# Characteristic functions and the central limit theorem

Saifuddin Syed, #20387835
University of Waterloo
Submitted to Dr. Kathryn Hare for PMATH 800

April 19, 2013

**Abstract**

The central limit theorem is one of the cornerstones of probability and statistics. It allows us to determine how the sample mean deviates from the the true mean mean regardless of the underlying probability distribution. In this paper, we discuss how we can analyse the limiting behaviour of sequences of random variables via characteristic functions. The Lindeberg-Lévy-Feller theorem is proven and consequences are provided. We mention applications of the central limit theorem, including the delta method and Stirling's formula. Rates of convergence and dependence of random variables are also discussed.

## 1 History/Importance

The central limit theorem is considered by some to be the most important theorem in probability theory and statistics. It allows us to determine long term behaviour of random variables, and approximate complicated problems. The strong law of large numbers tells us that in the case of iid random variables, the arithmetic mean converges to the expectation value a.s. The central limit theorem tell us the asymptotic distribution of the error between sample mean and the expectation value. It allows us the determine the rate at which strong law of large numbers holds.

**Theorem 1.1** (Central limit theorem)**.** *Suppose $X_1, X_2, \dots$ are iid random variables such that $E(X_n) = \mu, \mathrm{Var}(X_n) = \sigma^2 < \infty$. Let $S_n = \sum_{i=1}^n X_i$, and let $N$ be a normally distributed random variable, with $E(N) = 0, \mathrm{Var}(N) = 1$. Then,*

$$\frac{\frac{S_n}{n} - E\left(\frac{S_n}{n}\right)}{\sqrt{\mathrm{Var}\left(\frac{S_n}{n}\right)}} = \frac{S_n - n\mu}{\sqrt{n}\sigma} \Longrightarrow N$$

This theorem is important because of its simplicity. It implies that when $n$ is large, $\frac{S_n}{n}$ is approximately normally distributed with mean $\mu$ and variance $\frac{\sigma^2}{n}$.

The central limit theorem was originally postulated by Abraham deMoivre in 1733 when he used the normal distribution to approximate the distribution of the number of heads that appear after tossing a fair coin. It was largely forgotten until 1812 when Laplace expanded on DeMoivre's work to use the normal distribution to approximate a binomial random variable. It was finally proved by Aleksandr Lyapunov in 1901 [**?**]. George Pólya coined the term "central limit theorem," referring to it as central due to its importance in probability theory [**?**].

In the coming sections, we will introduce characteristic functions, which will be handy tools when proving the central limit theorem and its generalizations. After proving the theorem, we will provide applications such as the delta method, and Stirling's approximation. Finally some further topics such as dependency of random variables and rate of convergence will be discussed.

## 2 Convergence in Distribution

Throughout this paper, assume we are dealing with an underlying probability space $(\Omega, \mathcal{M})$, and $N$ will denote a normally distributed random variable, with $E(N) = 0, \text{Var}(N) = 1$

**Definition 2.1.** *Let $X$ be a random variable. Then $P_X$ is the **probability distribution** of $X$, and $F_X$ is the **distribution function** of $X$.*

**Definition 2.2.** *Let $P, P_1, P_2, \ldots$ be probability measures with distribution functions $F, F_1, F_2, \ldots$ respectively. $P_n$ converges to $P$ **weakly** or in **distribution** if and only if for all bounded continuous real functions $f$,*

$$\int_\Omega f dP_n \to \int_\Omega f dP.$$

*We say $F_n$ converges to $F$ **weakly** or in **distribution** if and only if $P_n$ does. Weak convergence in probability distribution and distribution function is denoted by respectively*

$$P_n \Rightarrow P, \quad F_n \Rightarrow F \quad as \quad n \to \infty$$

**Definition 2.3.** *Let $X, X_1, X_2, \ldots$ be random variables. We say that $X_n$ converge to $X$ **weakly**, or in **distribution**, if $P_{X_n}$ converge weakly to $P_X$. This will be denoted by*

$$X_n \Rightarrow X \quad as \quad n \to \infty$$

I will state some important theorems regarding convergence in distribution that I will take advantage of throughout this paper. Their proofs will be omitted, as they are standard results.

**Theorem 2.4.** *Let $X, X_1, X_2, \ldots$ be random variables. If $X_n \xrightarrow{P} X$, then $X_n \Rightarrow X$.*

**Theorem 2.5** (Skorokod's Representation Theorem)**.** *Suppose $P_n$ and $P$ are probability measures and $P_n \Rightarrow P$. Then there exist random variables $X_n, X$ such that $P_n = P_{X_n}$ and $P = P_X$, and $X_n(\omega) \to X(\omega)$ for all $\omega \in \Omega$.*

**Theorem 2.6** (Continuous mapping theorem)**.** *Let $X, X_1, X_2, \ldots$ be random variables, $g : \mathbb{R} \to \mathbb{R}$ be continuous. If $X_n \Rightarrow X$ then $g(X_n) \Rightarrow g(X)$.*

**Theorem 2.7** (Slutskys' Theorem)**.** *Let $X, X_1, X_2, \ldots$ and $Y, Y_1, Y_2, \ldots$ be random variables, $a \in \mathbb{R}$. Suppose*

$$X_n \Rightarrow X \quad and \quad Y_n \xrightarrow{P} a \quad as \quad n \to \infty$$

*Then*

*(a)* $X_n + Y_n \Rightarrow X + a \quad as \quad n \to \infty$

*(b)* $X_n Y_n \Rightarrow X \cdot a \quad as \quad n \to \infty$

## 3 Characteristic functions

Before we can prove the central limit theorem we first need to build some machinery. To do this, we will transform our random variable from the space of measure functions to the space of continuous complex values function via a Fourier transform, show the claim holds in the function space, and then invert back. In this section, the goal is to develop the properties that will allow us to achieve our goal. We will develop tools to allow us to show uniqueness and to deal with sequences of random variables. So we begin by defining the Fourier transform of a measure.

| Distribution | Notation | Characteristic Function |
|---|---|---|
| One Point | $\delta_a$ | $e^{ita}$ |
| Binomial | $\text{Bin}(n,p)$ | $(1-p+pe^{it})^n$ |
| Poisson | $\text{Pois}(m)$ | $e^{m(e^{it}-1)}$ |
| Uniform | $U(a,b)$ | $\frac{e^{itb}-e^{ita}}{it(b-a)}$ |
| Exponential | $\text{Exp}(\theta)$ | $(1-it\theta)^{-1}$ |
| Gamma | $\text{Gam}(k,\theta)$ | $(1-it\theta)^{-k}$ |
| Normal | $N(\mu,\sigma^2)$ | $e^{it\mu-\frac{1}{2}t^2\sigma^2}$ |
| Standard Normal | $N(0,1)$ | $e^{-\frac{t^2}{2}}$ |
| Cantor Distribution | | $\prod_{k=1}^{\infty}\cos(\frac{t}{3^k})$ |

Table 3.1: Example of some common characteristic functions [?]

