The Central Limit Theorem

1 The Central Limit Theorem

Theorem 1.1 (The Central Limit Theorem)
*Let $\{ X_n : (\Omega, \mathcal{A}, \mathfrak{m}) \longrightarrow \mathbb{R} \}_{n\in\mathbb{N}}$ be a sequence of independent identically distributed (*iid*) $\mathbb{R}$-valued random variables whose common variance σ^2 is positive, i.e. $\sigma^2 \in (0,\infty)$. Let $\mu \in \mathbb{R}$ be the common mean of the X_n's. For each $n \in \mathbb{N}$, define*

$$Y_n \ := \ \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \ : \ (\Omega, \mathcal{A}, \mathfrak{m}) \ \longrightarrow \ \mathbb{R}.$$

Then, the sequence $\{ P_{Y_n} \}_{n\in\mathbb{N}}$ of distribution measures of the random variables Y_n converges weakly to the standard normal disctribution measure $P_{N(0,1)}$ on $\mathbb{R}$. In other words, the sequence of probability density functions (defined on $\mathbb{R} = \mathrm{codomain}(X_n)$) induced by Y_n converges weakly to the probability density function (also defined on $\mathbb{R}$) of the standard normal distribution.

Remark 1.2

- ***Note that the Central Limit Theorem (Theorem 1.1) does NOT say that the sequence $\{ Y_n \}_{n\in\mathbb{N}}$ of random variables converges to anything at all.***

  Instead, it is a statement about the (weak) convergence of the sequence $\{ P_{Y_n} \}_{n\in\mathbb{N}}$ of distribution measures of $\{ Y_n \}_{n\in\mathbb{N}}$. Recall that each distribution measure P_{Y_n} is a probability measure defined on the Lebesgue measure space $(\mathbb{R}, \mu_{\mathrm{Lebesgue}})$.

- *Recall that a <u>random variable</u> is simply a measurable function $X : (\Omega, \mathcal{A}, \mathfrak{m}) \longrightarrow (E, \mathcal{B})$, whose domain $(\Omega, \mathcal{A}, \mathfrak{m})$ is a probability space, and whose codomain $(E, \mathcal{B})$ is a measure space.*

  The <u>distribution measure</u> P_X of a random variable $X : (\Omega, \mathcal{A}, \mathfrak{m}) \longrightarrow (E, \mathcal{B})$ is, by definition, the probability measure P_X defined on $(E, \mathfrak{B})$ as follows: For each measurable $U \subset E$ (i.e. the subset $U \subset E$ is a member of the σ-algebra $\mathcal{B}$),

  $$P_X(U) \ := \ \mathfrak{m}\big(X^{-1}(U)\big) \ = \ \mathfrak{m}(\{ x \in \Omega \mid X(x) \in U \}).$$

- *In particular, the distribution measure of each X_n or Y_n as in Theorem 1.1 is a probability measure on $\mathbb{R} = \mathrm{codomain}(X_n) = \mathrm{codomain}(Y_n)$.*

- *Recall that the standard normal distribution measure $P_{N(0,1)}$ on $\mathbb{R}$ is given by:*

  $$P_{N(0,1)}(U) \ := \ \int_U \frac{1}{\sqrt{2\pi}}\, e^{-z^2/2}\, \mathrm{d}z, \ \ \textit{for each Lebesgue-measurable } U \subset \mathbb{R}.$$

  Equivalently, the probability density function $f_{N(0,1)}(z)$ of the standard normal distribution is the standard Gaussian function:

  $$f_{N(0,1)}(z) \ := \ \frac{1}{\sqrt{2\pi}}\, e^{-z^2/2}, \ \ \textit{for } z \in \mathbb{R}.$$

  Equivalently, the cumulative distribution function $F_{N(0,1)}(z)$ of the standard normal distribution is given by:

  $$F_{N(0,1)}(z) \ := \ P_{N(0,1)}(\,(-\infty, z)\,) \ = \ \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}}\, e^{-z^2/2}\, \mathrm{d}z, \ \ \textit{for } z \in \mathbb{R}.$$

