

Probability Theory & Mathematical Statistics

Limit Theorems

Limit Theorems

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

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

- Law of large numbers (LLN): sample mean \rightarrow expectation

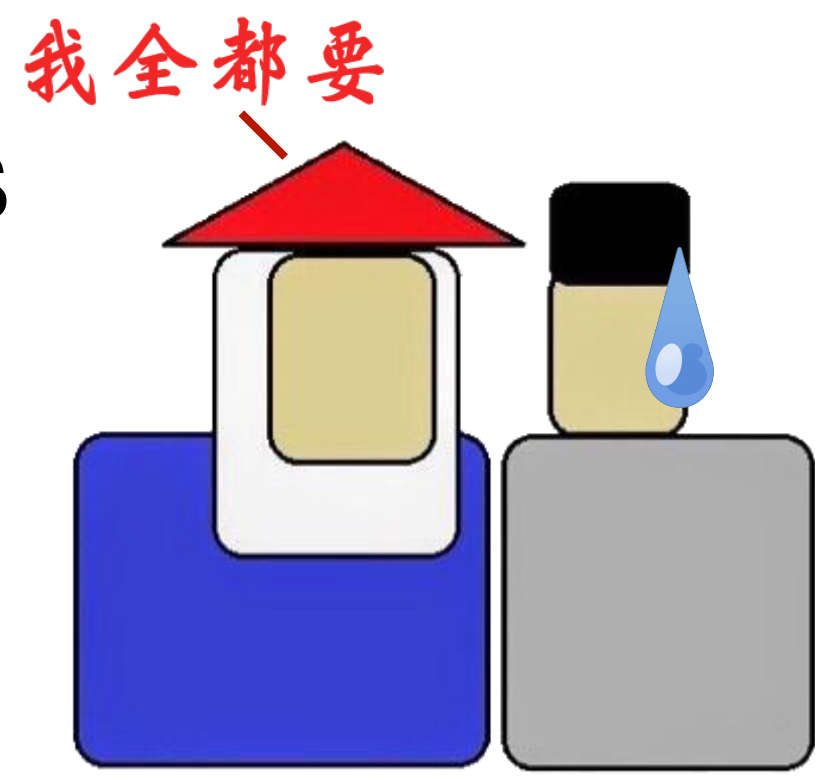
$$\bar{X}_n \longrightarrow \mu \quad \text{as } n \rightarrow \infty$$

- Central limit theorem (CLT): standardized sample mean \rightarrow standard normal

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

Convergence

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



Convergence of Random Variables

0.  $\rightarrow U_{[0,1]}$

Modes of Convergence

- Let $X, X_1, X_2, \dots : \Omega \rightarrow \mathbb{R}$ be random variables on prob. space (Ω, Σ, \Pr) .

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

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

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

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

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

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

Modes of Convergence

- Let X_1, X_2, \dots and X be random variables on probability space (Ω, Σ, \Pr) .
- $X_n \xrightarrow{D} X$ (convergence in distribution / in law / weak convergence of measure) if

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

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

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

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

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

$X_n \rightarrow X$
in measure

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

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

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

Convergence in Distribution

- Let X_1, X_2, \dots and X be random variables on probability space (Ω, Σ, \Pr) .
- $X_n \xrightarrow{D} X$ (convergence in distribution / in law / weak convergence of measure) if

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

$F_{X_n} \rightarrow F_X$ 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 weak convergence of measures

Convergence in Probability

- Let X_1, X_2, \dots and X be random variables on probability space (Ω, Σ, \Pr) .
- $X_n \xrightarrow{P} X$ (convergence in probability) if
$$\lim_{n \rightarrow \infty} \Pr(|X_n - X| > \epsilon) = 0 \quad \text{for all } \epsilon > 0$$

$X_n \rightarrow X$
in measure
- Functions $X_n : \Omega \rightarrow \mathbb{R}$ converges to $X : \Omega \rightarrow \mathbb{R}$ in measure \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:** $\Pr(|X_n - c| > \epsilon) = \Pr(X_n < c - \epsilon) + \Pr(X_n > \epsilon + c) \rightarrow 0$ if $X_n \xrightarrow{D} c$

Almost Sure Convergence

- Let X_1, X_2, \dots and X be random variables on probability space (Ω, Σ, \Pr) .

