Probability Theory & Mathematical Statistics

Limit Theorems

尹一通 Nanjing University, 2023 Spring

Limit Theorems

Let $X_1, X_2, \ldots$ be *i.i.d.* random variables with $\mu = \mathbb{E}[X_1]$ and $\mathbf{Var}[X_1] = \sigma^2$.

And let $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ be the sample mean.

- Law of large numbers (**LLN**): sample mean → expectation

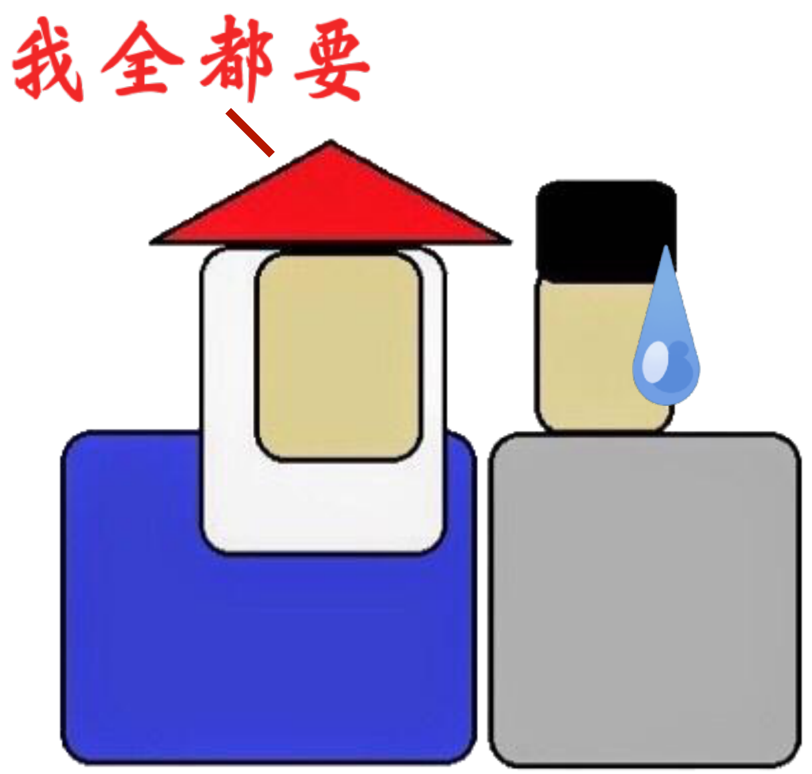
$$\overline{X}_n \longrightarrow \mu \quad \text{as } n \to \infty$$

- Central limit theorem (**CLT**): standardized sample mean → standard normal

$$\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \longrightarrow N(0,1) \quad \text{as } n \to \infty$$

Convergence

- A real sequence $\{a_n\}$ converges to $a \in \mathbb{R}$, denoted $\lim_{n\to\infty} a_n = a$ or $a_n \to a$, if for all $\epsilon > 0$, there is N such that $|a_n - a| < \epsilon$ for all $n > N$
- A sequence $f_1, f_2, \ldots : \Omega \to \mathbb{R}$ is said to <u>**converge pointwise**</u> to $f : \Omega \to \mathbb{R}$, if and only if $\lim_{n\to\infty} f_n(x) = f(x)$ for all $x \in \Omega$
- For random variables $X_1, X_2, \ldots$ and X on probability space $(\Omega, \Sigma, \Pr)$:
  - random variables $X_1, X_2, \ldots : \Omega \to \mathbb{R}$ and $X : \Omega \to \mathbb{R}$ are functions
  - CDFs $F_{X_1}, F_{X_2}, \ldots : \mathbb{R} \to [0,1]$ and $F_X : \mathbb{R} \to [0,1]$ are functions
- Should $X_n \to X$ be: $X_n \to X$ pointwise or $F_{X_n} \to F_X$ pointwise?



Convergence of Random Variables

0.  $\cdots \to U_{[0,1]}$

Modes of Convergence

- Let $X, X_1, X_2, \ldots : \Omega \to \mathbb{R}$ be random variables on prob. space $(\Omega, \Sigma, \Pr)$.

- $\{X_n\}$ converges in distribution (依分布收敛) to X, denoted $X_n \xrightarrow{D} X$, if

$$F_{X_n}(x) = \Pr(X_n \leq x) \to F_X(x) = \Pr(X \leq x) \quad \text{as} \quad n \to \infty$$

  for all $x \in \mathbb{R}$ at which $F_X(x)$ is continuous

- $\{X_n\}$ converges in probability (依概率收敛) to X, denoted $X_n \xrightarrow{P} X$, if

$$\Pr(|X_n - X| > \epsilon) = 0 \quad \text{as} \quad n \to \infty \quad \text{for all } \epsilon > 0$$

- $\{X_n\}$ converges almost surely to X, denoted $X_n \xrightarrow{a.s.} X$, if $\exists A \in \Sigma$ such that

$$\lim_{n \to \infty} X_n(\omega) = X(\omega) \quad \text{for all } \omega \in A, \quad \text{and} \quad \Pr(A) = 1$$

Modes of Convergence

- Let $X_1, X_2, \ldots$ and X be random variables on probability space $(\Omega, \Sigma, \Pr)$.

- $X_n \xrightarrow{D} X$ (convergence <u>in distribution</u> / <u>in law</u> / weak convergence of measure) if

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

$F_{X_n} \to F_X$ pointwise on continuous set

for all $x \in \mathbb{R}$ at which $F_X(x)$ is continuous

- $X_n \xrightarrow{P} X$ (convergence <u>in probability</u> / <u>in measure</u>) if

$$\lim_{n\to\infty} \Pr(|X_n - X| > \epsilon) = 0 \quad \text{for all } \epsilon > 0$$

$X_n \to X$ in measure

- $X_n \xrightarrow{a.s.} X$ (convergence <u>almost surely</u> / <u>almost everywhere</u> / <u>w.p. 1</u>) if

$$\Pr\left(\lim_{n\to\infty} X_n = X\right) = 1$$

$X_n \to X$ pointwise on a set of measure 1

Convergence in Distribution

- Let $X_1, X_2, \ldots$ and X be random variables on probability space $(\Omega, \Sigma, \Pr)$.
- $X_n \xrightarrow{D} X$ (convergence <u>in distribution</u> / <u>in law</u> / weak convergence of measure) if

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

$F_{X_n} \to F_X$ pointwise on continuous set

for all $x \in \mathbb{R}$ at which $F_X(x)$ is continuous

- The restriction on continuity set is necessary, consider:

uniform X_n on $(0,1/n)$, which satisfies $X_n \xrightarrow{D} X$, where $\Pr(X=0)=1$

- $X_n \xrightarrow{D} X$ and $F_X = F_Y \implies X_n \xrightarrow{D} Y$ (convergence in distribution depends only on distribution)
- $X_n \xrightarrow{D} X$ is a <u>weak convergence of measures</u>

Convergence in Probability

- Let $X_1, X_2, \ldots$ and X be random variables on probability space $(\Omega, \Sigma, \Pr)$.
- $X_n \xrightarrow{P} X$ (convergence <u>in probability</u>) if

$$\lim_{n\to\infty} \Pr(|X_n - X| > \epsilon) = 0 \quad \text{for all } \epsilon > 0$$