**Definition 3.1.** *Let $X$ be a random variable, we define the **characteristic function** of $X$ by $\varphi_X : \mathbb{R} \to \mathbb{C}$ by*

$$\varphi_X(t) = E(e^{itX}) = E(\cos(tX)) + iE(\sin(tX))$$

**Remark 3.2.** *Aside from a negative sign in the exponential or the $2\pi$ factor, the characteristic function is the Fourier transform of the probability measure.*

Table 3.1 gives examples of some common characteristic functions. Note that these are standard distributions one would see in an elementary probability class, so their definitions are omitted.

## 3.1 Properties of Characteristic Functions

We begin our analysis of characteristic function by giving some elementary properties.

**Theorem 3.3.** [?] *Let $X, X_1, X_2$ be random variables and $a, b \in \mathbb{R}$. Then,*

(a) *$\varphi_X(t)$ exists, and $|\varphi_X(t)| \le \varphi_X(0) = 1$*

(b) *$\varphi_X(-t) = \varphi_{-X}(t) = \overline{\varphi_X(t)}$*

(c) *$\varphi_{aX+b}(t) = e^{ibt}\varphi_X(at)$*

(d) *If $X_1, X_2$ are independent then, $\varphi_{X_1+X_2} = \varphi_{X_1}\varphi_{X_2}$*

(e) *$\varphi_X(t)$ is uniformly continuous*

*Proof.* (a) $|\varphi_X(t)| \le E|e^{itX}| = E(1) = 1 = E(e^{i0X}) = \varphi_X(0)$. Hence $e^{itX}$ is integrable and $\varphi$ exists.

(b) Clear from definition of $\varphi$.

(c) $\varphi_{aX+b}(t) = E(e^{it(aX+b)}) = e^{itb}E(e^{i(at)X}) = e^{itb}\varphi_X(at)$.

(d) Let $Y_i = \cos(tX_i), Z_i = \sin(tX_i), i = 1, 2$. Then we have $\{Y_1, Z_1\}$ and $\{Y_2, Z_2\}$ are independent.

$$\begin{aligned}
\varphi_{X_1}(t)\varphi_{X_2}(t) &= (E[Y_1] + iE[Z_1])(E[Y_2] + iE[Z_2]) \\
&= E[Y_1]E[Y_2] - E[Z_1]E[Z_2] + i(E[Y_1]E[Z_2] + E[Z_1]E[Y_2]) \\
&= E[Y_1Y_2 - Z_1Z_2] + iE[Y_1Z_2 + Z_1Y_2] \\
&= E[\cos(tX_1 + tX_2)] + iE[\sin(tX_1 + tX_2)] \\
&= p_{X_1+X_2}(t)
\end{aligned}$$

(e)

$$\begin{aligned}|\varphi_X(t+h)-\varphi_X(t)| &= |E(e^{i(t+h)X}-e^{itX})| \\ &\le |e^{itX}|E(|e^{ihX}-1|) \\ &= E(|e^{ihX}-1|).\end{aligned}$$

Since $|e^{ihX}-1| \to 0$ as $h \to 0$, and is dominated by 2, we can apply dominated convergence theorem to get $E(|e^{ihX}-1|) \to 0$ as $h \to 0$. So for all $\epsilon > 0$, we can show $h$ to be sufficiently small so insure that $|\varphi_X(t+h)-\varphi_X(t)| < \epsilon$.

□

Properties (a),(e) are what make characteristic functions particularly nice. Characteristic functions allow us to view random variables as bounded, continuous, complex valued functions. The fact that they always exist with no restrictions on their domain makes them them more appealing than other similar transformations such as moment generating functions. Property (b) tells us that a characteristic function is real valued if and only if $X$ is symmetric about 0. In general, properties (c),(d) give us tools to compute a linear combination of random variables. In particular, if $X_1, X_2, \ldots$ are independent random variables and $a_k$ are real numbers, then

$$\varphi_{\sum_{k=1}^n a_k X_k} = \prod_{k=1}^n \varphi_{X_k}(a_k t). \tag{1}$$

**Lemma 3.4.** [?] *For $y \in \mathbb{R}, n \in \mathbb{N}$, we have*

$$\left| e^{iy} - \sum_{k=0}^n \frac{(iy)^k}{k!} \right| \le \min\left\{ \frac{2|y|^n}{n!}, \frac{|y|^{n+1}}{(n+1)!} \right\}. \tag{2}$$

I will not prove this since it is a standard result from complex analysis and is just a simple application of Taylor's theorem. We should pay close attention to the case where $n = 2$ and $y = tX$, where $X$ is a random variable

$$\left| e^{itX} - \left(1 + itX - \frac{1}{2}t^2X^2\right) \right| \le \min\left\{ |tX|^2, \frac{|tX^3|}{6} \right\}.$$

So after taking expectations we have

$$\left| \varphi_X(t) - \left(1 + itE(X) - \frac{1}{2}t^2E(X^2)\right) \right| \le E\left( \min\left\{ |tX|^2, \frac{|tX|^3}{6} \right\} \right). \tag{3}$$

This particular approximation will be useful when proving the central limit theorem.

At the end of the day, the characteristic functions do not have a meaning on their own; we use them as tools to deduce properties about $X$. This is possible since the distribution of $X$ is uniquely determined by its characteristic function.

