Lecture note 9: Limit theorems

Foundation of Probability Theory/STA 203

Zhuosong ZHANG

Department of Statistics and Data Science, SUSTech

Fall, 2023

Introduction

Limit Theorems in Probability Theory

- The most important theoretical results in probability theory are limit theorems.
- Among which the most important are those classified either under the heading laws of large numbers or under the heading central limit theorems.
- The Law of Large Numbers: As the sample size increases, the sample mean converges to the true mean.
- Central Limit Theorem: As the sample size increases, the distribution of the sample mean approaches a normal distribution.

Example: Tossing a Fair Coin

- Toss a fair coin repeatedly, record the proportion of times it comes up heads.
- The Law of Large Numbers tells us that as the number of coin tosses increases, the proportion of heads will approach 0.5.
- If we toss the coin a large number of times, we should expect the proportion of heads to be very close to 0.5.

Example: Rolling a Fair Die

南方科技大学
SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

- Roll a fair six-sided die repeatedly, record the average value of the rolls.
- The Central Limit Theorem tells us that as the number of rolls increases, the distribution of the average values will become approximately normal.
- If we roll the die a large number of times and calculate the average value of the rolls, the distribution of these average values will be bell-shaped, with a mean close to 3.5 and a standard deviation that decreases as the number of rolls increases.

Example: Measuring Heights of Students

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

- Measure the heights of a large number of students, record the average height.
- The Central Limit Theorem tells us that as the sample size increases, the distribution of the average heights will become approximately normal.
- If we measure the heights of a large number of students and calculate the average height, the distribution of these average values will be bell-shaped, with a mean close to the true average height of the population and a standard deviation that decreases as the sample size increases.

Some important inequalities

Markov's inequality

Proposition 1 (Markov's inequality)

If X is a random variable that takes only nonnegative values, then

$$\mathbb{P}\{X \geqslant a\} \leqslant \frac{\mathbb{E}[X]}{a} \quad \text{for all } a > 0.$$

Figure: A. A. Markov (马尔可夫, 1856–1922)

Proof.

Note that X is nonnegative, then,

$$\begin{aligned}\mathbb{P}(X \geqslant a) &= \mathbb{E}[\mathbb{1}_{\{X \geqslant a\}}] \\ &\leqslant \mathbb{E}\left[\frac{X}{a}\right] \\ &= \frac{\mathbb{E}[X]}{a}.\end{aligned}$$

■

Chebyshev's inequality

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Proposition 2

If X is a random variable with finite mean μ and variance σ^2, then, for any value $a > 0$,

$$\mathbb{P}\{|X - \mu| \geqslant a\} \leqslant \frac{\sigma^2}{a^2}.$$

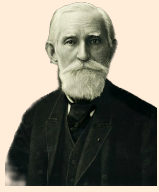


Figure: Chebyshev (切比雪夫)

Proof.

We will apply Markov's inequality to prove it. Note that

$$\mathbb{P}\{|X - \mu| \geqslant a\} = \mathbb{P}\{|X - \mu|^2 \geqslant a^2\} \leqslant \frac{\mathbb{E}[(X - \mu)^2]}{a^2} = \frac{\mathrm{Var}(X)}{a^2}.$$

■

Examples

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Example 3

Suppose that it is known that the number of items produced in a factory during a week is a random variable with mean 50.

(a) What can be said about the probability that this week's production will exceed 75?

(b) If the variance of a week's production is known to equal 25, then what can be said about the probability that this week's production will be between 40 and 60?

Solution.

Let X be the number of items produced during a weak. Then, we are going to estimate $\mathbb{P}\{X > 75\}$ and $\mathbb{P}\{40 < X < 60\}$.

Examples

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Solution (Cont'd).