$X_n \to X$ in measure

- Functions $X_n : \Omega \to \mathbb{R}$ converges to $X : \Omega \to \mathbb{R}$ <u>in measure</u> $\Pr$
- $X_n \xrightarrow{P} X \implies X_n \xrightarrow{D} X$
  - **Counterexample for converse**: X is uniform on $[0,1]$ and $X_n = 1 - X$
- If $X_n \xrightarrow{D} c$, where $c \in \mathbb{R}$ is a constant, then $X_n \xrightarrow{P} c$
  - **Proof**: $\Pr(|X_n - c| > \epsilon) = \Pr(X_n < c - \epsilon) + \Pr(X_n > \epsilon + c) \to 0$ if $X_n \xrightarrow{D} c$

Almost Sure Convergence

- Let $X_1, X_2, \ldots$ and X be random variables on probability space $(\Omega, \Sigma, \Pr)$.
- $X_n \xrightarrow{a.s.} X$ (convergence <u>almost surely</u> / <u>almost everywhere</u> / **w.p. 1**) if

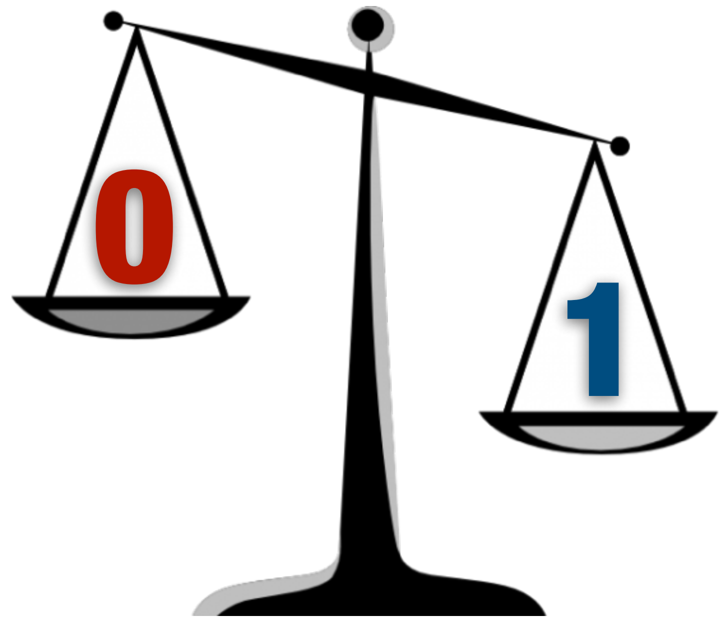
$$\Pr\left(\lim_{n\to\infty} X_n = X\right) = 1$$

$X_n \to X$ pointwise on a set of measure 1

- $X_n : \Omega \to \mathbb{R}$ converges to $X : \Omega \to \mathbb{R}$ **almost everywhere** except a null set
- The event $\lim_{n\to\infty} X_n = X$ is: $\bigcap_{m=1}^{\infty} \bigcup_{n_0=1}^{\infty} \bigcap_{n=n_0}^{\infty} \{\omega \in \Omega \mid |X_n(\omega) - X(\omega)| \le 1/m\}$
- $X_n \xrightarrow{a.s.} X \implies X_n \xrightarrow{P} X$
  - **Counterexample for converse**: $\{X_n\}$ are independent $\text{Bernoulli}(1/n)$. We have $X_n \xrightarrow{P} 0$, but we do not have $X_n = 0$ almost everywhere as $n \to \infty$.

Borel–Cantelli Lemmas*

(博雷尔－坎特利引理 / 波莱尔－坎泰利引理 / zero-one law)



- Let $A_1, A_2, \ldots$ be a sequence of events from a probability space $(\Omega, \Sigma, \Pr)$.
  Let A be the event that infinitely many of the A_n occurs:
  $$A = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m$$
  denoted A_n ***infinitely often***, or A_n ***i.o.***
- (1st lemma) $\sum_{n=1}^{\infty} \Pr(A_n) < \infty \implies \Pr(A) = 0$
- (2nd lemma) $\sum_{n=1}^{\infty} \Pr(A_n) = \infty$ and $A_1, A_2, \ldots$ are independent $\implies \Pr(A) = 1$

Continuity of Probability Measures*

- Let $A_1 \subseteq A_2 \subseteq A_3 \subseteq \ldots$ be an increasing sequence of events, and write A for their limit

$$A = \bigcup_{i=1}^{\infty} A_i = \lim_{i \to \infty} A_i \,.$$

Then $\Pr(A) = \lim_{i \to \infty} \Pr(A_i)$.

- **Proof**: Express A as a disjoint union $A = A_1 \uplus (A_2 \backslash A_1) \uplus (A_3 \backslash A_2) \uplus \cdots$. Then

$$\begin{aligned}
\Pr(A) &= \Pr(A_1) + \sum_{i=1}^{\infty} \Pr(A_{i+1} \backslash A_i) \\
&= \Pr(A_1) + \lim_{n \to \infty} \sum_{i=1}^{n-1} [\Pr(A_{i+1}) - \Pr(A_i)] \\
&= \lim_{n \to \infty} \Pr(A_n)
\end{aligned}$$

Continuity of Probability Measures*

- Let $A_1 \subseteq A_2 \subseteq A_3 \subseteq \ldots$ be an increasing sequence of events, and write A for their limit

$$A = \bigcup_{i=1}^{\infty} A_i = \lim_{i \to \infty} A_i \,.$$

  Then $\Pr(A) = \lim_{i \to \infty} \Pr(A_i)$.

- Let $B_1 \supseteq B_2 \supseteq B_3 \supseteq \ldots$ be an decreasing sequence of events, and write B for their limit

$$B = \bigcap_{i=1}^{\infty} B_i = \lim_{i \to \infty} B_i \,.$$

  Then $\Pr(B) = \lim_{i \to \infty} \Pr(B_i)$.

- **Proof**: Consider the complements $B_1^c \subseteq B_2^c \subseteq B_3^c \subseteq \ldots$ which is an increasing sequence.

Borel–Cantelli Lemmas*

$$A = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m$$

(1st lemma) $\sum_{n=1}^{\infty} \Pr(A_n) < \infty \implies \Pr(A) = 0$

Proof: By union bound, $\Pr\left(\bigcup_{m=n}^{\infty} A_m\right) \leq \sum_{m=n}^{\infty} \Pr(A_m)$, which $\to 0$ as $n \to \infty$,

assuming that $\sum_{n=1}^{\infty} \Pr(A_n) < \infty$ converges.

And by **continuity** of $\Pr$, we have $\Pr(A) = \lim_{n\to\infty} \Pr\left(\bigcup_{m=n}^{\infty} A_m\right) = 0$

Borel–Cantelli Lemmas*

$$A = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m$$

(2nd lemma) $\sum_{n=1}^{\infty} \Pr(A_n) = \infty$ and $A_1, A_2, \ldots$ are independent $\implies \Pr(A) = 1$

Proof: By independence, $\Pr\left(\bigcap_{m=n}^{\infty} A_m^c\right) = \prod_{m=n}^{\infty} \left(1 - \Pr(A_m)\right) \leq \exp\left(-\sum_{m=n}^{\infty} \Pr(A_m)\right) = 0,$

assuming the divergence of $\sum_{n=1}^{\infty} \Pr(A_n) = \infty$.

