CLT & Delta method

Jiahui Ren

October 21, 2025

This is a document prepared for BIOS8004.
Before we start, we need to review some basic properties of moment generating functions.

1 Reviews of Moment Generating Function

Moment generating function can give us a representation of all the moments.

Definition 1.1 (Moment Generating Function)**.** The moment generating function (MGF) of a random variable X is defined as:

$$M_X(t) = E[e^{tX}]$$

if the expectation exists for t in some neighborhood of 0.

Remark 1.1. The MGF of X does not always exist. However, if it exists, then $M_X(t)$ is continuously differentiable in some neighborhood of the origin.

1.1 Some properties of MGFs

Fact 1.1. If X and Y are independent random variables with MGFs $M_X(t)$ and $M_Y(t)$, then the MGF of $X + Y$ is given by:

$$M_{X+Y}(t) = M_X(t)M_Y(t)$$

Proof

Fact 1.2. If X has MGF $M_X(t)$, then the MGF of $aX + b$ is given by:

$$M_{aX+b}(t) = e^{bt}M_X(at)$$

Proof

Fact 1.3. The derivative of the mgf at $t = 0$ gives us moments.

Proof

Theorem 1.1 (Uniqueness of MGF)**.** *Let X and Y be random variables with MGFs $M_X(t)$ and $M_Y(t)$. Suppose their mgfs, $M_X(t)$ and $M_Y(t)$, both exist and are equal for all t in the interval $(h < t < h)$ for some $h > 0$. Then X and Y have the same distribution, say $F_X(\cdot) = F_Y(\cdot)$.*

Theorem 1.2. *Let $X_1, X_2, \ldots, X_n$ be a sequence of random variables with MGFs $M_{X_1}(t), M_{X_2}(t), \ldots, M_{X_n}(t)$.*
Let X be a random variable with MGF $M_X(t)$.
If for all t in an open interval around 0 we have

$$M_{X_n}(t) \to M_X(t) \text{ as } n \to \infty,$$

then $X_n \rightsquigarrow X$. i.e, the limiting distribution of X_n is equal to the distribution of X. (convergence in distribution, converge weakly, converge in law)

Fact 1.4 (MGF of Normal Distribution)**.** If $X \sim N(\mu, \sigma^2)$, then the MGF of X is given by:

$$M_X(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$$

Proof

Remark 1.2. The MGF of a standard normal distribution $Z \sim N(0, 1)$ is $M_Z(t) = e^{\frac{t^2}{2}}$.

Remark 1.3. If we use taylor expansion for e^{tX}, we obtain the series expansion of $M_X(t)$ in terms of the moments of X;

$$M_X(t) = \mathbb{E}[\sum_{i=0}^{\infty} \frac{t^i}{i!} X^i] = \sum_{i=0}^{\infty} \frac{t^i}{i!} \mathbb{E}(X^i) = \sum_{i=0}^{\infty} \frac{t^i}{i!} \mu_i' \tag{1.1}$$

2 Central Limit Theorem

The central limit theorem(CLT) is one of the most important theorems in statistics. Roughly speaking, it gives us an approximate distribution of an average without any distributional assumption(other than independence and finite mean and variances).

Theorem 2.1 (Central Limit Theorem)**.** *Let $X_1, X_2, \dots, X_n$ be a sequence of independent random variables with mean μ and variance σ^2. Assume that the mgf $\mathbb{E}[e^{tX_i}]$ is finite for t in a neighborhood around 0. Let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ be the sample mean. Let*

$$Z_n = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \tag{2.1}$$

Then the distribution of Z_n converges to the standard normal distribution $Z \sim N(0,1)$ as $n \to \infty$. i.e, $Z_n \rightsquigarrow Z$. Hence, as $n \to \infty$,

$$\mathbb{P}(Z_n \leq t) \to \Phi(t) \tag{2.2}$$

where $\Phi(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$ is the cumulative distribution function of the standard normal distribution.

Proof

Remark 2.1. The most general version of the CLT does not require any assumption about the mgf. The classic proofs use characteristic functions.