(a) As X is nonnegative, by Markov's equality,

$$\mathbb{P}\{X > 75\} \leqslant \frac{\mathbb{E}[X]}{75} = \frac{50}{75} \approx 0.667.$$

(b) By Chebyshev's inequality, it follows that

$$\mathbb{P}\{|X - 50| \geqslant 10\} \leqslant \frac{\mathrm{Var}(X)}{10^2} = \frac{25}{100} = 0.25,$$

and therefore,

$$\mathbb{P}\{40 < X < 60\} \geqslant 1 - 0.25 = 0.75.$$

■

Degenerate random variables

南方科技大学
SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Proposition 4

If $\mathrm{Var}(X) = 0$, then

$$\mathbb{P}\{X = \mathbb{E}[X]\} = 1.$$

In other words, the only random variables having variances equal to 0 are those which are constant with probability 1.

Weak law of large numbers

Convergence of a sequence of numbers

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Definition 5

A sequence $\{a_n\}$ has the limit a, written as

$$\lim_{n\to\infty} a_n = a \quad \text{or} \quad a_n \to a \text{ as } n \to \infty$$

if for every $\varepsilon > 0$, there exists a corresponding integer N such that

$$n > N \implies |a_n - a| < \varepsilon.$$

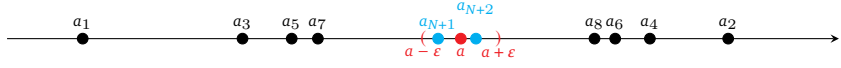


Figure: Convergence of a sequence

Convergence of a sequence of numbers

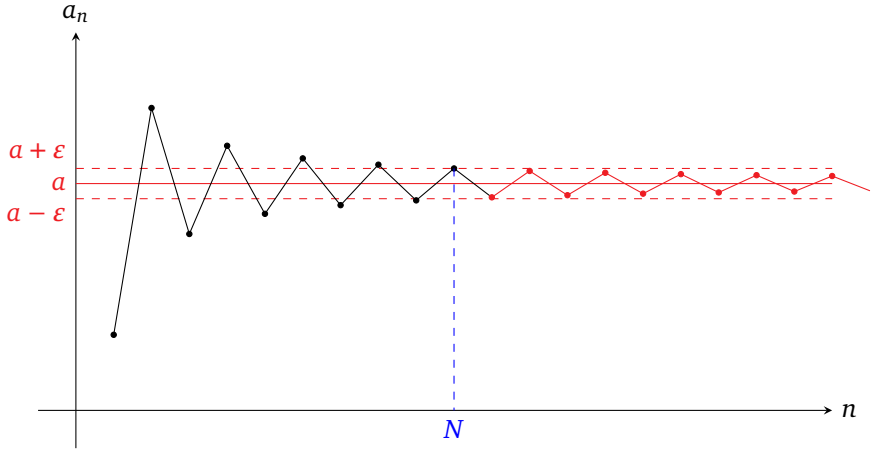


Figure: Convergence of a sequence

Convergence of a sequence of random variables

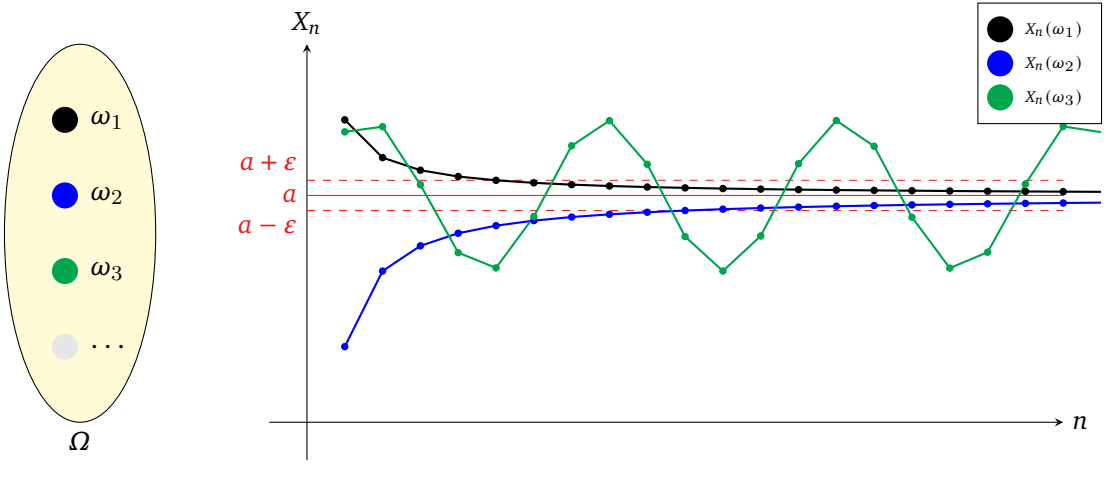


Figure: Convergence of a sequence of random variables. In this picture, $X_n(\omega_1)$ and $X_n(\omega_2)$ converges to a as $n \to \infty$.

Convergence in probability

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Definition 6

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $\{X_n\}$ be a sequence of random variables on it. We say X_n converges in probability to X, written as $X_n \xrightarrow{p} X$ as $n \to \infty$, if for every $\varepsilon > 0$,

$$\mathbb{P}\{|X_n - X| > \varepsilon\} \to 0 \quad \text{as } n \to \infty$$