**Theorem 3.5** (Inversion)**.** *If $X$ is a random variable and $a < b$ then*

$$P(a < X < b) + \frac{P(X=a)+P(X=b)}{2} = \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-itb}-e^{-ita}}{-it}\varphi(t)dt. \tag{4}$$

*Proof.* [?] We begin by fixing $T$. Since we have by Taylor theorem for complex variables

$$\begin{aligned}\Big|\int_{-T}^{T}\frac{e^{-itb}-e^{-ita}}{-it}\varphi(t)dt\Big| &\leq \int_{-T}^{T}\Big|\frac{e^{-itb}-e^{-ita}}{-it}\varphi(t)\Big|dt\\ &\leq \int_{-T}^{T}|e^{-ita}|\Big|\frac{e^{-it(b-a)}-1}{t}\Big|dt\\ &\leq \int_{-T}^{T}\frac{|t(b-a)|}{t}dt\\ &= 2T(b-a)\\ &\leq \infty.\end{aligned}$$

The second last inequality came from an application of lemma 3.4 where $n=0$. Hence we can apply Fubini's theorem.

$$\begin{aligned}I := \lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-itb}-e^{-ita}}{-it}\varphi(t)dt &= \lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-itb}-e^{-ita}}{-it}\int_{\Omega}e^{itX}dPdt\\ &= \lim_{T\to\infty}\frac{1}{\pi}\int_{-T}^{T}\int_{\Omega}\frac{e^{it(X-a)}-e^{it(X-b)}}{2it}dPdT\\ &= \lim_{T\to\infty}\frac{1}{\pi}\int_{\Omega}\int_{-T}^{T}\frac{e^{it(X-a)}-e^{it(X-b)}}{2it}dtdP\\ &= \lim_{T\to\infty}\frac{1}{\pi}\int_{\Omega}\int_{0}^{T}\frac{\sin(t(X-a))}{t}-\frac{\sin(t(X-b))}{t}dtdP.\end{aligned}$$

Since $|f(T)| := \big|\int_0^T\frac{\sin(x}{x}dx\big| \leq \big|\int_0^{\pi}\frac{\sin(x)}{x}dx\big| \leq \pi$, we can apply Lebesgue dominated convergence theorem. Using $\int_0^{\infty}\frac{\sin(x)}{x}dx=\frac{\pi}{2}$ we get,

$$\begin{aligned}I &= \int_{\Omega}\lim_{T\to\infty}\frac{1}{\pi}\int_{0}^{T}\frac{\sin(t(X-a))}{t}-\frac{\sin(t(X-b))}{t}dtdP\\ &= \int_{\Omega}g(X)dP,\end{aligned}$$

where

$$g(x)=\begin{cases}0 & \text{if } x<a\\ \frac{1}{2} & \text{if } x=a\\ 1 & \text{if } a<x<b\\ \frac{1}{2} & \text{if } x=b\\ 0 & \text{if } x>b\end{cases}$$

$$I = P(a<X<b)+\frac{P(X=a)+P(X=b)}{2} \tag{5}$$

□

**Corollary 3.6.** *If $X,Y$ are random variables such that $\varphi_X=\varphi_Y$, then $P_X=P_Y$ and $F_X=F_Y$.*

*Proof.* We have that $P_X([a,b])=P_Y([a,b])$ by the previous theorem, so the distribution of $X,Y$ agree on all the intervals. By the uniqueness part of Carathéodory's extension theorem, we have $P_X=P_Y$ or $F_X=F_Y$. □

**Example 3.7.** Suppose $X_1, \dots, X_n$ are identically distributed exponential random variables with parameter $\theta$, so they have the following distribution

$$F_{X_n}(x) = \begin{cases} 1 - e^{-\frac{x}{\theta}} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

Then after referring to Table 3.1 and using (1) we have

$$\varphi_{S_n} = \prod_{k=1}^{n} \varphi_{X_k} = \prod_{k=1}^{n} (1 - it\theta)^{-1} = (1 - it\theta)^{-n}$$

This is the same characteristic function as a gamma distributed random variable with parameters $\theta$ and $n$. Therefore by the uniqueness of the characteristic function, $S_n$ has the $\text{Gam}(\theta, n)$ distribution.

## 3.2 Continuity of Characteristic Functions

Our goal at the end of the day is to determine the distribution of $\frac{S_n}{n}$ as $n$ approaches infinity. So we need a better way to deal with sequences of random variables. It is natural to ask, "if we have a sequence of random variables $X_1, X_2, \dots$ such that their characteristic function converge, then do their distributions also converge?" The problem is that the limit of characteristic functions may not be a characteristic function. I will show that the desired result will follow if we have the added condition that the limit of characteristic functions is continuous at 0.

**Theorem 3.8** (Helly's Selection Theorem)**.** *For every sequence of $(F_n)_n$ of distribution functions, there exists a subsequence $(F_{n_k})_k$ and a non-decreasing right continuous $F$ such that $\lim_{k\to\infty} F_{n_k}(x) = F(x)$ at continuity points of $F$.*

This is a fairly standard result from analysis, and the proof is omitted. It can be shown relatively easily with an application of the diagonal method.

The $F$ resulting from the selection theorem is not necessarily a distribution function. For example, take $F_n$ as the distribution functions of $\delta_n$, the point mass probability measure at $n$. Then

$$F_n(x) = \begin{cases} 1 & \text{if } x \geq n \\ 0 & \text{if } x < n \end{cases}.$$

In this case $F(x) = 0$ for all $x$ and is clearly not valid distribution function. The issue in this example is that the probability masses are not "well behaved", in the sense there is no one bounded interval that contains majority of the probability for each distribution. For our purposes, we need a condition that will ensure we get a valid distribution. This leads to our next definition.

**Definition 3.9.** *A sequence of distributions functions $\{F_n\}_n$ is **tight** if $\forall \epsilon > 0$ there exist $a, b \in \mathbb{R}$ such that $F_n(a) < \epsilon, F_n(b) > 1 - \epsilon$. Equivalently, a sequence of probability measures $\{P_n\}_n$ is **tight** if $\forall \epsilon > 0$ there is a half open interval $(a, b]$ such that $P((a, b]) > 1 - \epsilon$. A sequence of random variables $\{X_n\}_n$ is **tight** if $\{P_{X_n}\}_n$ is tight.*

The term tightness refers to the fact that the mass of a family of distributions does not diverge.

**Example 3.10.** Let $\{a_n\}_n$ be a sequence of real numbers, and let $X_n$ are random variables with $P_{X_n} = \delta_{a_n}$. Then $X_n$ are tight if and only if $\{a_n\}_n$ is bounded.