Example 2.1 (Confidence interval for the unknown mean and known variance)**.** We would like to construct a confidence interval for the unknown mean μ with known variance σ^2. $X_1, X_2, \dots, X_n$ are iid random variables. $\hat{\mu} = \bar{X}$ is the sample mean.
We would like to find a random set C such that

$$\mathbb{P}(\mu \in C) \geq 1 - \alpha. \tag{2.3}$$

Take

$$C = [\hat{\mu} - t, \hat{\mu} + t]. \tag{2.4}$$

Then

$$\mathbb{P}(\mu \in C) = \mathbb{P}(\hat{\mu} - t \leq \mu \leq \hat{\mu} + t) \tag{2.5}$$

$$\mathbb{P}(\frac{\sqrt{n}|\hat{\mu} - \mu|}{\sigma} \leq \frac{\sqrt{n}t}{\sigma}) \approx \mathbb{P}(|Z| \leq \frac{\sqrt{n}t}{\sigma}) \tag{2.6}$$

where $Z \sim N(0,1)$. Let $\Phi(t)$ be the cdf of Z. Then define

$$z_\alpha = \Phi^{-1}(1-\alpha) \tag{2.7}$$

where Φ^{-1} is the inverse of Φ. We know that

$$\mathbb{P}(Z > z_{\alpha/2}) = \mathbb{P}(Z < -z_{\alpha/2}) = \alpha/2 \tag{2.8}$$

Therefore,

$$\frac{\sqrt{n}t}{\sigma} = z_{\frac{\alpha}{2}} \tag{2.9}$$

Therefore,

$$t = \frac{z_{\frac{\alpha}{2}}\sigma}{\sqrt{n}} \tag{2.10}$$

Therefore,

$$C = [\hat{\mu} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \hat{\mu} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}] \tag{2.11}$$

is a confidence interval for μ with confidence level $1-\alpha$.

3 Lyapnov Central Limit Theorem

Suppose $X_1, X_2, \ldots, X_n$ are independent but not identically distributed. Let $X_i \overset{\text{IID}}{\sim} [\mu_i, \sigma_i^2]$.

Definition 3.1 (Lyapnov Condition)**.** The sequence $\{X_i\}$ satisfies the Lyapnov condition if

$$\sum_{i=1}^{n} \mathbb{E}|\frac{X_i - \mu_i}{s_n}|^{2+\delta} \to 0 \text{ for some } \delta > 0, \text{ where } s_n^2 := \sum_{i=1}^{n} \sigma_i^2 \tag{3.1}$$

Theorem 3.1 (Lyapnov Central Limit Theorem)**.** *If the sequence $\{X_i\}$ satisfies the Lyapnov condition, then*

$$\frac{1}{s_n}\sum_{i=1}^{n}(X_i - \mu_i) \rightsquigarrow N(0,1) \tag{3.2}$$

Remark 3.1. When $\delta = 1$, then the Lyapnov condition becomes

$$\lim_{n\to\infty} \frac{1}{s_n^3}\sum_{i=1}^{n} \mathbb{E}|X_i - \mu_i|^3 = 0. \tag{3.3}$$

Remark 3.2 (Interpretation of the Lyapnov condition)**.** The Lyapnov condition

$$\frac{1}{s_n^{2+\delta}}\sum_{i=1}^{n} \mathbb{E}|X_i - \mu_i|^{2+\delta} \to 0$$

means that, after scaling by the natural variance scale $s_n^2 = \sum_i \sigma_i^2$, the tail contribution of each summand is negligible.
- **Easy sufficient conditions**:
1) Uniformly bounded $(2+\delta)$-moments: $\sup_i \mathbb{E}|X_i - \mu_i|^{2+\delta} < \infty$;
2) No dominant variance: $\max_i \sigma_i^2/s_n^2 \to 0$.
Then $\sum_i \mathbb{E}|X_i - \mu_i|^{2+\delta} = O(n)$ while $s_n^{2+\delta} \gg n$, hence the ratio $\to 0$.
- **When it fails**: a few terms with variances comparable to s_n^2 or with very heavy tails can violate the condition; then a single summand can drive the limit and the normal approximation may fail.