Remark

Note that

$$\mathbb{P}\{|X_n - X| > \varepsilon\} = \mathbb{P}\{\omega : |X_n(\omega) - X(\omega)| > \varepsilon\}.$$

We can imagine that each ω is a trial, or an experiment, and we consider $|X_n - X| \leqslant \varepsilon$ is a success of convergence. Then, convergence in probability can be understood as the frequency of failures is very small.

Examples

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Example 7

Let $X_1, X_2, \ldots$ be i.i.d. random variables with $X_n \sim \text{Uniform}(0,1)$. Define $Y_n = \min(X_1, \ldots, X_n)$. Show that Y_n converges in probability to 0.

Proof.

For any $\varepsilon > 0$,

$$\begin{aligned}\mathbb{P}\{|Y_n - 0| > \varepsilon\} &= \mathbb{P}\{\min(X_1, \ldots, X_n) > \varepsilon\} \\ &= \mathbb{P}\{\text{all of } X_i\text{'s are greater than } \varepsilon\} \\ &= \prod_{i=1}^{n} \mathbb{P}\{X_i > \varepsilon\} = (1-\varepsilon)^n \to 0 \quad \text{as } n \to \infty.\end{aligned}$$

Therefore, $Y_n \xrightarrow{p} 0$ as $n \to \infty$. ■

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Example 8

Let $X \sim \text{Exp}(1)$ and $Y_n = X/n$. Show that Y_n converges to 0 in probability.

Proof.

For any $\varepsilon > 0$,

$$\begin{aligned}\mathbb{P}\{|Y_n - 0| \geqslant \varepsilon\} &= \mathbb{P}\{X \geqslant n\varepsilon\} \\ &= \int_{n\varepsilon}^{\infty} e^{-x} dx \\ &= e^{-n\varepsilon} \to 0 \quad \text{as } n \to \infty.\end{aligned}$$

Therefore, $Y_n \xrightarrow{p} 0$ as $n \to \infty$. ■

Properties

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Theorem 9

Let $\{X_n\}$ and $\{Y_n\}$ be two sequences of random variables, and suppose that

$$X_n \xrightarrow{p} a, \quad Y_n \xrightarrow{p} b,$$

then

(a) $X_n \pm Y_n \xrightarrow{p} a \pm b$,

(b) $X_n Y_n \xrightarrow{p} ab$, and

(c) if $b \neq 0$, $\frac{X_n}{Y_n} \xrightarrow{p} \frac{a}{b}$.