- ***Definition.*** *A sequence $\{\mu_n\}_{n\in\mathbb{N}}$ of probability measures on $\mathbb{R}^d$, $d \geq 1$, is said to <u>converge weakly</u> to a probability measure μ on $\mathbb{R}^d$ if μ_n converges weakly to μ as linear functionals on the vector space $C^0(\mathbb{R}^d, \mathbb{R}) \bigcap L^\infty(\mathbb{R}^d, \mathbb{R})$, i.e.*

  $$\langle\, \mu_n \,,\, f \,\rangle := \int_{\mathbb{R}^d} f(x)\, \mu_n(\mathrm{d}x) \ \longrightarrow \ \langle\, \mu \,,\, f \,\rangle := \int_{\mathbb{R}^d} f(x)\, \mu(\mathrm{d}x), \ \ \textit{as } n \longrightarrow \infty,$$

  for each $f \in C^0(\mathbb{R}^d, \mathbb{R}) \bigcap L^\infty(\mathbb{R}^d, \mathbb{R})$.

  Here, $C^0(\mathbb{R}^d, \mathbb{R})$ denotes the vector space of all continuous $\mathbb{R}$-valued functions defined on $\mathbb{R}^d$, and $L^\infty(\mathbb{R}^d, \mathbb{R})$ denotes the vector space of all bounded $\mathbb{R}$-valued functions defined on $\mathbb{R}^d$. Hence, $C^0(\mathbb{R}^d, \mathbb{R}) \bigcap L^\infty(\mathbb{R}^d, \mathbb{R})$ is the vector space of all bounded continuous $\mathbb{R}$-valued functions defined on $\mathbb{R}^d$.

Outline of proof of Theorem 1.1

CLAIM: $\widehat{P}_{Y_n}$ converges pointwise to $\widehat{P}_{N(0,1)}$, as $n \longrightarrow \infty$.

Here, $\widehat{P}_{Y_n} : \mathbb{R} \longrightarrow \mathbb{C}$ is the Fourier transform of the distribution measure P_{Y_n} of the random variable Y_n, and $\widehat{P}_{N(0,1)} : \mathbb{R} \longrightarrow \mathbb{C}$ is the Fourier transform of the standard normal distribution measure $P_{N(0,1)}$.

By Levy's Continuity Theorem C.3 and the continuity of $\widehat{P}_{N(0,1)}(\theta) = e^{-\theta^2/2}$ at $\theta = 0 \in \mathbb{R}$, the CLAIM implies that there exists a probability measure ν on $\mathbb{R}$ such that $\widehat{\nu} = \widehat{P}_{N(0,1)}$, as $\mathbb{C}$-valued functions on $\mathbb{R}$, and that P_{Y_n} converges weakly to ν. By the injectivity of the Fourier transform on the set of probability measures on $\mathbb{R}$ (Theorem C.1), we must have $\nu = P_{N(0,1)}$. Therefore, granting the validity of the CLAIM, the proof of Theorem 1.1 is complete. $\square$

Proof of Theorem 1.1. Given the Outline above, in order to complete the proof of Theorem 1.1, it remains only to establish the CLAIM in the Outline.

We will show that, for each $\theta \in \mathbb{R}$, $\widehat{P}_{Y_n}(\theta) \longrightarrow \widehat{P}_{N(0,1)}(\theta) = \exp\left(-\dfrac{\theta^2}{2}\right)$, as $n \longrightarrow \infty$.

$$\widehat{P}_{Y_n}(\theta) \;=\; \widehat{P}_{\frac{1}{\sigma\sqrt{n}}\sum\limits_{i=1}^{n}(X_i-\mu)}(\theta) \;=\; \widehat{P}_{\sum\limits_{i=1}^{n}\frac{1}{\sigma}(X_i-\mu)}\left(\frac{\theta}{\sqrt{n}}\right) \;=\; \prod_{i=1}^{n}\widehat{P}_{\frac{1}{\sigma}(X_i-\mu)}\left(\frac{\theta}{\sqrt{n}}\right) \;=\; \prod_{i=1}^{n}\psi\left(\frac{\theta}{\sqrt{n}}\right) \;=\; \left(\psi\left(\frac{\theta}{\sqrt{n}}\right)\right)^n,$$

where $\psi(\theta) := \widehat{P}_{\frac{1}{\sigma}(X_i-\mu)}(\theta)$. The function $\psi(\theta)$ is well-defined (i.e. independent of i) because the X_i's are identically distributed. The second equality follows from Theorem A.5, whereas the third equality follows from Theorem A.6 and the independence hypothesis on the X_i's.