**Theorem 3.11.** [?] *Let $\{P_n\}_n$ be probability measures, then $\{P_n\}_n$ are tight if and only if for every subsequence $\{P_{n_k}\}_k$, there exists a sub-subsequence $\{P_{n_{k(j)}}\}_j$ and probability measure $P$, such that $P_{n_{k(j)}} \Rightarrow P$ as $j \to \infty$.*

*Proof.* ($\Rightarrow$) Suppose $\{P_n\}_n$ are tight, let $F_n$ be the distribution functions of $P_n$. We can apply Helly's theorem to the subsequence $\{F_{n_k}\}_k$, so there exist a sub-subsequence $\{F_{n_{k(j)}}\}_j$, and right-continuous non-decreasing $F$ such that $\lim_{j\to\infty} F_{n_{k(j)}}(x) = F(x)$ at continuity points of $F$. Let $\epsilon > 0$, then there exists $a, b$ such that $F_n(a) > \epsilon$ and $F_n(b) > 1-\epsilon$ for all $n$. Thus we have $F(a) < \epsilon$, and $F(b) > 1-\epsilon$, so F is a valid probability. We denote the probability associated with $F$ by $P$. Thus we have $P_n \Rightarrow P$.

($\Leftarrow$) If $\{P_n\}_n$ are not tight, then there is some $\epsilon > 0$ such that for all $(a, b]$, $P_n((a, b]) < 1-\epsilon$ for some $n$. Lets pick a subsequence $P_{n_k}$ such that $P_{n_k}((-k, k]) \leq 1-\epsilon$. Suppose some sub-subsequence $\{P_{n_{k(j)}}\}$ were to converge weakly to some $P$, then we can pick $a, b$ such that $P((a, b]) > 1-\epsilon$, and $P(\{a\}) = P(\{b\}) = 0$. Then for a large enough $j$, $(a, b] \subset (-k(j), k(j)]$ so

$$1-\epsilon \geq P_{n_{k(j)}}((-k(j), k(j)]) \geq P_{n_{k(j)}}((a, b]) \to P((a, b]) < 1-\epsilon$$

This is a contradiction, hence we are done. □

**Corollary 3.12.** [?] *If $\{P_n\}_n$ is a tight sequence of probability measures, and if each subsequence that converges weakly at all converges to the probability measure $P$, then $P_n \Rightarrow P$.*

*Proof.* By theorem 3.11, each sub-subsequence $\{P_{n_{k(j)}}\}_j$converges weakly to the $P$ in the hypothesis. If $P_n$ does not weakly converge to $P$, then there exists an $x$ such that $P(\{x\}) = 0$ but $P_n(-\infty, x]$ does not converge to $P(-\infty, x]$. Then there is an $\epsilon > 0$ such that $|P(-\infty, x] - P_{n_k}(-\infty, x]| \geq \epsilon$ for any subsequence. Thus no sub-subsequence $P_{n_{k(j)}}$ can converge to $P$, this is a contadiction, hence $P_n \Rightarrow P$. □

**Theorem 3.13** (Continuity Theorem)**.** [?] *Let $X_1, X_2, \dots$ be random variables. If $\varphi_{X_n}(t) \to g(t)$ at each $t$ to some $g$ continuous at 0, then there exists a $X$ such that $X_n \Rightarrow X$ and $\varphi_X = g$.*

*Proof.* By Fubini's theorem,

$$\begin{aligned}
\frac{1}{u}\int_{-u}^{u}(1-\varphi_{X_n}(t))dt &= \int_\Omega \left[\frac{1}{u}\int_{-u}^{u}(1-e^{itX}dt\right] dP_{X_n} \\
&= 2\int_\Omega \left(1-\frac{\sin(ux)}{ux}\right) dP_{X_n} \\
&\geq 2\int_{|X|>2/u} \left(1-\frac{1}{ux}\right) dP_{X_n} \\
&\geq 2\int_{|X|\geq 2/u} \left(1-\frac{1}{2}\right) dP_{X_n} \\
&= P_{X_n}\left[|X| \geq \frac{2}{u}\right].
\end{aligned}$$

First note that the first integral is real. Since $g$ is continuous at 0 and $g(0) = 1$, for all $\epsilon > 0$ there is a $u$ such that $\frac{1}{u}\int_{-u}^{u}(1-g(t))dt < \epsilon/2$. Since $|1-\varphi_n(t)| \leq 2$ by dominated convegence theorem, there is a $n_0$ such that for all $n > n_0$, $\frac{1}{u}\int_{-u}^{u}(1-\varphi_{X_n}(t))dt < \epsilon$. Let $\alpha := 2/u$, then

$$P_{X_n}[|X| \geq \alpha] < \epsilon, \quad \forall n > n_0. \tag{6}$$

We can we increase $\alpha$ so that $P_{X_n}(|X| > \alpha) < \epsilon$ for all $n$, since there are only many $n \leq n_0$. So $P_{X_n}$ are tight.

Let $\{P_{X_{n_k}}\}_k$ be a subsequence such that $P_{X_{n_k}} \Rightarrow P$ as $k \to \infty$, for some probability measure $P$. By Skorokod's representation theorem, there exists a random variable $X$ such , $X_{n_k} \Rightarrow X$ as $k \to \infty$ and $P = P_X$. Since $e^{itX}$ is a bounded continuous function, by definition of weak convergence of probability measures we get

$$\int_\Omega e^{itX} dP_{X_{n_k}} \to \int_\Omega e^{itX} dP_X, \quad \text{as} \quad n \to \infty.$$

This is equivalent to $\lim_{k\to\infty} \varphi_{X_{n_k}}(t) = \varphi_X(t) = g(t)$. Therefore $g$ is a valid characteristic function for some random variable $Y$. Since $g = \varphi_X$, by the corollary to the inversion formula we get $P_Y = P_X = P$. Therefore any subsequence that converges weakly at all, converge to $P$. By corollary 3.12 we get $P_{X_n} \Rightarrow P$, or equivalently, $X_n \Rightarrow X$. □

# 4 Lindeberg-Lévy-Feller CLT

The classic central limit theorem requires the random variables to be identically distributed. However, the condition of identically distributed can be weakened as long as the random variables are "nice enough", which we will going into more detail.