By **continuity** of $\Pr$, $\Pr(A^c) = \Pr\left(\bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m^c\right) = \lim_{n\to\infty} \Pr\left(\bigcap_{m=n}^{\infty} A_m^c\right) = 0 \implies \Pr(A) = 1$

Strength of Convergence

- $(X_n \xrightarrow{a.s.} X) \implies (X_n \xrightarrow{P} X) \implies (X_n \xrightarrow{D} X)$

Proof* $(X_n \xrightarrow{a.s.} X \implies X_n \xrightarrow{P} X)$: Let $A_n(\epsilon) = \{\,|X_n - X| > \epsilon\}$. Then for any $\epsilon > 0$

$$\lim_{n\to\infty} X_n = X \implies \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m^c(\epsilon)$$

Assume $X_n \xrightarrow{a.s.} X$. Then $1 = \Pr\left(\lim_{n\to\infty} X_n = X\right) = \Pr\left(\bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m^c(\epsilon)\right)$

$$\implies 0 = \Pr\left(\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m(\epsilon)\right) = \lim_{n\to\infty} \Pr\left(\bigcup_{m=n}^{\infty} A_m(\epsilon)\right)$$ (by continuity of probability measure)

$$\implies \Pr(\,|X_n - X| > \epsilon) = \Pr(A_n(\epsilon)) \le \Pr\left(\bigcup_{m=n}^{\infty} A_m(\epsilon)\right) \to 0 \text{ as } n \to \infty$$

$$\implies X_n \xrightarrow{P} X$$

Strength of Convergence

- $(X_n \xrightarrow{a.s.} X) \implies (X_n \xrightarrow{P} X) \implies (X_n \xrightarrow{D} X)$

Proof* $(X_n \xrightarrow{P} X \implies X_n \xrightarrow{D} X)$: Fix any $\epsilon > 0$. It holds that

$$\{X_n \leq x\} \subseteq \{X \leq x + \epsilon\} \cup \{|X_n - X| > \epsilon\} \implies F_{X_n}(x) \leq F_X(x+\epsilon) + \Pr(|X_n - X| > \epsilon)$$

$$\{X \leq x - \epsilon\} \subseteq \{X_n \leq x\} \cup \{|X_n - X| > \epsilon\} \implies F_X(x-\epsilon) \leq F_{X_n}(x) + \Pr(|X_n - X| > \epsilon)$$

$$\implies F_X(x-\epsilon) - \Pr(|X_n - X| > \epsilon) \leq F_{X_n}(x) \leq F_X(x+\epsilon) + \Pr(|X_n - X| > \epsilon)$$

Assume $X_n \xrightarrow{P} X$. Then $\Pr(|X_n - X| > \epsilon) \to 0$ as $n \to \infty$ for all $\epsilon > 0$. Therefore,

$$F_X(x-\epsilon) \leq \liminf_{n\to\infty} F_{X_n}(x) \leq \limsup_{n\to\infty} F_{X_n}(x) \leq F_X(x+\epsilon) \quad \text{for all } \epsilon > 0$$

Furthermore, if F_X is continuous at x, then

$$F_X(x-\epsilon) \uparrow F_X(x) \text{ and } F_X(x+\epsilon) \downarrow F_X(x) \quad \text{as } \epsilon \downarrow 0.$$

Condition for Almost Sure Convergence*

- If $\sum_{n=1}^{\infty} \Pr(|X_n - X| > \epsilon) < \infty$ for all $\epsilon > 0$, then $X_n \xrightarrow{a.s.} X$

Proof: For any $\epsilon > 0$, let $A_n(\epsilon) = \{|X_n - X| > \epsilon\}$. Then due to **Borel–Cantelli**: $\forall \epsilon > 0$

$$\sum_{n=1}^{\infty} A_n(\epsilon) < \infty \implies \Pr\left(\bigcap_{n=1}^{\infty}\bigcup_{m=n}^{\infty} A_m(\epsilon)\right) = \Pr\left(A_n(\epsilon) \text{ infinitely often}\right) = 0$$

$$\implies \Pr\left(\bigcup_{k=1}^{\infty}\bigcap_{n=1}^{\infty}\bigcup_{m=n}^{\infty} A_m(1/k)\right) = 0 \text{ by countable additivity}$$

$$\implies \Pr\left(\lim_{n\to\infty} X_n = X\right) = \Pr\left(\bigcap_{k=1}^{\infty}\bigcup_{n=1}^{\infty}\bigcap_{m=n}^{\infty} A_m^c(1/k)\right) = 1$$

$$\implies X_n \xrightarrow{a.s.} X$$

Almost Sure vs. In Probability Convergence*

- Let $\{X_n\}$ be independent Bernoulli trials with parameter $1/n$. Then

$$X_n \xrightarrow{P} 0, \text{ but it does not hold } X_n \xrightarrow{a.s.} 0$$

Proof: For any $\epsilon > 0$, $\Pr(|X_n| > \epsilon) \leq \Pr(X_n = 1) = \frac{1}{n} \to 0$ as $n \to \infty$ $\implies X_n \xrightarrow{P} 0$

$\{X_n\}$ are independent and $\sum_{n=1}^{\infty} \Pr(X_n = 1) = \sum_{n=1}^{\infty} \frac{1}{n} = \infty$, then by **Borel–Cantelli:**

$\Pr(X_n = 1 \text{ infinitely often}) = 1 \implies \Pr\left(\lim_{n\to\infty} X_n = 0\right) = 0 \implies X_n \xrightarrow{a.s.} 0$ does not hold

Coupling*

- <u>Skorokhod's representation theorem</u>:

  If $X_n \xrightarrow{D} X$, then there exist random variables $Y_1, Y_2, \ldots$ and Y on some $(\Omega', \mathscr{F}, \mathbb{P})$, satisfying $F_{X_n} = F_{Y_n}$ for all $n \geq 1$ and $F_X = F_Y$, such that $Y_n \xrightarrow{a.s.} Y$

Proof: Apply inverse transform sampling. Let $\Omega' = [0,1]$, $\mathscr{F}$ the Borel σ-field on $[0,1]$, and $\mathbb{P}$ the uniform law. For $u \in \Omega' = [0,1]$, let

$$Y_n(u) = \inf\{x \mid u \leq F_{X_n}(x)\} \text{ and } Y(u) = \inf\{x \mid u \leq F_X(x)\}$$

Due to inverse transform sampling, $F_{X_n} = F_{Y_n}$ for all $n \geq 1$ and $F_X = F_Y$.