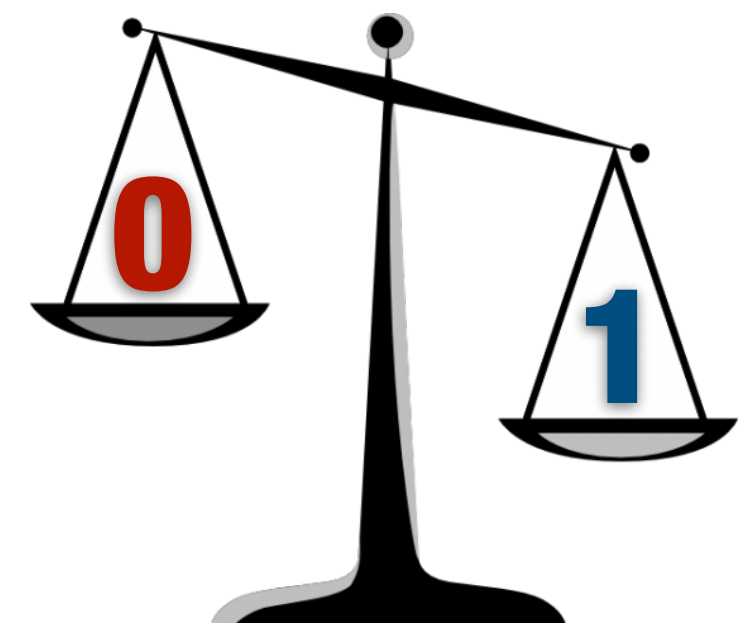
- $X_n \xrightarrow{a.s.} X$ (convergence almost surely / almost everywhere / w.p. 1) if

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

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

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

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, \dots be a sequence of events from a probability space (Ω, Σ, \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, \dots are independent $\implies \Pr(A) = 1$

Continuity of Probability Measures*

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

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

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

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

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

Continuity of Probability Measures*

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

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

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

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

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

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

- Proof:** Consider the complements $B_1^c \subseteq B_2^c \subseteq B_3^c \subseteq \dots$ 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 $\rightarrow 0$ as $n \rightarrow \infty$,

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

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

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, \dots are independent $\implies \Pr(A) = 1$

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

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

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

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 $\epsilon > 0$

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

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

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

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

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

Strength of Convergence

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

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 $\Pr(|X_n - X| > \epsilon) \rightarrow 0$ as $n \rightarrow \infty$ for all $\epsilon > 0$. Therefore,

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

Furthermore, if F_X is continuous at x , then

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

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: $\forall \epsilon > 0$

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

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

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

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

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 $\epsilon > 0$, $\Pr(|X_n| > \epsilon) \leq \Pr(X_n = 1) = \frac{1}{n} \rightarrow 0$ as $n \rightarrow \infty \implies X_n \xrightarrow{P} 0$

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

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

Coupling*

- Skorokhod's representation theorem:

If $X_n \xrightarrow{D} X$, then there exist random variables Y_1, Y_2, \dots and Y on some $(\Omega', \mathcal{F}, \mathbb{P})$, 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 $\Omega' = [0,1]$, \mathcal{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 $Y_n(u) \rightarrow Y(u)$ 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*

- Continuous mapping theorem: If $g : \mathbb{R} \rightarrow \mathbb{R}$ 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 ,