Theorem 10

Let $r > 0$ be a positive number. If $\mathbb{E}[|X|^r] < \infty$, $\mathbb{E}[|X_n|^r] < \infty$ for each $n \geqslant 1$, and $\mathbb{E}[|X_n - X|^r] \to 0$ as $n \to \infty$, then

$$X_n \xrightarrow{p} X.$$

Weak law of large numbers

- Let $\xi, \xi_1, \xi_2, \ldots$ be a sequence of i.i.d. random variables, with expected value μ.
- The sample mean is defined by

$$X_n := \bar{\xi}_n = \frac{1}{n} \sum_{i=1}^{n} \xi_i.$$

- In practice, we "think" that X_n will converge to the true expected value μ.
- However, the value of X_n may differ.
- If we are extremely unlucky , $|X_n - \mu|$ may be large: that is, $|X_n - \mu| > \varepsilon$ for some given $\varepsilon > 0$.
- With the help of convergence in probability , we can describe this phenomenon.

Weak law of large numbers

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Theorem 11 (Weak law of large numbers)

Let $\xi_1, \xi_2, \ldots$ be a sequence of independent and identically distributed random variables, each having the finite mean $\mu = \mathbb{E}[\xi_i]$. Then,

$$\bar{\xi}_n \xrightarrow{p} \mu \quad \text{as } n \to \infty,$$

in other words, for any $\varepsilon > 0$,

$$\mathbb{P}\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i - \mu\right| > \varepsilon\right\} \to 0 \quad \text{as } n \to \infty.$$

Bernoulli
伯努利
1600s

Poisson
泊松
1837

Markov
马尔可夫
around 1900

Khinchin
辛钦
1929

Time

- 1600s: Bernoulli proved a special form of LLN for binary random variables, named it as “Golden theorem”.
- 1837: Poisson described it under the name “la loi des grands nombres”.
- 1900: Markov showed that the law can apply to a weaker condition (second moment is not necessary).
- 1929: “Finite mean” condition is enough for i.i.d. case.

Proof

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Proof.

We only prove the theorem under the additional assumption that $\mathrm{Var}(\xi_1) = \sigma^2$ is finite. Now,

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\xi_i\right] = \mu, \quad \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i\right) = \frac{\sigma^2}{n},$$

and it follows from the Chebyshev's inequality that

$$\mathbb{P}\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i - \mu\right| > \varepsilon\right\} \leqslant \frac{\sigma^2}{n\varepsilon^2} \to 0 \quad \text{as } n \to \infty.$$

For the proof under a weaker condition, we omit the details. ■

Example 12 (Probability and Frequency)

Consider an event A defined in the context of some probabilistic experiment. Let $p = \mathbb{P}(A)$ be the probability of this event. We consider n independent repetitions of the experiment, and let M_n be the fraction of time that event A occurs; in this context, M_n is often called the empirical frequency of A. Note that

$$M_n = \frac{1}{n} \sum_{i=1}^{n} \xi_i.$$

where ξ_i is 1 whenever A occurs, and 0 otherwise; in particular, $\mathbb{E}[\xi_i] = p$. The weak law applies and shows that when n is large, the empirical frequency is most likely to be within ε of p. Loosely speaking, this allows us to conclude that empirical frequencies are faithful estimates of p. Alternatively, this is a step towards interpreting the probability p as the frequency of occurrence of A.

Multiple-choice Questions

1. Markov's Inequality, $\mathbb{P}(X \geqslant a) \leqslant \frac{\mathbb{E}[X]}{a}$, is applicable when:

   A. X is a normally distributed random variable.

   B. X is a continuous random variable.

   C. X is a discrete random variable.

   D. X is a non-negative random variable and $a > 0$.

Multiple-choice Questions

2. The sequence of random variables $X_1, X_2, X_3, \ldots$ converges in probability to a constant c if:

   A. For all $\varepsilon > 0$, $\mathbb{P}(|X_n - c| > \varepsilon)$ approaches 0 as n approaches infinity.

   B. For all $\varepsilon > 0$, $\mathbb{P}(|X_n - c| > \varepsilon)$ approaches 1 as n approaches infinity.

   C. For all $\varepsilon > 0$, $\mathbb{E}(|X_n - c| > \varepsilon)$ approaches 0 as n approaches infinity.

   D. For all $\varepsilon > 0$, $\mathbb{E}(|X_n - c| > \varepsilon)$ approaches 1 as n approaches infinity.

Multiple-choice Questions

3. Which of the following statements is correct regarding Chebyshev's inequality?

   A. It provides an upper bound on the probability that the absolute deviation of a random variable from its mean is more than k standard deviations.