It can also be verified that $Y_n(u) \to Y(u)$ for all points u of continuity of Y, meanwhile the set $D \subseteq [0,1]$ of discontinuities of Y is countable , thus $\mathbb{P}(D) = 0$, which implies

$$Y_n \xrightarrow{a.s.} Y$$

Continuous Mapping Theorem*

- <u>Continuous mapping theorem</u>: If $g : \mathbb{R} \to \mathbb{R}$ is continuous, then

$$X_n \xrightarrow{D} X \implies g(X_n) \xrightarrow{D} g(X)$$

$$X_n \xrightarrow{P} X \implies g(X_n) \xrightarrow{P} g(X)$$

$$X_n \xrightarrow{a.s.} X \implies g(X_n) \xrightarrow{a.s.} g(X)$$

Proof (for convergence in distribution):

Construct $\{Y_n\}$ and Y as in Skorokhod's representation theorem. By continuity of g,

$$Y_n(u) \to Y(u) \implies g(Y_n(u)) \to g(Y(u)) \implies g(Y_n) \xrightarrow{a.s.} g(Y) \implies g(X_n) \xrightarrow{D} g(X)$$

Other Convergence Modes*

- $X_n \xrightarrow{1} X$ (convergence <u>in mean</u>) if

$$\lim_{n\to\infty} \mathbb{E}\left[|X_n - X|\right] = 0$$

- $X_n \xrightarrow{r} X$ (convergence <u>in rth mean</u> / <u>in the L^r-norm</u>) if

$$\lim_{n\to\infty} \mathbb{E}\left[|X_n - X|^r\right] = 0$$

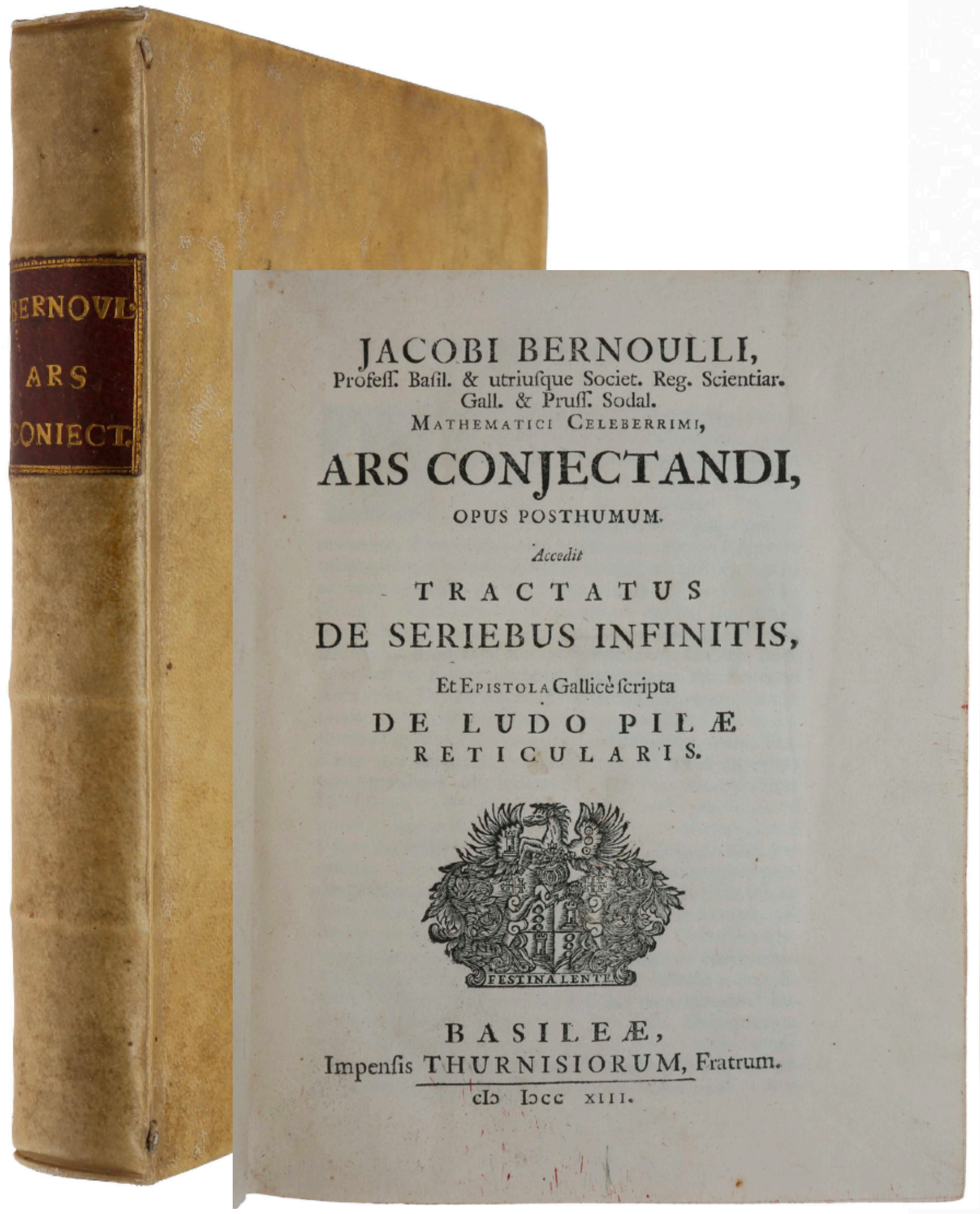
$$(X_n \xrightarrow{a.s.} X) \implies (X_n \xrightarrow{P} X) \implies (X_n \xrightarrow{D} X)$$