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

Other Convergence Modes*

- $X_n \xrightarrow{1} X$ (convergence in mean) if

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

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

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

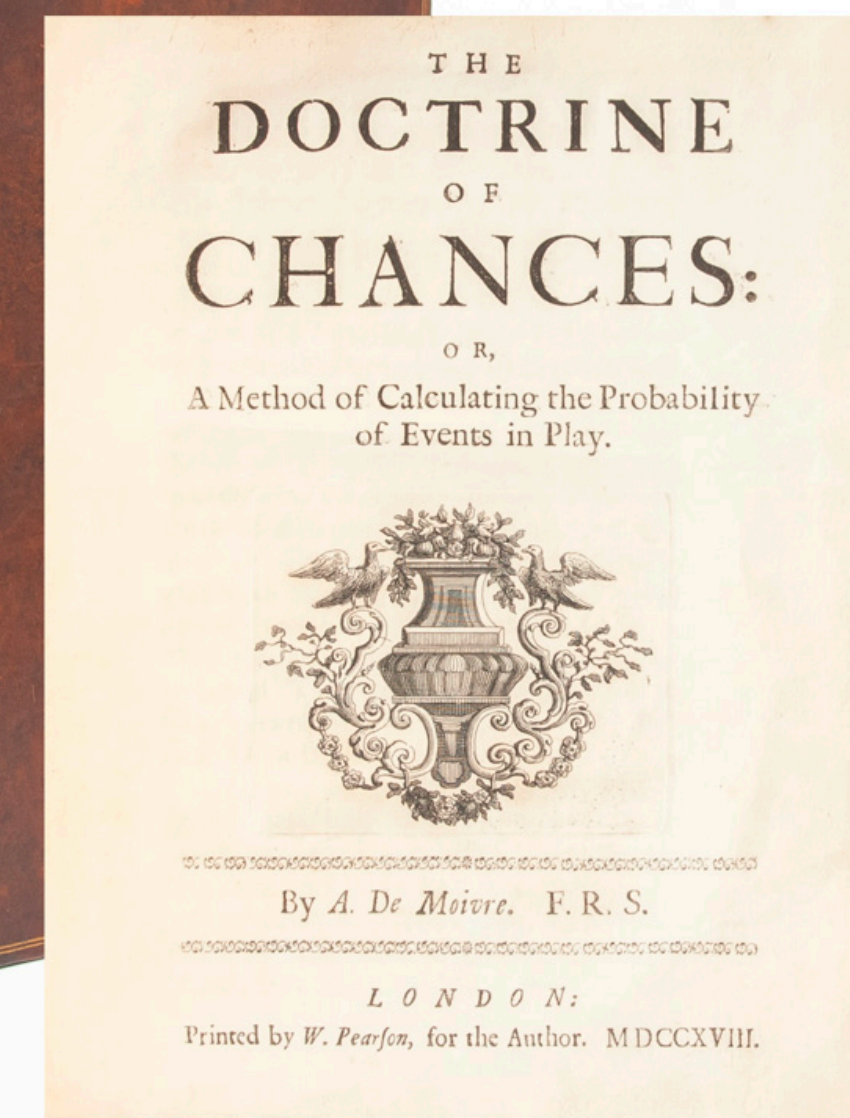
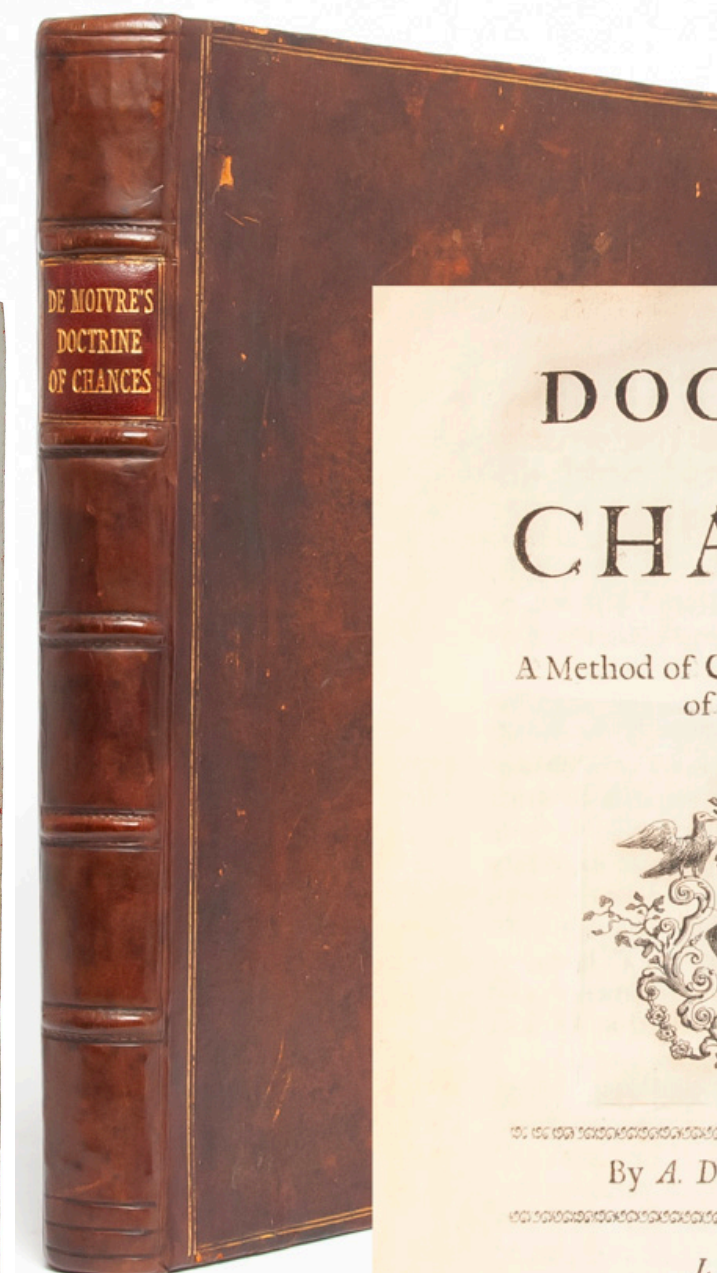
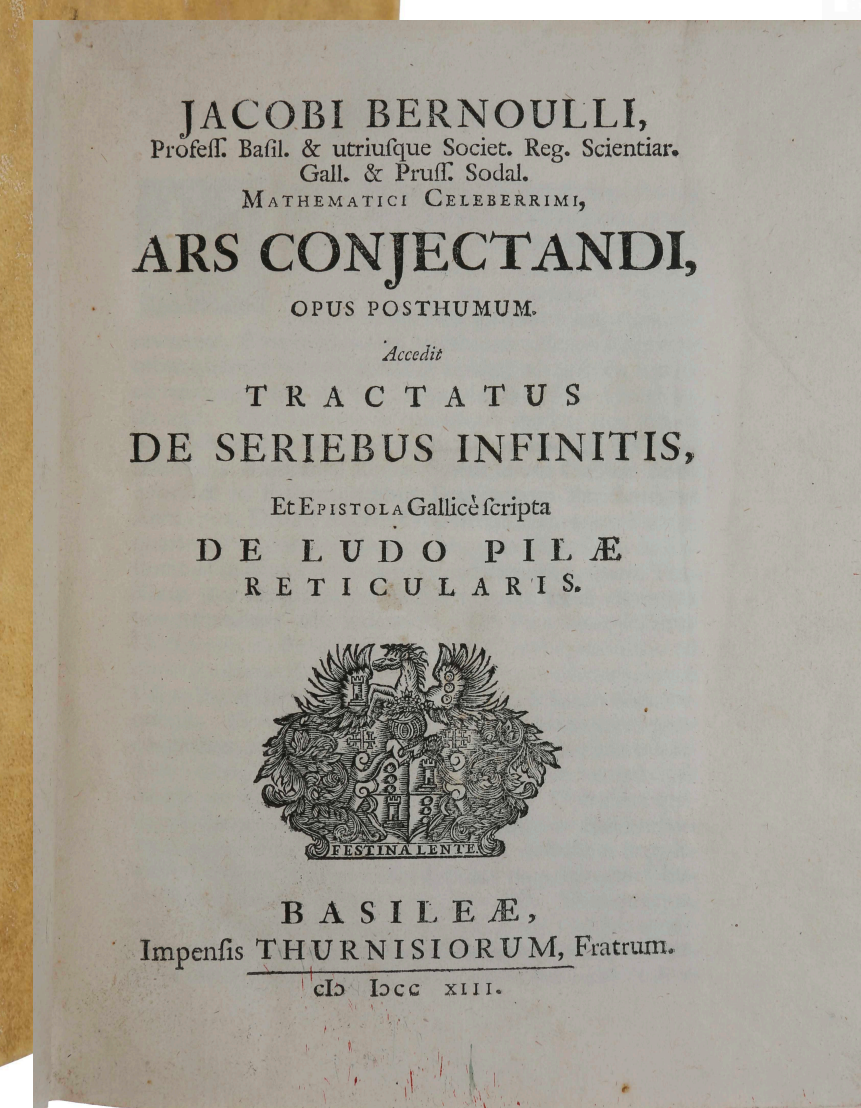
$$(X_n \xrightarrow{a.s.} X) \implies (X_n \xrightarrow{P} X) \implies (X_n \xrightarrow{D} X)$$