**Definition 4.1.** *Let $\{r_n\}_n$ be a sequence of natural numbers, for each $n \in \mathbb{N}$, let $X_{n,1}, \ldots, X_{n,r_n}$ be random variables. Such a collection of random variables is called a **triangular array**. Given a triangular array we define $S_n = \sum_{k=1}^{r_n} X_{n,k}$.*

Note that every sequence of random variables $(X_n)_n$ can be viewed as a triangular array, in the case where $r_n = n$ and $X_{n,k} = X_k$. Triangular arrays can be useful as they give us a means to describe how a sequence of random variables could change with $n$.

**Theorem 4.2** (Lindeberg-Levy-Feller)**.** *Let $\{X_{n,k}\}_{n,k}$ be a triangular array and for each $n$, let $X_{n,1}, \ldots, X_{n,r_n}$ be independent. Denote*

$$E(X_{n,k}) = \mu_{n,k}, \mathrm{Var}(X_{n,k}) = \sigma^2_{n,k}, \mathrm{Var}(S_n) = s_n^2 = \sum_{k=1}^{r_n} \sigma^2_{n,k}.$$

*(Disregard the degenerate case where all the variances are 0). Further suppose*

$$\forall \epsilon > 0 \quad \lim_{n\to\infty} L_n^\epsilon := \frac{1}{s_n^2} \sum_{k=1}^{r_n} (|X_k - \mu_k|^2 I_{\{|X_k - \mu_l| > \epsilon s_n\}} \to 0, \tag{7}$$

*then we have*

$$\frac{1}{s_n} \sum_{k=1}^{r_n} (X_k - \mu_k) \Rightarrow N \tag{8}$$

We will refer to (7) as the Lindeberg condition.

Since we are trying to determine the behaviour of how $S_n/n$ deviates from the mean, we can assume WLOG that $\mu_{n,k} = 0$. In the case $\mu_{n,k} \neq 0$ one can replace $X_{n,k}$ with $X_{n,k} - \mu_{n,k}$ and the proof will remain the same.

This more general theorem is much stronger, since it implies the classical central limit theorem.

**Corollary 4.3** (Central Limit Theorem). *If $X, X_1, X_2, \ldots$ are iid random variables with $E(X) = 0, \mathrm{Var}(X) = \sigma^2 < \infty$ the $\frac{S_n}{\sqrt{n}\sigma} \Rightarrow N$.*

*Proof.* Let $r_n = n$, and $X_{n,k} = X_k$. Then the Lindeberg condition is reduced to the following

$$L_n^\epsilon = \frac{1}{s_n^2}\sum_{k=1}^{n} E(X_k^2 I_{\{X_k > \epsilon s_n\}}) = \frac{1}{\sigma^2}E(X^2 I_{\{X > \epsilon\sqrt{n}\sigma\}}) \to 0, \quad \text{as} \quad n \to \infty$$

□

Here is a quick example of the central limit theorem to demonstrate its power.

**Example 4.4.** Let $X_1, X_2, \ldots$ be iid Poison random variables with mean $\mu$, then $\frac{(X_n - n\mu)}{\sqrt{n}} \Rightarrow N$.

This example shows that the normal distribution can be used to approximate even discrete random variables.

## 4.1 Proof of CLT

Before we begin our proof, we will need a lemma

**Lemma 4.5.** *Let $w_1, z_1, \ldots, w_n, z_n \in \mathbb{C}$ where $|w_k|, |z_k| \leq 1\ \forall k$, then*

$$\left|\prod_{k=1}^{n} w_k - \prod_{k=1}^{n} z_k\right| \leq \sum_{k=1}^{n} |w_k - z_k|. \tag{9}$$

*Proof.* The claim is trivially true when $n = 1$. Suppose its true for $n = m$ then we have

$$\begin{aligned}\left|\prod_{k=1}^{m+1} w_k - \prod_{k=1}^{m+1} z_k\right| &= \left|(w_{m+1} - z_{m+1})\prod_{k=1}^{m} w_k + z_{m+1}\left(\prod_{k=1}^{m} w_k - \prod_{k=1}^{m} z_k\right)\right| \\ &\leq |w_{m+1} - z_{m+1}| + \left|\prod_{k=1}^{m} w_k - \prod_{k=1}^{m} z_k\right| \\ &\leq \sum_{k=1}^{m+1} |w_k - z_k|.\end{aligned}$$

By induction we are done. □

We can now prove theorem 4.2. The following proof took the ideas presented in [**?**] and applied them to the triangular array setting in [**?**].

*Proof.* By the continuity theorem, it is enough to show that $\varphi_{S_n/s_n}$ converges to $\varphi_N = e^{-t^2/2}$ point wise. This is the same as showing

$$|\varphi_{S_n/s_n}(t) - e^{-t^2/2}| \to 0 \qquad \text{as} \qquad n \to \infty.$$

We begin by noting that

$$\varphi_{S_n/s_n} = \prod_{k=1}^{r_n} \varphi_{X_{n,k}}(t/s_n) \quad \text{and} \quad e^{-t^2/2} = \prod_{k=1}^{r_n} \exp\left\{\frac{\sigma_{n,k}^2 t^2}{2s_n^2}\right\}.$$

by definition of $s_n$. Hence we need to show that

$$\left| \prod_{k=1}^{r_n} \varphi_{X_{n,k}}(t/s_n) - \prod_{k=1}^{r_n} \exp\left\{ \frac{\sigma_{n,k}^2 t^2}{2s_n^2} \right\} \right| \to 0 \quad \text{as} \quad n \to \infty.$$

However, by an application of the lemma 4.5 and triangle inequality we have

$$\begin{aligned}
&\left| \prod_{k=1}^{r_n} \varphi_{X_{n,k}}(t/s_n) - \prod_{k=1}^{r_n} \exp\left\{ \frac{\sigma_{n,k}^2 t^2}{2s_n^2} \right\} \right| \\
&\le \sum_{k=1}^{r_n} \left| \varphi_{X_{n,k}}(t/s_n) - \exp\left\{ -\frac{\sigma_{n,k}^2 t^2}{2s_n^2} \right\} \right| \\
&\le \sum_{k=1}^{r_n} \left| \varphi_{X_{n,k}}(t/s_n) - \left( 1 - \frac{\sigma_{n,k}^2 t^2}{2s_n^2} \right) \right| \\
&+ \sum_{k=1}^{r_n} \left| \exp\left\{ -\frac{\sigma_{n,k}^2 t^2}{2s_n^2} \right\} - \left( 1 - \frac{\sigma_{n,k}^2 t^2}{2s_n^2} \right) \right|.
\end{aligned}$$