   B. It only applies to normally distributed random variables.

   C. It provides a lower bound on the probability that a random variable takes on values within k standard deviations of its mean.

   D. It states that the sum of the probabilities of all possible outcomes of a random variable is equal to 1.

True/False Questions

4. If X is a random variable with finite mean μ and variance σ^2, then for any $k > 0$, Chebyshev's Inequality states that $\mathbb{P}(|X - \mu| \geqslant k\sigma) \leqslant \frac{1}{k}$.
   - True.
   - False.

5. The Weak Law of Large Numbers states that the sample average of a sequence of independent and identically distributed (i.i.d.) random variables with finite mean μ and variance σ^2 converges in probability to μ.
   - True.
   - False.

Strong law of large numbers

Introduction

- Consider the sequence of random varibles:

$$X_1 = \xi_1, \quad X_2 = \frac{\xi_1 + \xi_2}{2}, \quad \ldots, \quad X_n = \frac{\xi_1 + \cdots + \xi_n}{n}, \quad X_{n+1} = \frac{\xi_1 + \cdots + \xi_{n+1}}{n+1}, \quad \ldots$$

- We will be interested in the limit of the sequence X_n.
- Let $X_\infty = \lim_{n\to\infty} X_n$, then X_∞ is also a random variable.
- The Strong law of large numbers guarantees that

$$X_\infty = \mu \quad \text{almost surely.}$$

- Note that $X_\infty, X_1, X_2, \ldots$ are all functions from Ω to $\mathbb{R}$, then $X_\infty = \mu$ is the convergence of a sequence of functions.

Another way to tell the difference

- From Weak law of large numbers，

$$\mathbb{P}\big(|\bar{X}_n - \mu| < \varepsilon\big) \quad \text{is close to 1 when } n \text{ is large,}$$

- It is still possible that at least one of the events

$$B_{n+1} = \{|\bar{X}_{n+1} - \mu| \geqslant \varepsilon\}, \quad B_{n+2} = \{|\bar{X}_{n+2} - \mu| \geqslant \varepsilon\}, \quad \ldots, \quad B_{n+m} = \{|\bar{X}_{n+m} - \mu| \geqslant \varepsilon\}, \quad \ldots$$

  may happen.

- We need to consider a stronger version of law of large numbers.

Convergence of function sequences

- Let $\Omega = [-1, 1]$ and let $f_n : \Omega \to \mathbb{R}$ be a sequence of functions.
- What is the limit of f_n?
- For example,

$$f_n(x) = \frac{1}{n}[nx + (-1)^n \sin(1 + x^2)],$$

and it can be shown that $\lim_{n\to\infty} f_n(x) = f$, where $f(x) = x$.

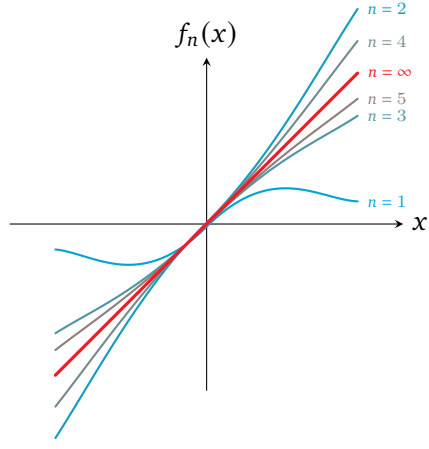


Figure: $\{f_n\}$ and its limit

Definition 13

Let $\{f_n\}$ be a sequence of function on Ω. Then $\{f_n\}$ converges pointwise to $f : \Omega \to \mathbb{R}$ if for each $x \in \Omega$ and each $\varepsilon > 0$ there exists $N > 0$ such that