\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



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

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

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

Proof: By Chebyshev's inequality, $\Pr(|\bar{X}_n - p| > \epsilon) \leq \frac{p(1-p)}{n\epsilon^2} \rightarrow 0$ as $n \rightarrow \infty$

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



Law of Large Numbers (LLN)

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

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

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

$$\bar{X}_n \xrightarrow{P} \mu \text{ as } n \rightarrow \infty$$

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

$$\bar{X}_n \xrightarrow{a.s.} \mu \text{ as } n \rightarrow \infty$$

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

Weak LLN Assuming Bounded Variance

- Let X_1, X_2, \dots 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 $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ has

$$\bar{X}_n \xrightarrow{P} \mu \text{ as } n \rightarrow \infty$$

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

De Moivre–Laplace Theorem

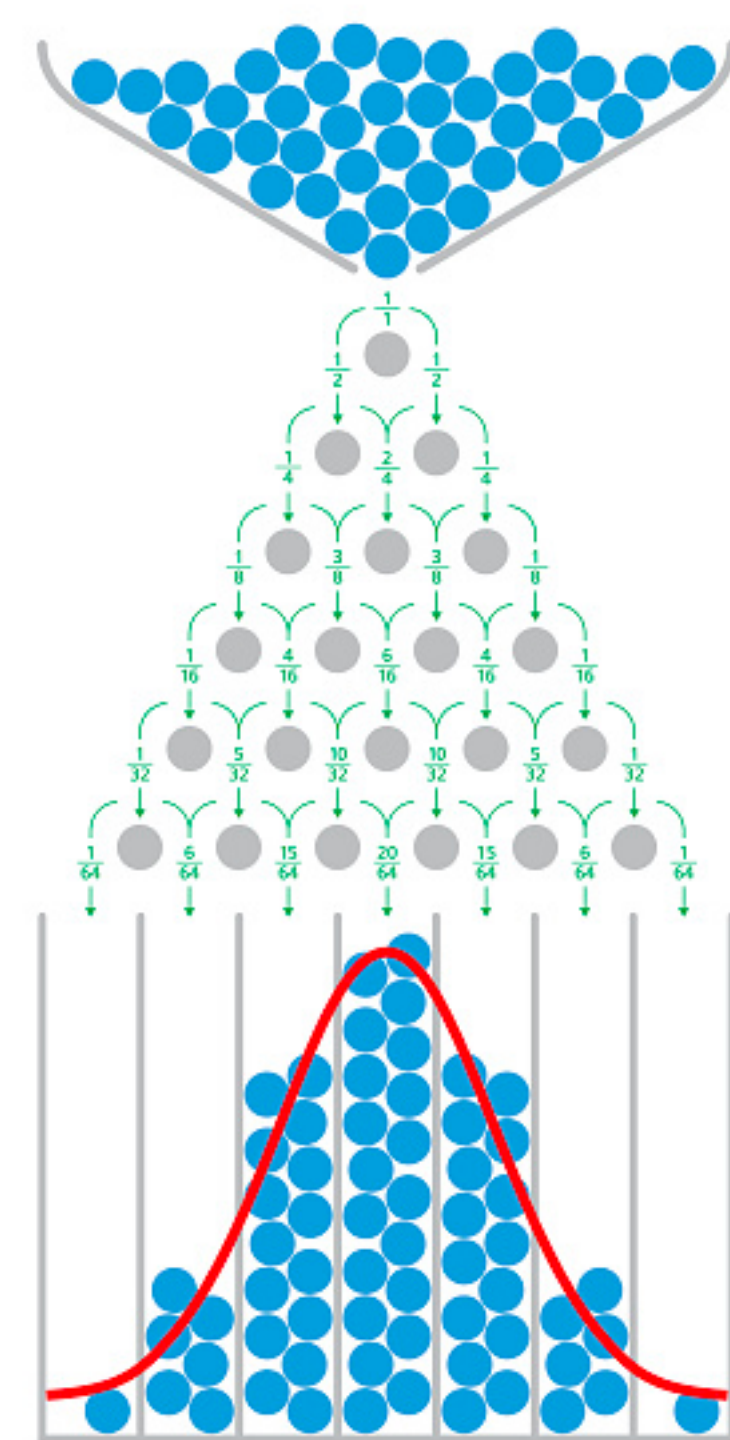
(棣莫弗–拉普拉斯定理)