So we are done if we can show that

$$\sum_{k=1}^{r_n} \left| \varphi_{X_{n,k}}(t/s_n) - \left( 1 - \frac{\sigma_{n,k}^2 t^2}{2s_n^2} \right) \right| \to 0 \quad \text{as} \quad n \to \infty, \tag{10}$$

$$\sum_{k=1}^{r_n} \left| \exp\left\{ -\frac{\sigma_{n,k}^2 t^2}{2s_n^2} \right\} - \left( 1 - \frac{\sigma_{n,k}^2 t^2}{2s_n^2} \right) \right| \to 0 \quad \text{as} \quad n \to \infty. \tag{11}$$

Note that the (11) is a special case of the (10), in the case where $X_{n,k}$ are normally distributed with mean 0 and variance $\frac{\sigma_{n,k}^2}{s_n^2}$. So if we can show the first claim, we can apply the identical reasoning to get the second one.

Let $\epsilon > 0$, then we have

$$\begin{aligned}
&\sum_{k=1}^{r_n} \left| \varphi_{X_{n,k}}(t/s_n) - \left( 1 - \frac{\sigma_{n,k}^2 t^2}{2s_n^2} \right) \right| \\
&\le \sum_{k=1}^{r_n} E\left( \min\left\{ \frac{t^2 X_{n,k}^2}{s^2}, \frac{|t|^3 |X_{n,k}|^3}{6s_n^3} \right\} \right) \\
&\le \sum_{k=1}^{r_n} E\left( \frac{|t|^3 |X_{n,k}|^3}{6s_n^3} I_{\{|X_{n,k}| \le \epsilon s_n\}} \right) + \sum_{k=1}^{r_n} E\left( \frac{|t|^2 X_{n,k}^2}{s_n^2} I_{\{|X_{n,k}| > \epsilon s_n\}} \right) \\
&= \frac{|t|^3}{6s_n^3} \sum_{k=1}^{r_n} E\left( |X_{n,k}|^3 I_{\{|X_{n,k}| \le \epsilon s_n\}} \right) + t^2 L_n^\epsilon \\
&\le \frac{|t|^3}{6s_n^3} \sum_{k=1}^{r_n} E\left( \epsilon s_n |X_{n,k}|^2 I_{\{|X_{n,k}| \le \epsilon s_n\}} \right) + t^2 L_n^\epsilon \\
&\le \frac{|t|^3 \epsilon}{6s_n^2} \sum_{k=1}^{r_k} E\left( |X_{n,k}|^2 \right) + t^2 L_n^\epsilon \\
&= \frac{|t|^3 \epsilon}{6} + t^2 L_n^\epsilon
\end{aligned}$$

By the Lindeberg condition we get $L_n^\epsilon \to 0$, therefore

$$\limsup_{n\to\infty} \sum_{k=1}^{n} \left| \varphi_{X_k}(t/s_n) - \left(1 - \frac{\sigma_k^2 t^2}{2s_n^2}\right) \right| \leq \frac{|t|^3 \epsilon}{6}$$

since $\epsilon$ is arbitrary, we are done. □

## 4.2 Lyapunov Condition

The Lindeberg although powerful can be difficult to show. There is an alternative condition one can show to get the same result. The Lyapunov condition offers a sufficient condition that is easier to verify.

**Theorem 4.6** (Lyapunov CLT)**.** *Let $\{X_{n,k}\}_{n,k}$ be a triangular array. Suppose there is some $\delta > 0$ such that $|X_{n,k}|^{2+\delta}$ are integrable for all $n \in \mathbb{N}, 1 \leq k \leq r_n$. Further suppose that*

$$\lim_{n\to\infty} \frac{1}{s_n^{2+\delta}} \sum_{k=1}^{r_n} E(|X_{n,k}|^{2+\delta}) = 0. \tag{12}$$

*Then $\frac{S_n}{s_n} \Rightarrow N$.*

*Proof.* It is enough to show that the Lindeberg condition holds.

$$\begin{aligned}
\lim_{n\to\infty} L_n^\epsilon &= \lim_{n\to\infty} \frac{1}{s_n^2} \sum_{k=1}^{r_n} E\left(|X_{n,k}|^2 I_{\{|X_{n,k}|>\epsilon s_n\}}\right) \\
&\leq \lim_{n\to\infty} \frac{1}{s_n^2} \sum_{k=1}^{r_n} E\left(\frac{|X_{n,k}|^{2+\delta}}{(\epsilon s_n)^\delta} I_{\{|X_{n,k}|>\epsilon s_n\}}\right) \\
&\leq \lim_{n\to\infty} \frac{1}{\epsilon^\delta s_n^{2+\delta}} \sum_{k=1}^{r_n} E\left(|X_{n,k}|^{2+\delta} I_{\{|X_{n,k}|>\epsilon s_n\}}\right) \\
&\leq \lim_{n\to\infty} \frac{1}{\epsilon^\delta s_n^{2+\delta}} \sum_{k=1}^{r_n} E\left(|X_{n,k}|^{2+\delta}\right) \\
&= 0.
\end{aligned}$$

Thus by Lindeberg-Lévy-Feller, $\frac{S_n}{s_n} \Rightarrow N$. □

**Example 4.7.** Suppose that $X_1, X_2, \ldots$ are independent random variables such that $E(X_n) = 0$ and are uniformly bounded, as in there exists by some constant $K$ such that $|X_n| \leq K, \forall n \in \mathbb{N}$. Also suppose that $s_n^2 = \sum_{k=1}^n E((X_k)^2) \to \infty$ as $n \to \infty$. Then we have $S_n/n \Rightarrow N$.