4 Multivariate Central Limit Theorem

We may extend the CLT to the multivariate case.

Theorem 4.1 (Multivariate Central Limit Theorem)**.** *Let $X_1, X_2, \ldots, X_n$ be a sequence of independent random variables with mean $\mu \in \mathbb{R}^d$ and covariance matrix $\Sigma \in \mathbb{R}^{d\times d}$. Let $\hat{\mu} = \frac{1}{n}\sum_{i=1}^n X_i$ be the sample mean. Then*

$$\sqrt{n}(\hat{\mu} - \mu) \rightsquigarrow N_d(0, \Sigma). \tag{4.1}$$

Proof

5 CLT with Estimated Variance

Sometimes we may not know the variance of the random variables, but we can estimate it from the data.

$$\hat{\sigma_n}^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \hat{\mu})^2. \tag{5.1}$$

If we replace σ^2 with $\hat{\sigma}^2$, we still have the CLT.

Theorem 5.1. *Let $X_1, X_2, \ldots, X_n$ be a sequence of independent random variables with mean μ and variance σ^2. Let $\hat{\sigma_n}^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \hat{\mu})^2$. Then*

$$\frac{\sqrt{n}(\hat{\mu} - \mu)}{\hat{\sigma_n}} \rightsquigarrow N(0, 1). \tag{5.2}$$

Proof

6 Berry-Esseen Theorem

The quality of the CLT:

Theorem 6.1 (Berry–Esseen). *If $\{X_i\}$ are i.i.d. with $\mathbb{E}X_i = 0$, $\mathsf{Var}(X_i) = \sigma^2 > 0$, and $\mathbb{E}|X_i|^3 = \rho_3 < \infty$, then with $Z_n = \frac{\sum_{i=1}^n X_i}{\sigma\sqrt{n}}$,*

$$\sup_{t\in\mathbb{R}} \left|\mathbb{P}(Z_n \le t) - \Phi(t)\right| \le C\,\frac{\rho_3}{\sigma^3\sqrt{n}},$$

where C is a positive constant.

Remark 6.1 (Value of C). Calculated upper bounds on the constant C have decreased markedly over the years, from the original value of 7.59 by Esseen in 1942. Numbers like 9, 1, 0.8, 0.56, 0.4748 all appear in the literature; theyre all valid bounds with varying sharpness.

7 Delta Method

The delta method is a technique for approximating the distribution of a function of a random variable.

Theorem 7.1 (Delta Method). *Suppose*

$$\frac{\sqrt{n}(X_n - \mu)}{\sigma} \rightsquigarrow N(0,1) \tag{7.1}$$

and g is a continuously differentiable function such that $g'(\mu) \neq 0$. Then

$$\frac{\sqrt{n}(g(X_n) - g(\mu))}{\sigma} \rightsquigarrow N(0, g'(\mu)^2) \tag{7.2}$$

Proof

Example 7.1 (Univariate: $g(x) = \exp(x)$). Suppose we have $X_1, \dots, X_n \sim P$ with $E[X] = \mu$, $\mathsf{Var}(X) = \sigma^2 < \infty$ and let $\hat{\mu}_n = \bar{X}$. Consider $Y_n = \exp(\hat{\mu}_n)$. Since $g'(\mu) = \exp(\mu)$, by the Delta method,

$$\sqrt{n}\,\frac{\exp(\hat{\mu}_n) - \exp(\mu)}{\sigma} \;\rightsquigarrow\; N\big(0,\, \exp(2\mu)\big).$$

Theorem 7.2 (Multivariate Delta Method). *Suppose we have random vectors $X_1, \dots, X_n \in \mathbb{R}^d$, and $g : \mathbb{R}^d \mapsto \mathbb{R}$ is a continuously differentiable function, then*

$$\sqrt{n}\,(g(\bar{X}_n) - g(\mu)) \rightsquigarrow N(0, \tau^2)$$

where $\bar{X}_n = n^{-1}\sum_i X_i$,

$$\tau^2 = \nabla_\mu(g)^T \Sigma \nabla_\mu(g))$$

and

$$\nabla_\mu(g) = \begin{pmatrix} \frac{\partial g(x)}{\partial x_1} \\ \vdots \\ \frac{\partial g(x)}{\partial x_d} \end{pmatrix}_{x=\mu}$$

is the gradient of g evaluated at μ.