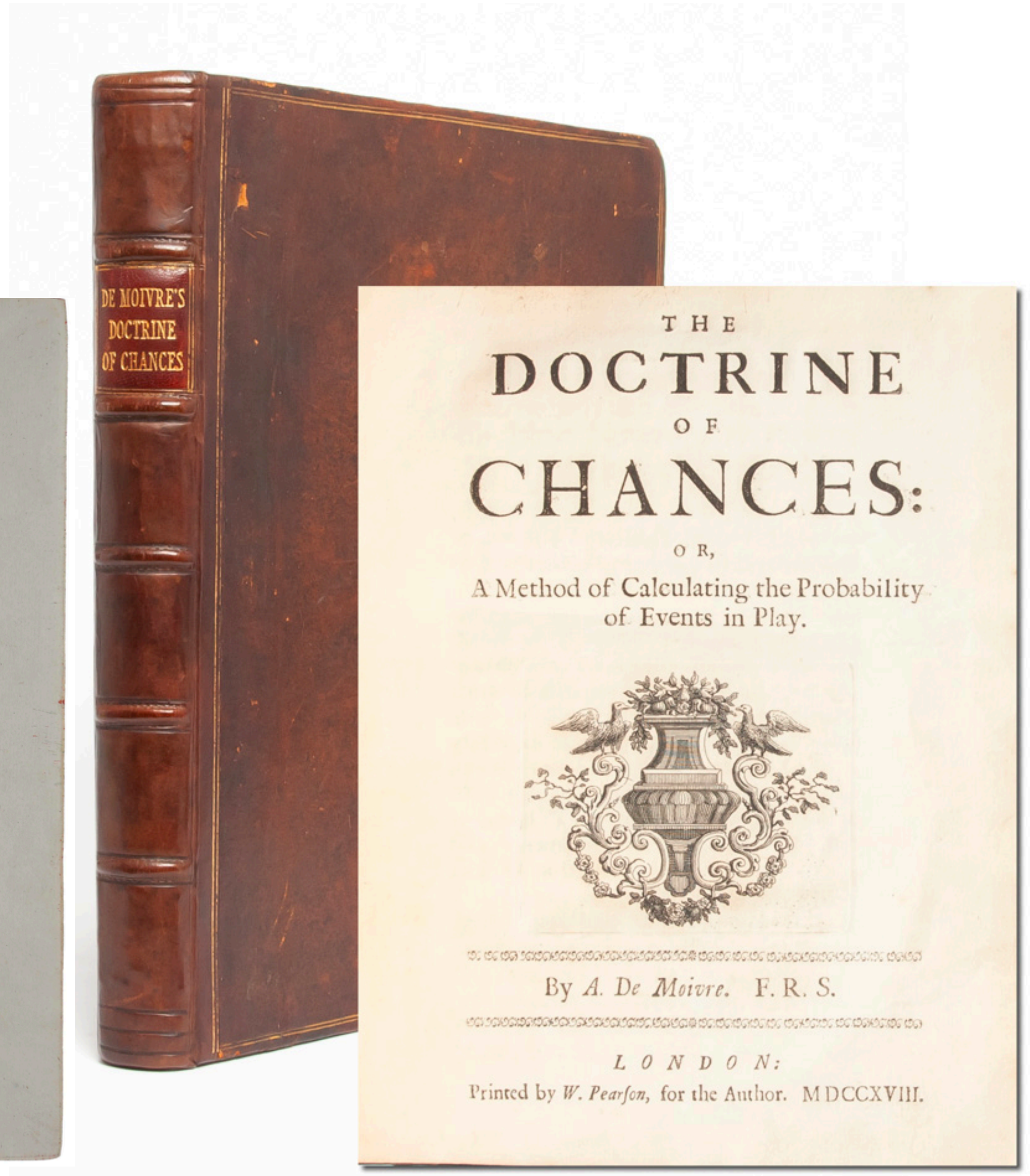
$$\Uparrow$$

$$(X_n \xrightarrow{s} X) \implies (X_n \xrightarrow{r} X) \implies (X_n \xrightarrow{1} X)$$

(for $s \geq r \geq 1$)

LLN and CLT


ARS
JACOBI BERNOULLI,
Profeff. Bafil. & utriufque Societ. Reg. Scientiar.
Gall. & Pruff. Sodal.
MATHEMATICI CELEBERRIMI,
ARS CONJECTANDI,
OPUS POSTHUMUM.
Accedit
TRACTATUS
DE SERIEBUS INFINITIS,
Et EPISTOLA Gallicè fcripta
DE LUDO PILÆ
RETICULARIS.
BASILEÆ,
Impenfis THURNISIORUM, Fratrum.
ciↃ Iↄcc XIII.


DE MOIVRE'S
DOCTRINE
OF CHANCES
THE
DOCTRINE
OF
CHANCES:
OR,
A Method of Calculating the Probability
of Events in Play.
By A. De Moivre. F. R. S.
LONDON:
Printed by W. Pearfon, for the Author. MDCCXVIII.

Bernoulli's Law of Large Number

In *Ars Conjectandi* (1713)

- Let $X_1, X_2, \dots$ be *i.i.d.* Bernoulli trials with $\mathbb{E}[X_1] = p \in [0,1]$. Then

$$\Pr\left(\left| \frac{X_1 + X_2 + \cdots + X_n}{n} - p \right| > \epsilon \right) \to 0 \quad \text{as } n \to \infty \quad \text{for all } \epsilon > 0$$

i.e. $\overline{X}_n \xrightarrow{P} p$, where $\overline{X}_n$ is the sample mean $\overline{X}_n = \dfrac{X_1 + X_2 + \cdots + X_n}{n}$

Proof: By Chebyshev's inequality, $\Pr(|\overline{X}_n - p| > \epsilon) \le \dfrac{p(1-p)}{n\epsilon^2} \to 0$ as $n \to \infty$

(This is of course not the original proof of Bernoulli.)

Law of Large Numbers (LLN)

Let $X_1, X_2, \ldots$ be *i.i.d.* random variables with finite mean $\mathbb{E}[X_1] = \mu$.

And let $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ be the **sample mean.**

- **Weak law** (Khinchin's law) **of large number**:

$$\overline{X}_n \xrightarrow{P} \mu \text{ as } n \to \infty$$

- **Strong law** (Kolmogorov's law) **of large number**:

$$\overline{X}_n \xrightarrow{a.s.} \mu \text{ as } n \to \infty$$

(*The deviation* $|\overline{X}_n - \mu|$ *is always small for all sufficiently large* n)

Weak LLN Assuming Bounded Variance

- Let $X_1, X_2, \ldots$ be independent random variables with finite mean $\mathbb{E}[X_i] = \mu$ and finitely bounded variance $\mathbf{Var}[X_i] \leq \sigma^2$.

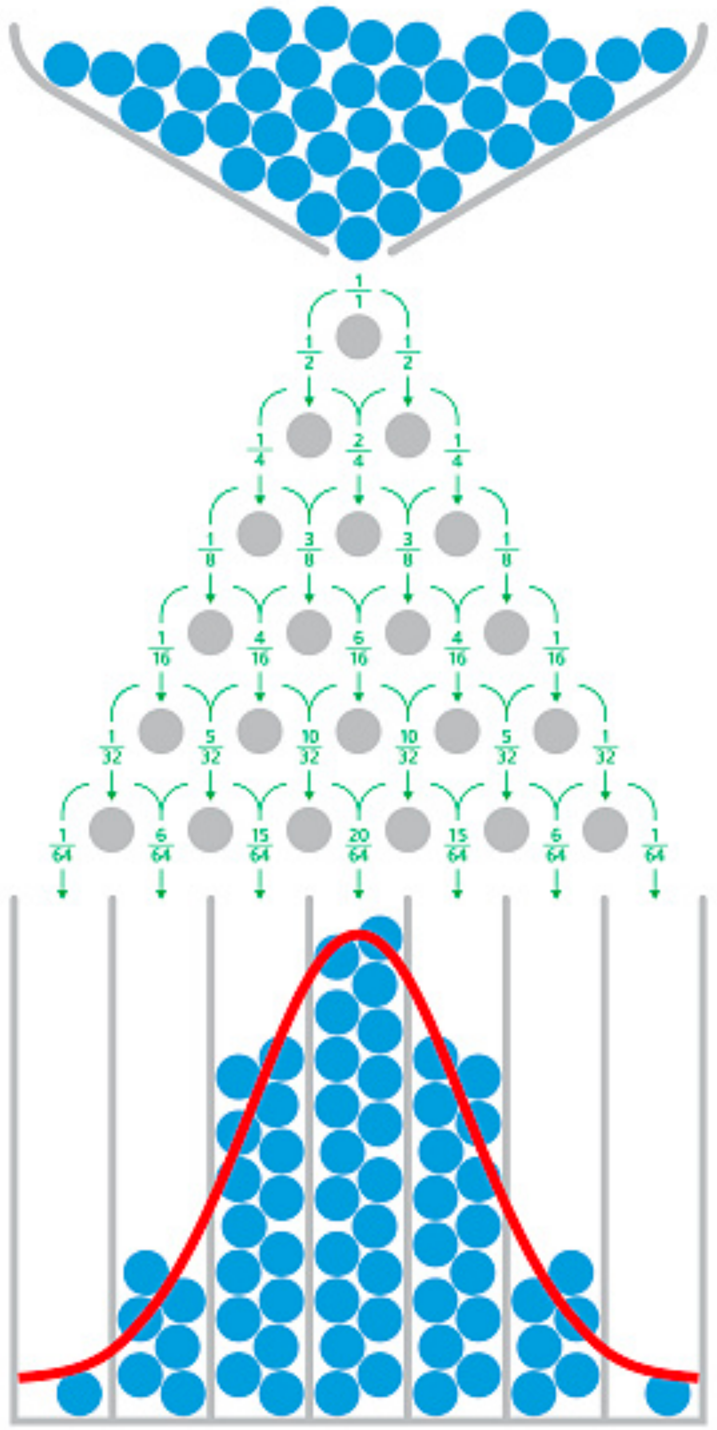
  Then the sample mean $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ has