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

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

- 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 $\left| (k - np) / \sqrt{np(1-p)} \right| < r$,

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



Central Limit Theorem (CLT)

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

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

- Classical (Lindeberg–Lévy) central limit theorem:

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

Convergence Rate of CLT

(Berry–Esseen theorem)

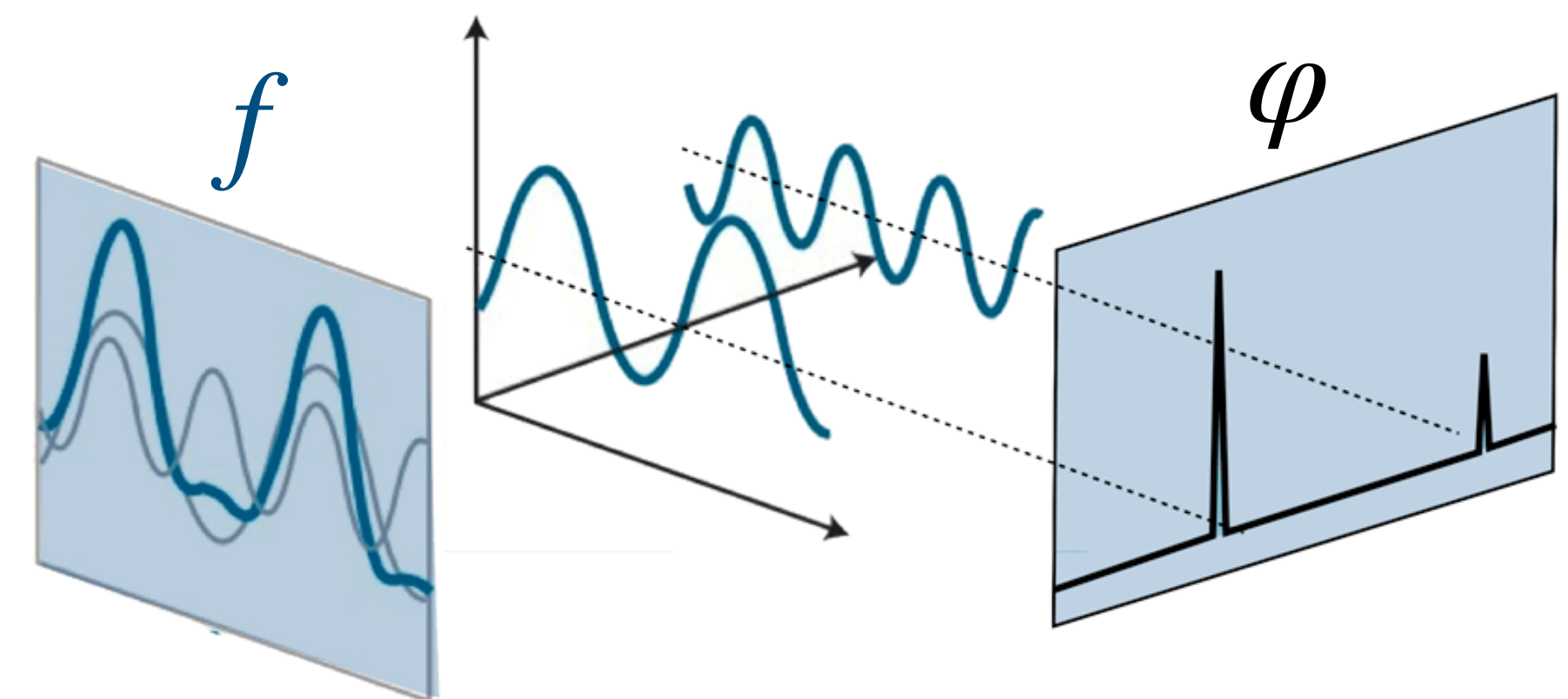
- Berry–Esseen theorem: Let X_1, X_2, \dots be *i.i.d.* random variables with $\mathbb{E}[X_1] = \mu$, $\mathbf{Var}[X_1] = \sigma^2$, and $\rho = \mathbb{E}[|X_1 - \mu|^3]$. And let $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$.