$$n \geqslant N \implies |f_n(x) - f(x)| < \varepsilon.$$

Remark

This definition can be rewritten as a language of set theory:

$$x \in \Omega \implies x \in \bigcap_{\varepsilon>0} \bigcup_{N\geqslant 1} \bigcap_{n>N} \{x : |f_n(x) - f(x)| \leqslant \varepsilon\},$$

or, equivalently,

$$\Omega = \bigcap_{\varepsilon>0} \bigcup_{N\geqslant 1} \bigcap_{n>N} \{x \in \Omega : |f_n(x) - f(x)| \leqslant \varepsilon\}.$$

Example

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Example 14

Consider the sequence $\{f_n\}$ defined on $[0,1]$ by $f_n(x) = x^n$. What is $\lim_{n\to\infty} f_n$?

Solution.

Since for all $n \geqslant 1$, $f_n(1) = 1^n = 1$, and therefore, $\lim_{n\to\infty} f_n(1) = 1$. On the other hand, if $x \in [0,1)$ then

$$f_n(x) = x^n \to 0 \quad \text{as } n \to \infty.$$

Therefore, the limit is

$$f(x) = \begin{cases} 0 & \text{if } x \in [0,1) \\ 1 & \text{if } x = 1. \end{cases}$$

■

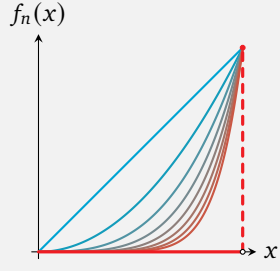


Figure: f_n and its limit

Almost sure convergence

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Definition 15

A sequence of random variables $\{X_n\}$ converges almost surely to X, written as $X_n \xrightarrow{a.s.} X$ $(n \to \infty)$, if

$$\mathbb{P}\left\{\lim_{n\to\infty} X_n = X\right\} = 1.$$

Example 16

Let $\Omega = [0,1]$, $\mathscr{F} = \mathscr{B}(\Omega)$ and $\mathbb{P}$ is the Lebesgue measure. Let $U(\omega) = \omega$. Then, we have $U \sim \text{Uniform}(0,1)$.
Let $X_n = U^n$. By Example 14, we have

$$X(\omega) := \lim_{n\to\infty} X_n(\omega) = \begin{cases} 0 & \text{if } \omega \in [0,1) \\ 1 & \text{if } \omega = 1. \end{cases}$$

In this case,

$$\{\lim_{n\to\infty} X_n = 0\} = \{\omega : 0 \leqslant \omega < 1\} := E,$$

and $\mathbb{P}(E) = 1$. Therefore,

$$X_n \xrightarrow{a.s.} 0 \quad \text{as } n \to \infty.$$

Remark

Note that $E = \{\lim_{n\to\infty} X_n = X\}$ means that, for any $\omega \in E$, and for any $\varepsilon > 0$ ($M > 0$), there exists a number $N > 0$ (depending on ε or M) such that for all $n > N$,

$$|X_n(\omega) - X| \leqslant \varepsilon \quad (\text{or } \frac{1}{M}),$$

which is equivalent to $E = \bigcap_{\varepsilon>0} \bigcup_{N\geqslant 1} \bigcap_{n>N} \{\omega : |X_n(\omega) - X(\omega)| \leqslant \varepsilon\}$. Therefore,

$$\{\lim_{n\to\infty} X_n = X\} = \bigcap_{\varepsilon>0} \bigcup_{N\geqslant 1} \bigcap_{n>N} \{|X_n - X| \leqslant \varepsilon\},$$

and

$$\{\lim_{n\to\infty} X_n = X\}^c = \bigcup_{\varepsilon>0} \bigcap_{N\geqslant 1} \bigcup_{n>N} \{|X_n - X| > \varepsilon\}.$$

Infinitely often

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Definition 17 (Infinitely often)

Let $\{A_n\}_{n\geqslant 1}$ be an infinite sequence of events. We say that the events in the sequence occur infinitely often, written as A_n i.o. if A_n holds for an infinite number of indices $n \in \{1, 2, 3, \dots\}$. Conversely, we say $\{A_n\}$ happens finitely often, written as A_n f.o. if they do not occur infinitely often.