*Proof.* We will show this by applying Lyapunov condition for $\delta = 1$.

$$\frac{1}{s_n^3} \sum_{k=1}^{n} E(|X_k|^3) \leq \frac{KE(|X_k|^2}{s_n^3} = \frac{K}{s_n} \to 0 \quad \text{as} \quad n \to \infty.$$

□

The Lyapunov condition is not as strong as the Lindeberg condition, or in some cases even the classical one. There exist random variables where the variance is finite, but higher-order moments are not. The following example demonstrates this:

**Example 4.8.** Let $X_1, X_2, \ldots$ be iid random variables with density

$$f(x) = \begin{cases} \frac{c}{|x|^3(\log|x|)^2} & \text{if } |x| > 2 \\ 0 & \text{otherwise} \end{cases}$$

where $c$ is a normalizing constant. It is trivial to verify that $f$ is a valid density, with $E(X_n) = 0$ and $E(X_n^2) = 2c\int_2^\infty \frac{dx}{x(\log x)^2} < \infty$, but for all $\delta > 0, E(|X_n|^{2+\delta}) = \infty$. So the Lyapunov central limit theorem is inconclusive. However, $X_n$ satisfies the conditions of the classical central limit theorem, so $\frac{S_n}{s_n} \Rightarrow N$.

# 5 Applications

## 5.1 Delta-Method

In practice the true strength of the central limit theorem comes in statistics where the sequence of random variables often represents observed data. It is fair to ask how the limiting behaviour changes if the data is transformed by some $g : \mathbb{R} \to \mathbb{R}$. If $g$ is "nice enough" we can determine the limiting behaviour of $g\left(\frac{S_n}{n}\right)$. Let $X_1, X_2, \ldots$ be iid random variables with $E(X_n) = \mu, \mathrm{Var}(X_n) = \sigma^2$. The delta method allows us to determine the behaviour of $g\left(\frac{S_n}{n}\right)$ as $n \to \infty$.

When $g$ is continuous, by strong law of large numbers we get $g\left(\frac{S_n}{n}\right) \overset{a.s.}{\to} g(\mu)$. To determine the distribution of $g\left(\frac{S_n}{n}\right)$ as $n \to \infty$ we require a bit more structure on $g$. The following result is a generalization of the one presented in [**?**].

**Theorem 5.1** (Delta Method)**.** *Let $X_1, X_2, \ldots$ be iid random variables with $E(X_n) = \mu, \mathrm{Var}(X_n) = \sigma^2, g : \mathbb{R} \to \mathbb{R}$, $N$ is a random variable with the standard normal distribution. If $g$ has derivative up to order $m$, $g^{(m)}$ is continuous at $\mu$ and $g^{(i)}(\mu) = 0$ for $1 \le i \le m-1$, then we have*

$$n^{\frac{m}{2}}\left(g\left(\frac{S_n}{n}\right) - g(\mu)\right) \Rightarrow \frac{g^{(m)}\sigma^m}{m!}N^m \tag{13}$$

*Proof.* by Taylor's theorem, there exist random variables $D_n$ such that $\left|D_n - \frac{S_n}{n}\right| \le \left|\mu - \frac{S_n}{n}\right|$ and

$$\begin{aligned} g\left(\frac{S_n}{n}\right) - g(\mu) &= \sum_{k=1}^{m-1} \frac{g^{(k)}}{k!}(\mu)\left(\frac{S_n}{n} - \mu\right)^k + \frac{g^{(m)}(D_n)}{m!}\left(\frac{S_n}{n} - \mu\right)^m \\ &= \frac{g^{(m)}(D_n)}{m!}\left(\frac{S_n}{n} - \mu\right)^m \end{aligned}$$

We can rewrite the above equality by

$$n^{\frac{m}{2}}\left(g\left(\frac{S_n}{n}\right) - g(\mu)\right) = \frac{g^{(m)}(D_n)\sigma^m}{m!}\left(\frac{S_n - n\mu}{\sqrt{n}\sigma}\right)^m \tag{14}$$

By strong law of large numbers we have $\frac{S_n}{n} \overset{a.s.}{\to} \mu$, and since $g^{(m)}$ is continuous at $\mu$, we have $g(m)(D_n) \overset{a.s.}{\to} g^{(m)}(\mu)$. Since $\frac{S_n - n\mu}{\sqrt{n}\sigma} \Rightarrow N$ and $x^m$ is continuous, by continuous mapping theorem we have $\left(\frac{S_n - n\mu}{\sqrt{n}\sigma}\right)^m \Rightarrow N^m$. Finally, by application of Slutsky's theorem we get

$$n^{\frac{m}{2}}\left(g\left(\frac{S_n}{n}\right) - g(\mu)\right) \Rightarrow \frac{g^{(m)}(\mu)\sigma^m}{m!}N^m \quad \text{as} \quad n \to \infty \tag{15}$$

□

## 5.2 Stirling's Formula

Another application of the central limit theorem is that it allows us to approximate numerical quantities using probabilistic methods. For example, we can prove Stirling's formula, which approximates the value of $n!$ for large $n$.

**Theorem 5.2** (Moment Convergence Theorem)**.** *Let* $\{X_n\}_n$ *be a sequence of random variables such that* $X_n \Rightarrow X$ *for some random variable* $X$*. If* $\limsup_{n\to\infty} X_n^2 < \infty$*, then*

$$\lim_{n\to\infty} E(|X_n|^r) = E(|X|^r) \quad \textit{for} \quad 0 \le r \le 2. \tag{16}$$

The proof of the moment convergence theorem follows routine measure theory arguments and is omitted. Suppose you have iid random variables $X_n$ with $E(X_n) = 0, E(X_n^2) = \sigma^2$. Since $X_n$ satisfy the conditions of the central limit theorem and the moment convergence theorem we get

$$\lim_{n\to\infty} E\left(\left|\frac{S_n}{\sqrt{n}\sigma}\right|\right) = E(|N| = \sqrt{\frac{2}{\pi}}. \tag{17}$$

We now prove the Stirling's Formula.