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

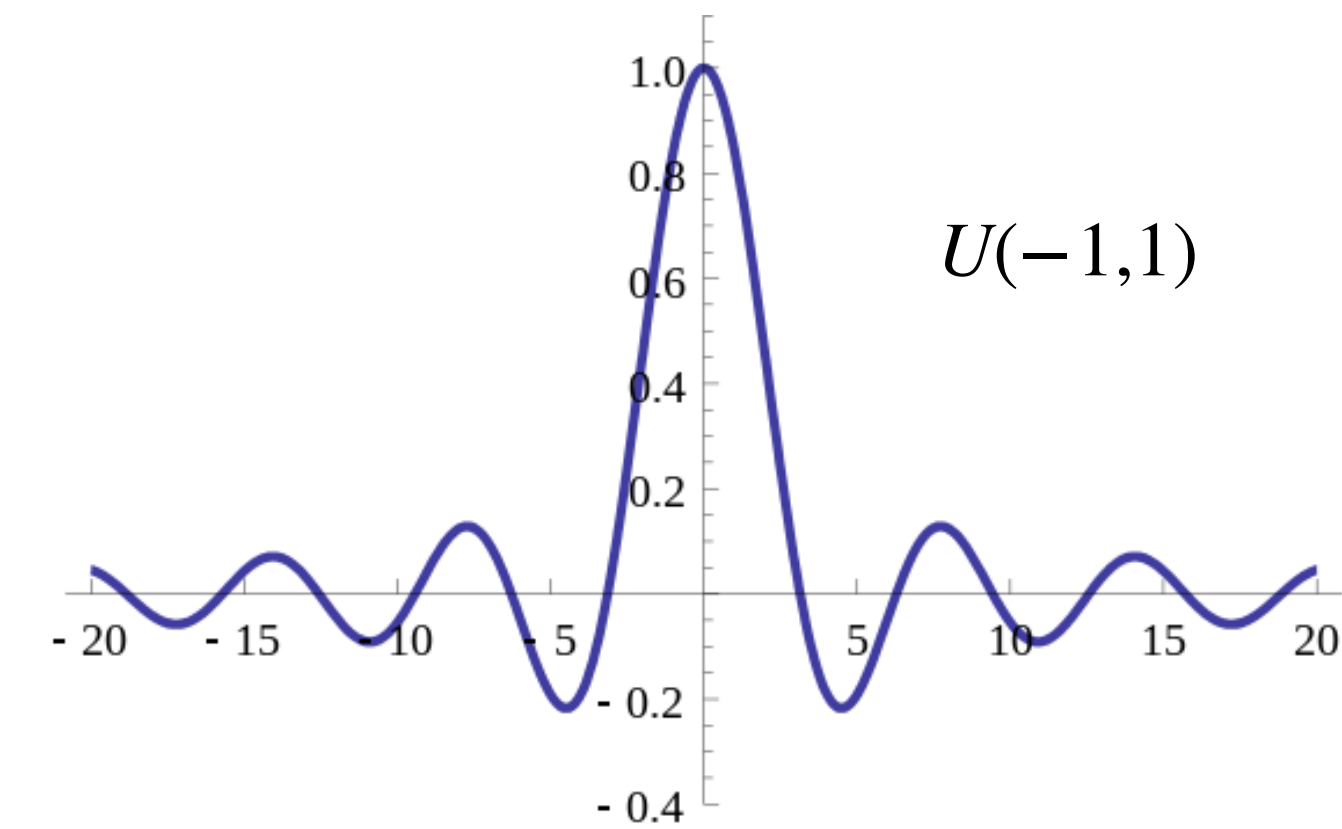
$$\left| \Pr \left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq z \right) - \Phi(z) \right| \leq \frac{C\rho}{\sigma^3\sqrt{n}}$$