$$\overline{X}_n \xrightarrow{P} \mu \text{ as } n \to \infty$$

Proof: By Chebysev's inequality, $\Pr(|\overline{X}_n - \mu| > \epsilon) \leq \frac{\sigma^2}{n\epsilon^2} \to 0$ as $n \to \infty$

De Moivre—Laplace Theorem

(棣莫弗－拉普拉斯定理)



- Let $p \in (0,1)$ and $X_n \sim B(n, p)$. Then its standardization

$$\frac{X_n - np}{\sqrt{np(1-p)}} \xrightarrow{D} N(0,1) \quad \text{as} \quad n \to \infty$$

- For any $p \in (0,1)$, any radius $r > 0$, and any $\epsilon > 0$, there is an n_0 such that for all $n > n_0$ and all k such that $\left|(k - np)/\sqrt{np(1-p)}\right| < r$,

$$\binom{n}{k} p^k (1-p)^{n-k} \in (1 \pm \epsilon) \frac{1}{\sqrt{2\pi np(1-p)}} \mathrm{e}^{-\frac{(k-np)^2}{2np(1-p)}}$$

Central Limit Theorem (CLT)

- Let $X_1, X_2, \ldots$ be *i.i.d.* random variables with $\mathbb{E}[X_1] = \mu$ and $\mathbf{Var}[X_1] = \sigma^2$.

And let $\overline{X}_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$ be the **sample mean.**

- **<u>Classical</u>** (Lindeberg–Lévy) **<u>central limit theorem</u>**:

$$\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{D} N(0,1) \quad \text{as } n \to \infty$$

Convergence Rate of CLT

(Berry–Esseen theorem)

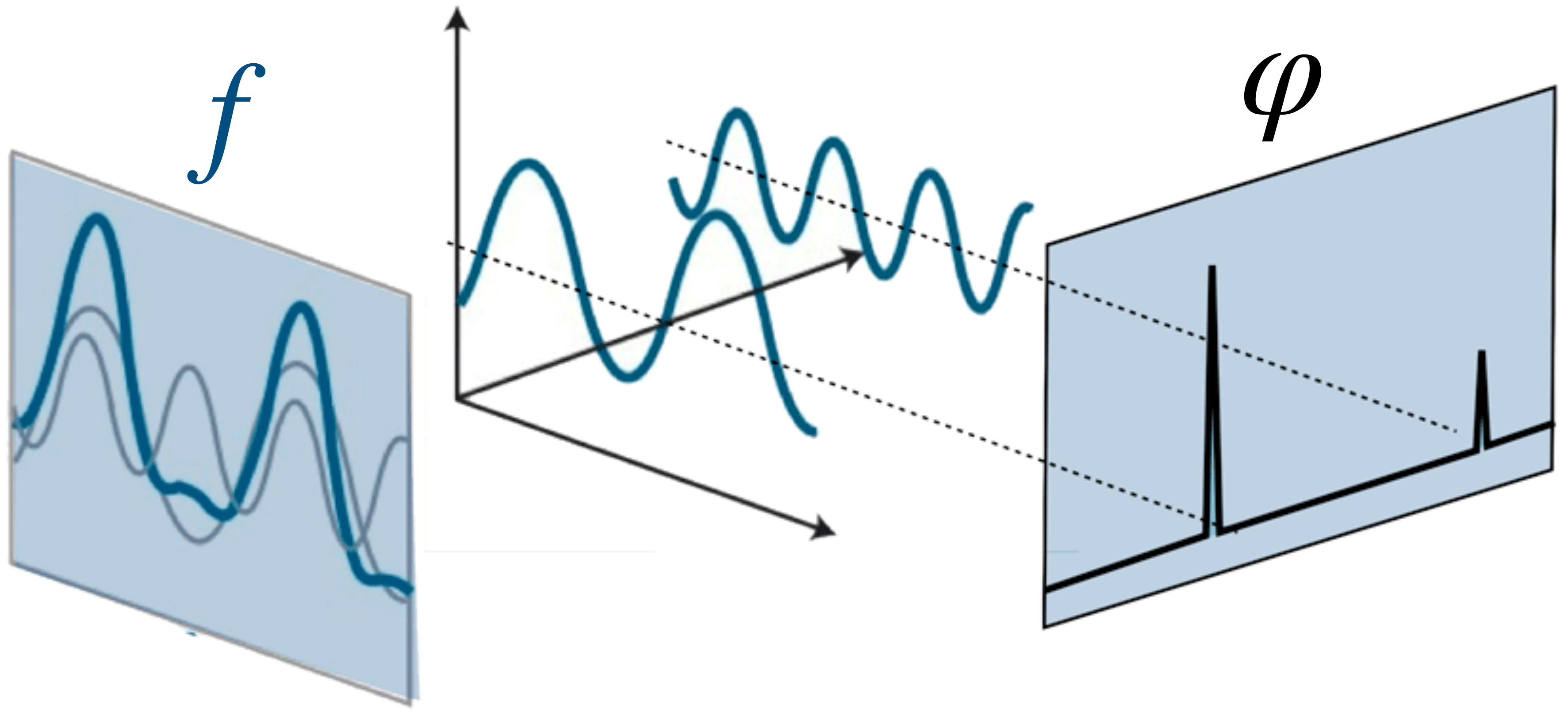
- Berry–Esseen theorem: Let $X_1, X_2, \ldots$ be *i.i.d.* random variables with $\mathbb{E}[X_1] = \mu$, $\mathbf{Var}[X_1] = \sigma^2$, and $\rho = \mathbb{E}[|X_1 - \mu|^3]$. And let $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.