By hypothesis on the X_i's, $E\left\{\dfrac{X_i-\mu}{\sigma}\right\} = 0$, and $E\left\{\left(\dfrac{X_i-\mu}{\sigma}\right)^2\right\} = 1$. Corollary A.8 thus implies

$$\begin{aligned}
E\left\{\frac{X_i-\mu}{\sigma}\right\} = 0 < \infty \quad &\Longrightarrow \quad \psi'(0) \;=\; \mathbf{i}\,E\left\{\frac{X_i-\mu}{\sigma}\right\} \;=\; 0,\\
E\left\{\left(\frac{X_i-\mu}{\sigma}\right)^2\right\} = 1 < \infty \quad &\Longrightarrow \quad \psi''(0) \;=\; -\,E\left\{\left(\frac{X_i-\mu}{\sigma}\right)^2\right\} \;=\; -1.
\end{aligned}$$

By Taylor's Theorem,

$$\psi(\theta) \;=\; \psi(0) + \psi'(0)\theta + \frac{1}{2}\psi''(r)\theta^2 \;=\; 1 + \frac{\theta^2}{2}\psi''(r), \;\; \text{for some } \; |r| < |\theta|.$$

Thus,

$$\begin{aligned}
\lim_{n\to\infty}\left(\psi\left(\frac{\theta}{\sqrt{n}}\right)\right)^n \;&=\; \lim_{n\to\infty}\left(1 + \frac{\theta^2}{2n}\psi''(s)\right)^n, \quad \text{where} \quad |s| < \frac{|\theta|}{\sqrt{n}}\\
&=\; \exp\left\{\log\left[\lim_{n\to\infty}\left(1 + \frac{\theta^2}{2n}\psi''(s)\right)^n\right]\right\} \;=\; \exp\left\{\lim_{n\to\infty}\left[n\,\log\left(1 + \frac{\theta^2}{2n}\psi''(s)\right)\right]\right\}\\
&=\; \exp\left\{\lim_{n\to\infty}\left[n\cdot\frac{\theta^2}{2n}\psi''(s)\cdot\frac{\log\left(1 + \frac{\theta^2}{2n}\psi''(s)\right) - \log(1)}{\frac{\theta^2}{2n}\psi''(s)}\right]\right\}\\
&=\; \exp\left\{\frac{\theta^2}{2}\cdot\left(\lim_{n\to\infty}\psi''(s)\right)\cdot\left(\lim_{n\to\infty}\frac{\log\left(1 + \frac{\theta^2}{2n}\psi''(s)\right) - \log(1)}{\frac{\theta^2}{2n}\psi''(s)}\right)\right\}
\end{aligned}$$

Now, $s \to 0$ as $n \to \infty$, since $|s| < \dfrac{|\theta|}{\sqrt{t}}$. By Theorem A.7 and the hypothesis that $E\left\{\left(\dfrac{X_i - \mu}{\sigma}\right)^2\right\} = 1$, we see that ψ'' is continuous; furthermore, Corollary A.8 implies

$$\lim_{n\to\infty} \psi''(s) \;=\; \psi''(0) \;=\; -1.$$

Now that we know $\displaystyle\lim_{n\to\infty} \psi''(s)$ is finite, we immediately see that

$$\lim_{n\to\infty} \frac{\log\left(1 + \dfrac{\theta^2}{2n}\psi''(s)\right) - \log(1)}{\dfrac{\theta^2}{2n}\psi''(s)} \;=\; \lim_{h\to 0} \frac{\log(1+h) - \log(1)}{h} \;=\; \frac{\mathrm{d}}{\mathrm{d}x}\log(x)\bigg|_{x=1} \;=\; \frac{1}{x}\bigg|_{x=1} \;=\; 1.$$

Hence, we may now deduce that

$$\begin{aligned}
\lim_{n\to\infty}\left(\psi\left(\frac{\theta}{\sqrt{n}}\right)\right)^n \;&=\; \exp\left\{\frac{\theta^2}{2}\cdot\left(\lim_{n\to\infty}\psi''(s)\right)\cdot\left(\lim_{n\to\infty}\frac{\log\left(1+\dfrac{\theta^2}{2n}\psi''(s)\right) - \log(1)}{\dfrac{\theta^2}{2n}\psi''(s)}\right)\right\} \\
&=\; \exp\left\{\frac{\theta^2}{2}\cdot(-1)\cdot(1)\right\} \;=\; \exp\left\{-\frac{\theta^2}{2}\right\} \\
&=\; \widehat{P}_{N(0,1)}(\theta)
\end{aligned}$$