Proof. By the multivariate CLT, $\sqrt{n}(\bar{X}_n - \mu) \rightsquigarrow N_d(0, \Sigma)$. By Taylor's expansion,

$$g(\bar{X}_n) - g(\mu) = \nabla g(\tilde{\mu}_n)^T(\bar{X}_n - \mu),$$

for some random $\tilde{\mu}_n$ on the line segment between $\bar{X}_n$ and μ. Since $\bar{X}_n \xrightarrow{\text{pr}} \mu$ and ∇g is continuous at μ, we have $\nabla g(\tilde{\mu}_n) \xrightarrow{\text{pr}} \nabla g(\mu)$. Hence

$$\sqrt{n}(g(\bar{X}_n) - g(\mu)) = \nabla g(\tilde{\mu}_n)^T \sqrt{n}(\bar{X}_n - \mu) \rightsquigarrow \nabla_\mu(g)^T Z,$$

where $Z \sim N_d(0, \Sigma)$. The limit is univariate normal with variance $\nabla_\mu(g)^T \Sigma \nabla_\mu(g) = \tau^2$. Slutsky's theorem justifies replacing $\nabla g(\tilde{\mu}_n)$ by $\nabla_\mu(g)$ in the limit. □

Example 7.2 (Multivariate: product $g(x_1, x_2) = x_1x_2$)**.** Let $X_i = (X_{i1}, X_{i2})$ have mean $\mu = (\mu_1, \mu_2)$ and covariance matrix Σ. The CLT gives $\sqrt{n}(\bar{X}_n - \mu) \rightsquigarrow N_d(0, \Sigma)$. For $g(x_1, x_2) = x_1x_2$, we have $\nabla g(x_1, x_2) = (x_2, x_1)^T$, so $\nabla g(\mu) = (\mu_2, \mu_1)^T$. Therefore,

$$\sqrt{n}\,(\bar{X}_1\bar{X}_2 - \mu_1\mu_2) \;\rightsquigarrow\; N\big(0,\, \tau^2\big), \qquad \tau^2 = (\mu_2, \mu_1)^T \Sigma (\mu_2, \mu_1).$$

8 Stocahstic Order Notation

Deterministic o and O notations

Notation	Definition	Intuitive meaning
$a_n = o(b_n)$	$\lvert\frac{a_n}{b_n}\rvert \to 0$ as $n \to \infty$	a_n goes to 0 faster than b_n
$a_n = O(b_n)$	$\exists\, M, s.t. \mathbb{1}(\lvert a_n/b_n\rvert > M) = 0\ \forall$ large n	a_n/b_n is bounded eventually
$a_n \asymp b_n$	$a_n = O(b_n)$ and $b_n = O(a_n)$	a_n and b_n are of the same order

Stochastic o and O notations

Notation	Definition	Intuitive meaning
$X_n = o_p(A_n)$	$\lvert\frac{X_n}{A_n}\rvert \xrightarrow{\text{pr}} 0$ as $n \to \infty$	$X_n \xrightarrow{\text{pr}} 0$ faster than A_n
$X_n = O_p(A_n)$	$\forall\, \varepsilon > 0\ \exists\, M, s.t. \mathbb{P}\Big(\lvert\frac{X_n}{A_n}\rvert > M\Big) < \varepsilon\ \forall$ large n	X_n/A_n is bounded in probability eventually

Example 8.1. Let $X_1, X_2, \ldots, X_n$ be iid random variables with finite variance. Define $\hat{\mu} = \frac{1}{n}\sum_{i=1}^n X_i$. Then

$$\hat{\mu} - \mu = o_p(1) \tag{8.1}$$

$$\hat{\mu} - \mu = O_p(1/\sqrt{n}) \tag{8.2}$$

Proposition 8.1. *Let $\{X_n\}_{n\in\mathbb{N}}$ and $\{Y_n\}_{n\in\mathbb{N}}$ be two sequences of random variables. Let a_n be a sequence of real numbers.*