  There is an absolute constant C, such that for any z

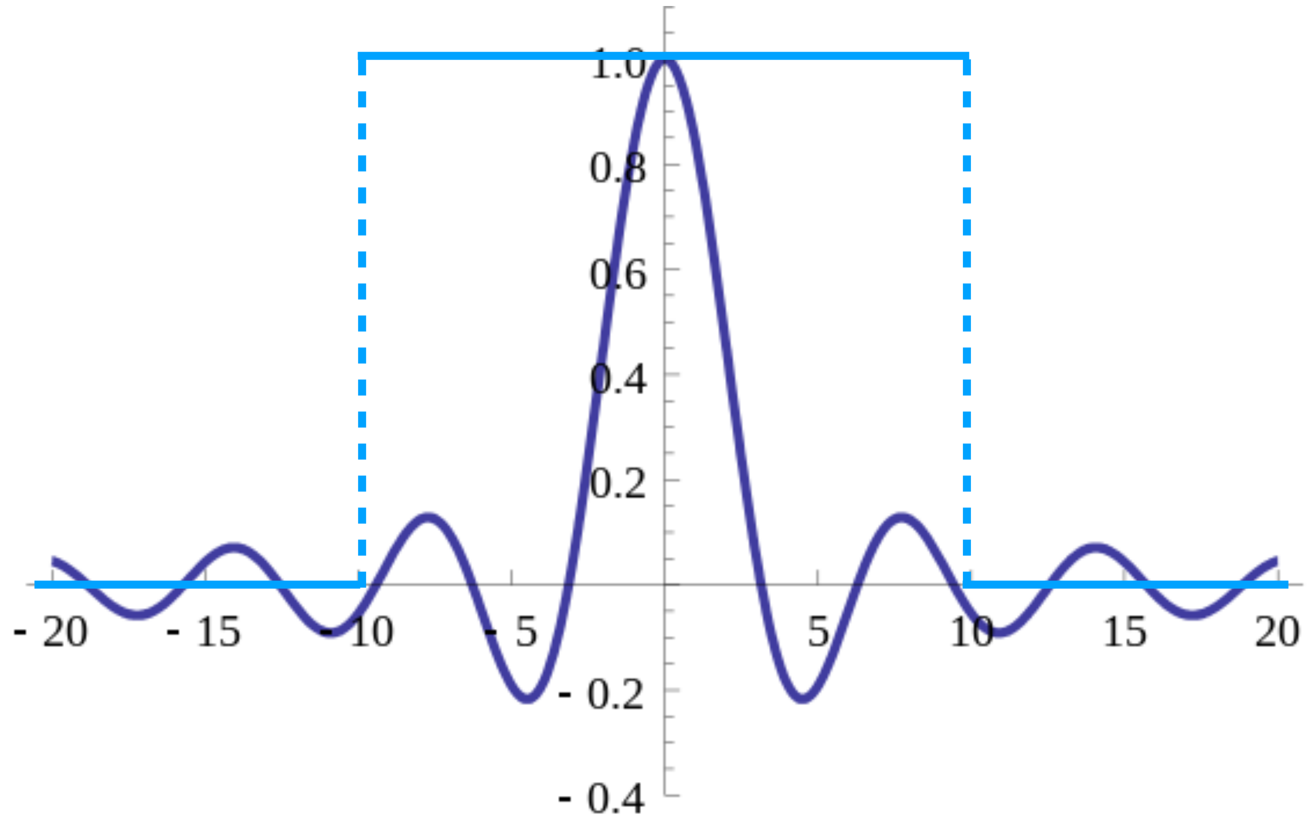
$$\left| \Pr\left(\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \leq z \right) - \Phi(z) \right| \leq \frac{C\rho}{\sigma^3\sqrt{n}}$$

  where Φ stands for the CDF for standard normal distribution $N(0,1)$

Characteristic Function



Characteristic Functions



- The moment generating function (MGF) of X is the function $M_X : \mathbb{R} \to \mathbb{R}_{\geq 0}$

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

- The characteristic function (特征函数) of X is the function $\varphi_X : \mathbb{R} \to \mathbb{C}$

$$\varphi_X(t) = \mathbb{E}[\mathrm{e}^{itX}], \text{ where } i = \sqrt{-1}$$

- Fourier transform: $\varphi_X(t) = \int \mathrm{e}^{itx}\, \mathrm{d}F_X(x) = \mathbb{E}[\cos tX] + i\mathbb{E}[\sin tX]$

- Unlike MGF, φ_X always exists and is finite, because $|\mathrm{e}^{itx}| = 1$

Boundedness of Characteristic Function

$$\varphi_X(t) = \mathbb{E}[\mathrm{e}^{itX}]$$

- $|\varphi_X(t)| \le 1$ for all $t \in \mathbb{R}$
- If $\mathbb{E}[|X^k|] < \infty$, then

$$\varphi_X(t) = \sum_{j=0}^{k} \frac{\mathbb{E}[X^j]}{j!}(it)^j + o(t^k)$$

$(\varphi_X(t) = 1 + i\mu t + o(t)\ )$

$(\varphi_X(t) = 1 + i\mu t - \frac{\sigma^2 t^2}{2} + o(t^2))$

Proof: $|\varphi_X(t)| \le \int |\mathrm{e}^{itx}| \, \mathrm{d}F_X(x) = \int \mathrm{d}F_X(x) = 1$ (for Lebesgue-Stieltjes integral)

Taylor's expansion: $\varphi_X(t) = \mathbb{E}[\mathrm{e}^{itX}] = \mathbb{E}\left[\sum_{j=0}^{k} \frac{X^j}{j!}(it)^j + o(t^k)\right] = \sum_{j=0}^{k} \frac{\mathbb{E}[X^j]}{j!}(it)^j + o(t^k)$

Normal Characteristic Function

=

- If $X \sim N(0,1)$, then

$$\varphi_X(t) = \mathbb{E}[\mathrm{e}^{itX}] = \mathrm{e}^{-t^2/2}$$

Proof (using complex integration):

$$\varphi_X(t) = \mathbb{E}[\mathrm{e}^{itX}] = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{itx - x^2/2}\, \mathrm{d}x = \mathrm{e}^{-t^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}(x-it)^2}\, \mathrm{d}x = \mathrm{e}^{-t^2/2}$$

because $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}(x-it)^2}\, \mathrm{d}x = 1$ via contour integration

Normal Characteristic Function

=

- If $X \sim N(0,1)$, then

$$\varphi_X(t) = \mathbb{E}[\mathrm{e}^{itX}] = \mathrm{e}^{-t^2/2}$$

Proof (without using complex integration): $\varphi_X(t) = \mathbb{E}[\mathrm{e}^{itX}] = \mathbb{E}[\cos tX] + i\mathbb{E}[\sin tX]$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \cos(tx)\mathrm{e}^{-x^2/2}\,\mathrm{d}x + \frac{i}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \underbrace{\sin(tx)\mathrm{e}^{-x^2/2}}_{\text{odd function}}\,\mathrm{d}x = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \cos(tx)\mathrm{e}^{-x^2/2}\,\mathrm{d}x$$

$$\frac{\mathrm{d}\varphi_X(t)}{\mathrm{d}t} = \mathbb{E}\left[\frac{\mathrm{d}\mathrm{e}^{itX}}{\mathrm{d}t}\right] = \mathbb{E}[iX\mathrm{e}^{itX}] = i\mathbb{E}[\underbrace{X\cos tX}_{\text{odd function}}] - \mathbb{E}[X\sin tX] = \frac{-1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x\sin(tx)\mathrm{e}^{-x^2/2}\,\mathrm{d}x$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sin(tx)\,\mathrm{d}\mathrm{e}^{-x^2/2} = \frac{1}{\sqrt{2\pi}} \sin(tx)\mathrm{e}^{-x^2/2}\Big|_{-\infty}^{\infty} - \frac{t}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \cos(tx)\mathrm{e}^{-x^2/2}\,\mathrm{d}x = -\frac{t}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \cos(tx)\mathrm{e}^{-x^2/2}\,\mathrm{d}x$$

$$\Longrightarrow \quad \frac{\mathrm{d}\varphi_X(t)}{\mathrm{d}t} = -t\varphi_X(t) \quad \Longrightarrow \quad \varphi_X(t) = \mathrm{e}^{-t^2/2}$$ (solving the ODE subject to $\varphi_X(0) = \mathbb{E}[\mathrm{e}^{i\cdot 0\cdot X}] = 1$)