□

A The Fourier transform of a probability measure on $\mathbb{R}$ and the characteristic function of an $\mathbb{R}$-valued random variable

Definition A.1 (Fourier transform of a probability measure on $\mathbb{R}$)
The Fourier transform $\widehat{\mu}$ of a probability measure μ on $\mathbb{R}$ is the $\mathbb{C}$-valued function $\widehat{\mu} : \mathbb{R} \longrightarrow \mathbb{C}$ defined on $\mathbb{R}$ via

$$\widehat{\mu}(\theta) \;:=\; E\left\{\, e^{\mathbf{i}\theta X} \,\right\} \;=\; \int_{\mathbb{R}} e^{\mathbf{i}\theta x}\,\mu(\mathrm{d}x), \quad \text{for } \theta \in \mathbb{R}.$$

Definition A.2 (Characteristic function of a random variable)
Let X be an $\mathbb{R}$-valued random variable, and P_X its distribution measure on $\mathbb{R} = \mathrm{codomain}(X)$. The characteristic function of X is by definition the Fourier transform of P_X. Explicitly, the characteristic function of X is the function $\widehat{P}_X : \mathbb{R} \longrightarrow \mathbb{C}$ defined by:

$$\widehat{P}_X(\theta) \;=\; \int_{\mathbb{R}} e^{\mathbf{i}\theta x}\,P_X(\mathrm{d}x), \quad \text{for } \theta \in \mathbb{R}.$$

Theorem A.3 (Theorem 13.1, [1])
The Fourier transform $\widehat{\mu}$ of any probability measure μ on $\mathbb{R}$ is a bounded and continuous $\mathbb{C}$-valued function on $\mathbb{R}$, and $\widehat{\mu}(0) = 1$.

Remark A.4
The Fourier transform can thus be regarded as a map from the set[1] of all probability measures on $\mathbb{R}$ into the set of all bounded continuous $\mathbb{C}$-valued functions defined on $\mathbb{R}$.

Theorem A.5 (Theorem 13.3, [1])
Let X be an $\mathbb{R}$-valued random variable and $\alpha, \beta \in \mathbb{R}$. Then, for any $\theta \in \mathbb{R}$,

$$\widehat{P}_{\alpha X+\beta}(\theta) \;=\; e^{\mathbf{i}\beta\theta}\cdot\widehat{P}_X(\alpha\,\theta).$$

[1] Note that the set of probability measures on $\mathbb{R}$ does not form a vector space.

PROOF

$$\widehat{P}_{\alpha X+\beta}(\theta) \;=\; E\left\{e^{\mathbf{i}(\alpha X+\beta)\theta}\right\} \;=\; \int_{\mathbb{R}} e^{\mathbf{i}\beta\theta}\cdot e^{\mathbf{i}(\alpha\theta)x}\,P_X(\mathrm{d}x) \;=\; e^{\mathbf{i}\beta\theta}\cdot\int_{\mathbb{R}} e^{\mathbf{i}(\alpha\theta)x}\,P_X(\mathrm{d}x) \;=\; e^{\mathbf{i}\beta\theta}\cdot\widehat{P}_X(\alpha\,\theta)$$

□

Theorem A.6 (Theorem 15.2, [1])
The characteristic function of the sum of two independent $\mathbb{R}$-valued random variables is the product of their characteristic functions.