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 $M_X : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$

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

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

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

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

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

Boundedness of Characteristic Function

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

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

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

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

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

Proof: $|\varphi_X(t)| \leq \int |e^{itx}| dF_X(x) = \int dF_X(x) = 1$ (for Lebesgue-Stieltjes integral)

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

Normal Characteristic Function



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

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

Proof (using complex integration):

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

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

Normal Characteristic Function

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

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



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

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \cos(tx)e^{-x^2/2} dx + \frac{i}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sin(tx)e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \cos(tx)e^{-x^2/2} dx$$

odd function

$$\frac{d\varphi_X(t)}{dt} = \mathbb{E} \left[\frac{de^{itX}}{dt} \right] = \mathbb{E}[iXe^{itX}] = i\mathbb{E}[X \cos tX] - \mathbb{E}[X \sin tX] = \frac{-1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x \sin(tx)e^{-x^2/2} dx$$

odd function

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

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

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 $\varphi_Y(t) = e^{itb}\varphi_X(at)$

Proof: For independent X and Y ,

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

For $Y = aX + b$,

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

Continuity Theorem

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

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

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

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

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

- Lévy's continuity theorem: Let $\{X_n\}$ and X be random variables.

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

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

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

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

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

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

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

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

$$\bar{X}_n \xrightarrow{P} \mu \text{ as } n \rightarrow \infty$$

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

$$\bar{X}_n \xrightarrow{a.s.} \mu \text{ as } n \rightarrow \infty$$

Proof of the Weak Law of Large Numbers

- Let X_1, X_2, \dots be *i.i.d.* with finite mean $\mathbb{E}[X_1] = \mu$. Let $\bar{X}_n = \frac{1}{n} \sum_{j=1}^n X_j$
- The characteristic function $\varphi_{X_j}(t) = \mathbb{E}[e^{itX_j}] = 1 + i\mu t + o(t)$
 $\implies \varphi_{\bar{X}_n}(t) = \varphi_{X_1 + \dots + X_n}(t/n) = \prod_{j=1}^n \varphi_{X_j}(t/n) = \left(1 + \frac{i\mu t}{n} + o\left(\frac{t}{n}\right)\right)^n$
 $\rightarrow e^{it\mu}$ for all $t \in \mathbb{R}$ as $n \rightarrow \infty$
- Meanwhile, $\varphi_X(t) = \mathbb{E}[e^{itX}] = e^{it\mu}$ for constant $X = \mu$
- $\implies \bar{X}_n \xrightarrow{D} \mu$ by Lévy's continuity theorem $\implies \bar{X}_n \xrightarrow{P} \mu$ for constant μ

Central Limit Theorem (CLT)

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

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

- Classical (Lindeberg–Lévy) central limit theorem:

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

Proof of the Central Limit Theorem

- Let X_1, X_2, \dots be *i.i.d.* with finite $\mathbb{E}[X_1] = \mu$ and $\mathbf{Var}[X_1] = \sigma^2$. Let $\bar{X}_n = \frac{1}{n} \sum_{j=1}^n X_j$
- 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: $Z_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{Y_1 + \dots + Y_n}{\sqrt{n}}$
 $\implies \varphi_{Z_n}(t) = \varphi_{Y_1 + \dots + Y_n} \left(t/\sqrt{n} \right) = \prod_{j=1}^n \varphi_{Y_j} \left(t/\sqrt{n} \right) = \left(1 - \frac{t^2}{2n} + o \left(\frac{t^2}{n} \right) \right)^n$
 $\rightarrow e^{-t^2/2}$ for all $t \in \mathbb{R}$ as $n \rightarrow \infty$ (characteristic function of $N(0,1)$)
- $\implies \bar{Z}_n \xrightarrow{D} Z \sim N(0,1)$ by Lévy's continuity theorem