Linear Transformation

- If X and Y are independent, then $\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$
- If $Y = aX + b$ for $a, b \in \mathbb{R}$, then $\varphi_Y(t) = \mathrm{e}^{itb}\varphi_X(at)$

Proof: For independent X and Y,

$$\varphi_{X+Y}(t) = \mathbb{E}[\mathrm{e}^{it(X+Y)}] = \mathbb{E}[\mathrm{e}^{itX}]\mathbb{E}[\mathrm{e}^{itY}] = \varphi_X(t)\varphi_Y(t)$$

For $Y = aX + b$,

$$\varphi_Y(t) = \mathbb{E}[\mathrm{e}^{it(aX+b)}] = \mathrm{e}^{itb}\mathbb{E}[\mathrm{e}^{itaX}] = \mathrm{e}^{itb}\varphi_X(at)$$

Continuity Theorem

- If X is **continuous** with density function f_X and characteristic function φ_X, then (by **Fourier inversion theorem**)

$$\varphi_X(t) = \int_{-\infty}^{\infty} \mathrm{e}^{itx} f_X(x)\,\mathrm{d}x \quad \text{and} \quad f_X(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{e}^{-itx}\varphi_X(t)\,\mathrm{d}t$$

  Hence, the distribution of continuous X is **uniquely identified** by φ_X.

- For **general** random variables: (*it's more complicated, but similarly*)

$$F_X = F_Y \text{ iff } \varphi_X = \varphi_Y$$

- <u>**Lévy's continuity theorem**</u>: Let $\{X_n\}$ and X be random variables.

$$X_n \xrightarrow{D} X \quad \text{iff} \quad \varphi_{X_n} \to \varphi_X \text{ pointwise on } \mathbb{R} \text{ as } n \to \infty$$

Convolution of Normal Distribution

- If $X \sim N(\mu, \sigma^2)$ and $Y \sim N(\nu, \tau^2)$ are independent, then
  - $X + Y \sim N(\mu + \nu, \sigma^2 + \tau^2)$

Proof (by characteristic function): Let $Z \sim N(0,1)$.

$$X \sim N(\mu, \sigma^2) \implies X = \sigma Z + \mu \implies \varphi_X(t) = \mathrm{e}^{it\mu}\varphi_Z(\sigma t) = \mathrm{e}^{it\mu - \sigma^2 t^2/2}$$

By the same calculation: $Y \sim N(\nu, \tau^2) \implies \varphi_Y(t) = \mathrm{e}^{it\nu - \tau^2 t^2/2}$

$$\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t) = \mathrm{e}^{it\mu - \sigma^2 t^2/2} \cdot \mathrm{e}^{it\nu - \tau^2 t^2/2} = \mathrm{e}^{it(\mu+\nu) - (\sigma^2 + \tau^2)t^2/2}$$

which is the characteristic function of normal distribution $N(\mu + \nu, \sigma^2 + \tau^2)$.

Law of Large Numbers (LLN)

Let $X_1, X_2, \ldots$ be *i.i.d.* random variables with finite mean $\mathbb{E}[X_1] = \mu$.

And let $\overline{X}_n = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ be the **sample mean.**

- **Weak law** (Khinchin's law) **of large number**:

$$\overline{X}_n \xrightarrow{P} \mu \text{ as } n \to \infty$$

- **Strong law** (Kolmogorov's law) **of large number**:

$$\overline{X}_n \xrightarrow{a.s.} \mu \text{ as } n \to \infty$$

Proof of the Weak Law of Large Numbers

- Let $X_1, X_2, \ldots$ be *i.i.d.* with finite mean $\mathbb{E}[X_1] = \mu$. Let $\overline{X}_n = \dfrac{1}{n}\sum_{j=1}^{n} X_j$
- The characteristic function $\varphi_{X_j}(t) = \mathbb{E}[\mathrm{e}^{itX_j}] = 1 + i\mu t + o(t)$

$$\implies \varphi_{\overline{X}_n}(t) = \varphi_{X_1+\cdots+X_n}(t/n) = \prod_{j=1}^{n} \varphi_{X_j}(t/n) = \left(1 + \frac{i\mu t}{n} + o\left(\frac{t}{n}\right)\right)^n$$

$$\to \mathrm{e}^{it\mu} \quad \text{for all } t \in \mathbb{R} \quad \text{as } n \to \infty$$

- Meanwhile, $\varphi_X(t) = \mathbb{E}[\mathrm{e}^{itX}] = \mathrm{e}^{it\mu}$ for constant $X = \mu$
- $\implies \overline{X}_n \xrightarrow{D} \mu$ by **Lévy's continuity theorem** $\implies \overline{X}_n \xrightarrow{P} \mu$ for constant μ

Central Limit Theorem (CLT)

- Let $X_1, X_2, \ldots$ be *i.i.d.* random variables with $\mathbb{E}[X_1] = \mu$ and $\mathbf{Var}[X_1] = \sigma^2$.

  Let $Z_n = \dfrac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}$ be the **standardized sample mean**, where $\overline{X}_n = \dfrac{1}{n}\sum_{i=1}^{n} X_i$.

- **Classical** (Lindeberg–Lévy) **central limit theorem**:

$$Z_n \xrightarrow{D} N(0,1) \quad \text{as } n \to \infty$$

Proof of the Central Limit Theorem

- Let $X_1, X_2, \ldots$ be *i.i.d.* with finite $\mathbb{E}[X_1] = \mu$ and $\mathbf{Var}[X_1] = \sigma^2$. Let $\overline{X}_n = \dfrac{1}{n}\sum_{j=1}^{n} X_j$
- For standardized $Y_j = (X_j - \mu)/\sigma \implies \varphi_{Y_j}(t) = \mathbb{E}[e^{itY_j}] = 1 - \frac{t^2}{2} + o(t^2)$
- The **standardized sample mean:** $Z_n = \dfrac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} = \dfrac{Y_1 + \cdots + Y_n}{\sqrt{n}}$

$$\implies \varphi_{Z_n}(t) = \varphi_{Y_1+\cdots+Y_n}\left(t/\sqrt{n}\right) = \prod_{j=1}^{n} \varphi_{Y_j}\left(t/\sqrt{n}\right) = \left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right)^n$$

$\to e^{-t^2/2}$ for all $t \in \mathbb{R}$ as $n \to \infty$ (characteristic function of $N(0,1)$)

- $\implies \overline{Z}_n \xrightarrow{D} Z \sim N(0,1)$ by **Lévy's continuity theorem**