1. *If $X_n = O_p(1)$ and $Y_n = O_p(1)$, then $X_n + Y_n = O_p(1)$.*
2. *If $X_n = o_p(1)$ and $Y_n = o_p(1)$, then $X_n + Y_n = o_p(1)$.*
3. *If $X_n = o_p(1)$ and $Y_n = O_p(1)$, then $X_n + Y_n = O_p(1)$.*

4. *If* $X_n = o_p(1)$ *and* $Y_n = O_p(1)$*, then* $X_n Y_n = o_p(1)$.
5. *If* $X_n = O_p(1)$*, then* $o_p(X_n) = o_p(1)$.
6. *If* $X_n = o_p(a_n)$*, then* $X_n = a_n o_p(1)$.
7. *If* $X_n = O_p(a_n)$*, then* $X_n = a_n O_p(1)$.

Acknowledgement

When I prepared this document, I referred to the following resources:

- BIOS 8004 Advacned Statistical Theory, CityUHK
- STAT 4003 Statistical Inference, CUHK
- STAT 5010 Advanced Statistical Inference, CUHK
- STAT 5005 Advanced Probability Theory, CUHK
- https://en.wikipedia.org/wiki/Berry%E2%80%93Esseen_theorem
- http://parker.ad.siu.edu/Olive/lsch3.pdf

Appendix

The appendix includes something that do not cover in the course materials but would be useful for understanding.

Characteristic Functions

In this section, we introduce the classical tool of proving distributional approximations via characteristic functions.

Definition 8.1. The **characteristic function (ch.f.)** of a random variable X is defined to be

$$\varphi_X(t) := Ee^{itX} = E\cos(tX) + i \cdot E\sin(tX).$$

Properties.

1. $\varphi_X(0) = 1$, $|\varphi_X(t)| \le 1$.
2. $\varphi_X(-t) = \overline{\varphi_X(t)}$. (conjugate)
3. $|\varphi_X(t+h) - \varphi_X(t)| \le E|e^{ihX} - 1| \to 0$, as $h \to 0$. (by DCT). That is, $\varphi_X(t)$ is uniformly continuous.
4. $\varphi_{aX+b}(t) = e^{itb}\varphi_X(at)$.
5. If X_1 is independent of X_2, then $\varphi_{X_1+X_2}(t) = \varphi_{X_1}(t)\varphi_{X_2}(t)$.

Theorem 8.2. *If $E(X^2) < \infty$, then*

$$\varphi_X(t) = 1 + i \cdot tE(X) - \frac{t^2}{2}E(X^2) + o(t^2), \text{ as } t \to 0.$$

Proof. By Taylor's expansion,

$$\varphi_X(t) = Ee^{itX} = 1 + EiXe^{itX}|_{t=0} \cdot t + E\frac{(iX)^2}{2}e^{itX}|_{t=0} \cdot t^2 + \text{error},$$

where

$$\begin{aligned} |\text{error}| &\le CE\left[(t|X|)^3 \wedge (t^2X^2)\right] &(8.3)\\ &= Ct^2E\left[(t|X|)^3 \wedge (X^2)\right] &(8.4)\\ &= o(t^2), \text{ as } t \to 0 &(8.5)\end{aligned}$$

by DCT. ∎

Proof of CLT without assumption of mgf

Using characteristic functions, we now give the second proof of the CLT.

Theorem 8.3. *Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables such that $EX_i = \mu, \mathsf{Var}(X_i) = \sigma^2$. Let*

$$W_n = \sum_{i=1}^{n} \frac{X_i - \mu}{\sigma\sqrt{n}}.$$

Then

$$W_n \rightsquigarrow Z \sim N(0,1).$$

Proof. N.T.S. $Ee^{itW_n} \to e^{-\frac{t^2}{2}}$ for all $t \in \mathbb{R}$. We have, by the expression of W_n and independence,

$$Ee^{itW_n} = E\exp\left(it\left(\frac{X_1-\mu}{\sigma\sqrt{n}} + \cdots + \frac{X_n-\mu}{\sigma\sqrt{n}}\right)\right) = \prod_{j=1}^n Ee^{i\frac{t}{\sqrt{n}}(X_j-\mu)} \tag{8.6}$$