More precisely, if $X, Y : \Omega \longrightarrow \mathbb{R}$ are independent $\mathbb{R}$-valued random variables with respective characteristic functions $\widehat{P}_X, \widehat{P}_Y : \mathbb{R} \longrightarrow \mathbb{C}$, then the characteristic function $\widehat{P}_Z$ of the random variable $Z := X + Y$ is given in terms of $\widehat{P}_X$ and $\widehat{P}_Y$ by:

$$\widehat{P}_Z(\theta) \;=\; \widehat{P}_X(\theta)\cdot\widehat{P}_Y(\theta), \quad \text{for each } \theta\in\mathbb{R}.$$

Theorem A.7 (Theorem 13.2, [1])
Let X be an $\mathbb{R}$-valued random variable and suppose that $E\{|X|^m\} < \infty$ for some non-negative integer m. Then the Fourier transform $\widehat{P}_X$ of the distribution measure P_X has continuous derivatives up to order m, and

$$\frac{\mathrm{d}^m}{\mathrm{d}\theta^m}\widehat{P}_X(\theta) \;=\; \mathbf{i}^m E\left\{X^m e^{\mathbf{i}\,\theta X}\right\}$$

Corollary A.8
For an $\mathbb{R}$-valued random variable X,

$$\begin{aligned} E\{|X|\} < \infty &\implies E\{X\} = -\,\mathbf{i}\,\widehat{P}_X'(0), \\ E\left\{X^2\right\} < \infty &\implies E\left\{X^2\right\} = -\widehat{P}_X''(0). \end{aligned}$$

B The Fourier transform of the standard normal distribution measure on $\mathbb{R}$

Recall that probability density function of the standard normal (or standard Gaussian) distribution measure $P_{N(0,1)}$ is

$$f_{N(0,1)}(z) \;:=\; \frac{1}{\sqrt{2\pi}}e^{-z^2/2}, \;\; \textit{for} \;\; z\in\mathbb{R}.$$

The Fourier transform $\widehat{f}_{N(0,1)}$ of $f_{N(0,1)}$ is

$$\widehat{f}_{N(0,1)}(\theta) \;:=\; E\left(e^{\mathbf{i}\,\theta Z}\right) \;=\; \int_{-\infty}^{\infty} e^{\mathbf{i}\,\theta z}\,f_{N(0,1)}(z)\,\mathrm{d}z \;=\; \int_{-\infty}^{\infty}\left(\cos(\theta z)+\mathbf{i}\,\sin(\theta z)\right)\frac{1}{\sqrt{2\pi}}e^{-z^2/2}\,\mathrm{d}z \;=\; e^{-\theta^2/2}, \;\; \textit{for} \;\; \theta\in\mathbb{R},$$

The function $\widehat{f}_{N(0,1)}$ is also, by definition, the Fourier transform $\widehat{P}_{N(0,1)}$ of the standard Gaussian distribution measure $P_{N(0,1)}$. In other words, $\widehat{P}_{N(0,1)} = \widehat{f}_{N(0,1)}$.

C Injectivity and continuity of the Fourier transform on the space of probability measures on $\mathbb{R}$

Theorem C.1 (Injectivity of the Fourier transform on the space of probability measures on $\mathbb{R}$)
If the Fourier transforms of two probability measures on $\mathbb{R}^d$ are equal (as $\mathbb{C}$-valued functions on $\mathbb{R}^d$), then the two probability measures themselves are equal.

See Theorem 14.1, [1].

Remark C.2

Recall that the Fourier transform can be regarded a map from the set of all probability measures on $\mathbb{R}$ into the set of all bounded continuous $\mathbb{C}$-valued functions defined on $\mathbb{R}$. The above injectivity theorem states that this Fourier transform map is injective.

Theorem C.3 (Levy's Continuity Theorem of the Fourier transform)

Let $\{\,\mu_n\,\}_{n\in\mathbb{N}}$ be a sequence of probability measures on $\mathbb{R}^d$, $d \geq 1$, and let $\widehat{\mu_n} : \mathbb{R}^d \longrightarrow \mathbb{C}$ be the Fourier transform of μ_n.

- *If μ_n converges weakly to a measure μ, then $\widehat{\mu_n}$ converges pointwise to $\widehat{\mu}$, i.e. $\widehat{\mu_n}(\theta)$ converges to $\widehat{\mu}(\theta)$, for each $\theta \in \mathbb{R}^d$.*
- *If $\widehat{\mu_u}$ converges pointwise to some function $f : \mathbb{R}^d \longrightarrow \mathbb{C}$, and f is continuous at $\mathbf{0} \in \mathbb{R}^d$, then there exists a probability measure μ on $\mathbb{R}^d$ such that $\widehat{\mu} = f$, and μ_n converges weakly to μ.*

See Theorem 19.1, [1].

References

[1] Jacod, J., and Protter, P. *Probability Essentials*. Springer-Verlag, New York, 2004. Universitext.