**Theorem 5.3** (Stirling's Formula)**.**

$$\lim_{n\to\infty} \frac{\sqrt{2n\pi}n^n e^{-n}}{n!} = 1 \tag{18}$$

*Proof.* [**?**] Let $\{X_n\}_n$ be independent, identically distributed exponentially random variables with $\theta = 1$. We define $Y_n := X_n - 1, S_n = \sum_{k=1}^n X_k$, it is easily verified that $E(Y) = 0, E(Y^2) = 1$. It then follows that

$$\lim_{n\to\infty} E\left(\frac{|S_n - n|}{\sqrt{n}}\right) = \sqrt{\frac{2}{\pi}} \tag{19}$$

From example 3.7, we know that $S_n$ has the $\text{Gam}(1, n)$ distribution. In other words, $S_n$ has distribution function

$$F_{S_n}(x) = \begin{cases} \int_0^x \frac{1}{(n-1)!} t^{n-1} e^{-t} dt & \text{if } x \ge 0 \\ 0 & \text{if } x < 0 \end{cases}$$

Therefore

$$E(|S_n - n|) = \frac{1}{(n-1)!} \int_0^\infty |t - n| t^{n-1} e^{-t} dt$$

This is an easy integral to compute, and can be done by elementary calculus methods. The result is

$$E(|S_n - n|) = \frac{2n n^n e^{-n}}{n!} \tag{20}$$

Combining (19) and (20) we get

$$\lim_{n\to\infty} \frac{2n n^n e^{-n}}{\sqrt{n} n!} = \sqrt{\frac{2}{\pi}},$$

or equivalently,

$$\lim_{n\to\infty} \frac{\sqrt{2n\pi}n^n e^{-n}}{n!} = 1 \tag{21}$$

□

# 6 Further topics

Is it possible to weaken or remove the condition of independence in the central limit theorem? What is the the rate at which the sample mean of random variables converge to $N$? In this section I will provide a brief discussion of these complex topics.

## 6.1 Dependency

The classical central limit theorem assumes identically distributed random variables with finite variance. The Lindeberg-Lévy-Feller central limit theorem showed that we can weaken the condition of identically distributed random variables so long as they satisfy the Lindeberg condition. It is natural to ask what happens if instead of weakening the identically distributed hypothesis, we weaken dependence.

Janson showed that you cannot even weaken the condition for pairwise independent random variables. There exist identically distributed random variables $X_1, X_2, \ldots$ with $E(X_n) = 0, 0 < E(X_n^2) = \sigma^2 < \infty$ such that

$$\frac{S_n}{\sqrt{n}\sigma} \Rightarrow V, \quad \text{as} \quad n \to \infty,$$

where $V$ is a non-normal, non-degenerate distribution. [**?**]

One can show that if you have a sequence of random variables $X_1, X_2, \ldots$ that are "almost" independent then the central limit theorem holds. We will define what "almost" means.

**Definition 6.1.** *Let $X_1, X_2, \ldots$ be random variable. We say $\{X_n\}_n$ is **$\alpha$-mixing** if there is a sequence of non-negative numbers $\alpha_n$ such that $\alpha_n \to 0$ as $n \to \infty$, and for all $A \in \sigma(X_1, \ldots, X_k), B \in \sigma(X_{k+n}, X_{k+n+1}, \ldots), n \geq 1$ we have*

$$|P(A \cap B) - P(A)P(B)| \leq \alpha_n. \tag{22}$$

*If the distribution of the random vector $(X_n, X_{n+1}, \ldots, X_{n+j})$ does not depend on $n$, then the sequence is called **stationary**.*

**Theorem 6.2.** [**?**] *Suppose $X_1, X_2, \ldots$ is stationary and $\alpha$-mixing with $\alpha_n = O(n^{-5})$ and $E(X_n) = 0, E(X_n^{12}) < \infty$. Then*

$$\frac{Var(S_n)}{n} \to \sigma^2 = E(X_1^2) + 2\sum_{k=1}^{\infty} E(X_1 E_k), \tag{23}$$

*where the series converges absolutely. If $\sigma > 0$, then $\frac{S_n}{\sqrt{n}\sigma} \Rightarrow N$.*

The conditions of $\alpha_n = O(n^{-5})$ and $E(X_n^{12}) < \infty$ are much stronger than necessary but are stated to simplify the proof [**?**]. Even with these implications, the proof is rather long and subtle.

## 6.2 Berry-Esséen Theorem

Suppose $X_1, X_2, \ldots$ are iid random variables with $E(X_n) = \mu, \mathrm{Var}(X_n) = \sigma^2$. The strong law of large numbers tells us that $\frac{S_n}{n} \to \mu$, and the central limit theorem tells us that $\frac{S_n - n\mu}{\sqrt{n}\sigma} \Rightarrow N$ as $n \to \infty$. Therefore, we have that for large $n$, $\frac{S_n}{n}$ is approximately normally distributed with $E\left(\frac{S_n}{n}\right) = \mu$ and $\mathrm{Var}\left(\frac{S_n}{n}\right) \approx \frac{\sigma^2}{n}$. Thus the central limit theorem provides us with information on the rate at which the strong law of large numbers holds. Another concern we have is: exactly how fast is the rate of convergence in distribution for the central limit theorem?

Both Berry and Essées answered this question independently during the World War II when they came up with the following result.

**Theorem 6.3.** [?] [?] *Let* $X_1, X_2, \ldots$ *be iid random variables such that* $E(X_n) = \mu, \mathrm{Var}(X_n) = \sigma^2, E(|X|^3) = \gamma^3 < \infty$. *Then*

$$\sup_x |F_{\frac{S_n - n\mu}{\sqrt{n}\sigma}}(x) - F_N(x)| \leq C \frac{\gamma^3}{\sigma^3 \sqrt{n}} \tag{24}$$

$C$ *is a numerical constant.*

It has been shown that $0.4097 \leq C \leq 0.7655$. This is remarkable since it tells us that the distribution functions converge at a rate of $O(\frac{\gamma^3}{\sqrt{n}})$. The proof of this result is, again, very long, and is omitted.

# 7 Conclusion

The classical central limit theorem tells use how the sample mean of iid random variables $X_1, X_2, \ldots$ deviates from their expected value. In particular, the distribution of the sample mean approaches a normal distribution. Our primary tool to show this were characteristic functions. Characteristic functions allow us to represent the distributions of random variables as uniformly continuous bounded complex-valued functions. We also showed that the representation is unique, and respect sequences given the limit is continuous at the origin.

We then introduced and proved Lindeberg-Lévy-Feller theorem, which generalizes the classical central limit theorem. It states that you can still get convergence in distribution to the normal distribution, even when the $X_n$ are not identically distributed, as long as the Lindeberg condition is satisfied. After than we state the Lyapunov condition, a weaker condition that ensures the Lindeberg condition is satisfied. Finally we discussed applications, and advanced topics related to dependence and rate of convergence.