$$= \prod_{j=1}^n \left[1 + i\frac{t}{\sigma\sqrt{n}}E(X_j-\mu) - \frac{t^2}{2\sigma^2 n}E(X_j-\mu)^2 + o\left(\frac{t^2}{n}\right)\right] \quad \text{(from Theorem 8.3)} \tag{8.7}$$

$$= \left[1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right]^n \tag{8.8}$$

$$\to e^{-t^2/2}. \tag{8.9}$$

□

Lindeberg condition and Lindeberg–Feller Central Limit Theorem

Theorem 8.4 (The Lindeberg-Feller Theorem)**.** *Assume for each n, $\xi_{n1}, \xi_{n2}, \ldots, \xi_{nn}$ are independent with $E\xi_{ni} = 0$ for all i and $E\sum_{i=1}^n \xi_{ni}^2 = 1$. If*

$$\forall \varepsilon > 0, \quad \sum_{i=1}^n E\xi_{ni}^2 \mathbf{1}_{\{|\xi_{ni}|>\varepsilon\}} \to 0, \quad \textit{(Lindeberg's Condition)}$$

then

$$\sum_{i=1}^n \xi_{ni} \rightsquigarrow N(0,1).$$

Proof of the Lindeberg-Feller theorem. Let

$$\varphi_n(t) = Ee^{it\sum_{i=1}^n \xi_{ni}}.$$

We have

$$\varphi_n(t) = \prod_{i=1}^n Ee^{it\xi_{ni}} \tag{8.10}$$

$$= \prod_{i=1}^n E\left[1 + it\xi_{ni} - \frac{t^2}{2}\xi_{ni}^2 + O(t^2\xi_{ni}^2\mathbf{1}_{\{|\xi_{ni}|>\varepsilon\}}) + O(t^3|\xi_{ni}|^3\mathbf{1}_{\{|\xi_{ni}|\le\varepsilon\}})\right] \tag{8.11}$$

$$= \prod_{i=1}^n \left[1 - \frac{t^2}{2}E\xi_{ni}^2 + O(t^2E\xi_{ni}^2\mathbf{1}_{\{|\xi_{ni}|>\varepsilon\}}) + O(t^3E|\xi_{ni}|^3)\right] \tag{8.12}$$

$$\to \prod_{i=1}^n e^{-\frac{t^2}{2}E\xi_{ni}^2} = e^{-t^2/2}, \tag{8.13}$$

where we used Lemma: Let $z_1, \ldots, z_n$ and $w_1, \ldots, w_n$ be complex numbers with $|z_i| \le 1, |w_i| \le 1$ for all i. Then

$$\left|\prod_{i=1}^n z_i - \prod_{i=1}^n w_i\right| \le \sum_{i=1}^n |z_i - w_i|.$$

□

Remark 8.1. CLT for i.i.d. sequence is a corollary of the above theorem: For $X_1, X_2, \ldots$, i.i.d. with $EX_i = \mu$, $\mathsf{Var}(X_i) = \sigma^2$. Consider $\xi_{ni} = \frac{X_i-\mu}{\sigma\sqrt{n}}$ and $W_n := \sum_{i=1}^n \xi_{ni}$. It can be checked by DCT that the Lindeberg condition is satisfied and hence CLT. Then the Lindeberg condition is:

Definition 8.2 (Lindeberg condition)**.** Let $X_1, \dots, X_n$ be independent (not necessarily identical) with means μ_i and variances σ_i^2, and let $s_n^2 := \sum_{i=1}^n \sigma_i^2$. The Lindeberg condition holds if for every $\varepsilon > 0$,

$$\frac{1}{s_n^2} \sum_{i=1}^n \mathbb{E}\Big[(X_i - \mu_i)^2 \, \mathbf{1}\{|X_i - \mu_i| > \varepsilon s_n\}\Big] \xrightarrow[n\to\infty]{} 0.$$

and the theorem is:

Theorem 8.5 (Lindeberg–Feller Central Limit Theorem)**.** *Under the Lindeberg condition, we have*

$$\frac{1}{s_n} \sum_{i=1}^n (X_i - \mu_i) \ \rightsquigarrow \ N(0,1).$$

In particular, Lyapnov's condition implies Lindeberg's, hence also the conclusion above.

