^2}$ and $\sigma'^2_{nj} \to 0$ as $n \to \infty$

Hence we get:

$$\prod_{j=1}^{n}(1-\frac{t^2\sigma_{nj}'^2}{2}) \to \prod_{j=1}^{n} e^{-\frac{t^2\sigma_{nj}'^2}{2}}$$

$$= e^{-\frac{t^2}{2}} \quad [\because \sum_{j=1}^{n}\sigma_{nj}'^2 = 1]$$

$\therefore$ we get

$$E(e^{itS_n'}) = \prod_{j=1}^{n}\phi_{nj}(t) \to e^{-\frac{t^2}{2}}$$

Hence the proof.

References and Acknowledgement

References

1. Probability and Measure, Patrick Billingsley
2. Fundamentals of Probability Theory, Chandra & Gangopadhay
3. Measure Theory and Probability Theory, K. B. Athreya & S. N. Lahiri
4. Notes by B. V. Rao
5. Introduction to Measure Theory, Terrence Tao
6. Introduction to Probability Theory, William Feller

Acknowledgement

We are grateful to professor R. L. Karandikar to let us work on this presentation and give us enough time and support to complete this. Thank You Sir.