Proof of the Lyapnov Central Limit Theorem. Set $Y_{ni} = X_i - \mu_i$. Note $\mathbb{E}Y_{ni} = 0$ and $\sum_{i=1}^n \mathsf{Var}(Y_{ni}) = s_n^2$. It suffices to show the Lindeberg condition and then apply the Lindeberg–Feller CLT to S_n/s_n.

Fix $\varepsilon > 0$. By Markov's inequality, for any i,

$$Y_{ni}^2 \mathbf{1}\{|Y_{ni}| > \varepsilon s_n\} \le \frac{|Y_{ni}|^{2+\delta}}{(\varepsilon s_n)^\delta}.$$

Taking expectations and summing, we get

$$\frac{1}{s_n^2} \sum_{i=1}^n \mathbb{E}\big[Y_{ni}^2 \mathbf{1}\{|Y_{ni}| > \varepsilon s_n\}\big] \le \frac{1}{\varepsilon^\delta s_n^{2+\delta}} \sum_{i=1}^n \mathbb{E}|Y_{ni}|^{2+\delta} \xrightarrow[n\to\infty]{} 0,$$

by the Lyapnov condition. Hence the Lindeberg condition holds.

By the Lindeberg–Feller CLT, we conclude

$$\frac{1}{s_n} \sum_{i=1}^n (X_i - \mu_i) \rightsquigarrow N(0,1).$$

□

Remark 8.2. Implication: Lyapnov ⇒ Lindeberg. For any $\varepsilon > 0$,

$$\mathbb{E}\big[(X_i - \mu_i)^2 \, \mathbf{1}\{|X_i - \mu_i| > \varepsilon s_n\}\big] \le \frac{\mathbb{E}|X_i - \mu_i|^{2+\delta}}{(\varepsilon s_n)^\delta},$$

so dividing by s_n^2 and summing gives Lindeberg $\to 0$.

Proof of Berry-Esseen

Proof of Berry–Esseen. Let $\varphi_n(u) = \mathbb{E}e^{iuZ_n}$ be the characteristic function (CF) of Z_n and $\varphi(u) = e^{-u^2/2}$ that of $N(0,1)$. By Taylor expansion of $\mathbb{E}e^{iuX_1/(\sigma\sqrt{n})}$ and independence,

$$\varphi_n(u) = \Big(1 - \tfrac{u^2}{2n} + R_3(u/n^{1/2})\Big)^n, \qquad |R_3(v)| \le \tfrac{|u|^3}{6n^{3/2}} \, \mathbb{E}\Big|\tfrac{X_1}{\sigma}\Big|^3.$$

Hence for all real u,

$$|\varphi_n(u) - \varphi(u)| \le c_1 \, \frac{|u|^3}{\sqrt{n}} \, \frac{\mathbb{E}|X_1|^3}{\sigma^3} \, e^{-u^2/4},$$

for some absolute constant c_1. By Esseen's smoothing lemma, for any $T > 0$,

$$\sup_t |\mathbb{P}(Z_n \le t) - \Phi(t)| \le \frac{1}{\pi} \int_{-T}^{T} \frac{|\varphi_n(u) - \varphi(u)|}{|u|} \, du + \frac{c_2}{T},$$

where c_2 is absolute. Using the previous bound and integrating, we get

$$\sup_t |\mathbb{P}(Z_n \le t) - \Phi(t)| \le c_3 \, \frac{\mathbb{E}|X_1|^3}{\sigma^3 \sqrt{n}} \, T^2 + \frac{c_2}{T}.$$

Optimizing over $T \asymp n^{1/6}$ yields

$$\sup_t |\mathbb{P}(Z_n \le t) - \Phi(t)| \le C \, \frac{\mathbb{E}|X_1|^3}{\sigma^3 \sqrt{n}},$$

as claimed. □