Example 18

- Let $a_n = (-1)^n$, $n = 1, 2, \dots$ and let $A_n = \{a_n \text{ is positive}\}$.
- Let $a_n = \frac{1}{n}$, and let $A_n = \{|a_n - 0| > \varepsilon\}$.

Infinitely often

Solution.

- A_n i.o. because a_n is positive for $n = 2, 4, 6, \ldots$.
- A_n f.o. because $|a_n - 0| > \varepsilon$ only if $n = 1, 2, \ldots, \lfloor \frac{1}{\varepsilon} \rfloor$. ■

Infinitely often

Proposition 19

We have

$$\{A_n \text{ i.o.}\} = \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n = \lim_{m\to\infty} \underbrace{\bigcup_{n=m}^{\infty} A_n}_{:=\sup_{n\geqslant m} A_n} = \limsup_{n\to\infty} A_n,$$

$$\{A_n \text{ f.o.}\} = \bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} A_n^c = \lim_{m\to\infty} \underbrace{\bigcap_{n=m}^{\infty} A_n^c}_{:=\inf_{n\geqslant m} A_n^c} = \liminf_{n\to\infty} A_n.$$

Proposition 20

We have $X_n \xrightarrow{a.s.} X$ if and only if

$$\mathbb{P}\{|X_n - X| > \varepsilon, \text{i.o.}\} = 0 \quad \text{for any } \varepsilon > 0.$$

Proof.

$\Longrightarrow$: If $\mathbb{P}\{|X_n - X| > \varepsilon, \text{i.o.}\} = 0$ for any $\varepsilon > 0$, then, with $A_n(u) = \{|X_n - X| > u\}$ for $u > 0$, we have for any $M \geqslant 1$,

$$\mathbb{P}\left(\bigcap_{m=1}^{\infty} \bigcup_{n=1}^{\infty} A_n\left(\frac{1}{M}\right)\right) = 0,$$

and therefore,

$$\mathbb{P}(\{\lim_{n\to\infty} X_n = X\}^c) = \mathbb{P}\left(\bigcup_{M=1}^{\infty} \bigcap_{m=1}^{\infty} \bigcup_{n=1}^{\infty} A_n\left(\frac{1}{M}\right)\right) \leqslant \sum_{M=1}^{\infty} \mathbb{P}\left(\bigcap_{m=1}^{\infty} \bigcup_{n=1}^{\infty} A_n\left(\frac{1}{M}\right)\right) = 0.$$

$\Longleftarrow$: If $X_n \xrightarrow{a.s.} X$, then, by definition,

$$
\begin{aligned}
0 &= \mathbb{P}\left(\bigcup_{\varepsilon>0}\bigcap_{m=1}^{\infty}\bigcup_{n=1}^{\infty} A_n(\varepsilon)\right) \\
&\geqslant \mathbb{P}\left(\bigcap_{m=1}^{\infty}\bigcup_{n=1}^{\infty} A_n(\varepsilon)\right) \quad \text{for any } \varepsilon > 0 \\
&= \mathbb{P}\{|X_n - X| > \varepsilon, \text{i.o.}\}.
\end{aligned}
$$

■

Borel-Cantelli's lemma

The Borel-Cantelli's lemma gives a sufficient condition for $\mathbb{P}\{A_n, \text{i.o.}\} = 0$.

Proposition 21

Let $\{A_n\}_{n\geqslant 1}$ be an infinite sequence of events. If $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, then

$$\mathbb{P}\{A_n, \text{i.o.}\} = 1.$$

Proof.

Note that

$$\begin{aligned}\mathbb{P}\{A_n, \text{i.o.}\} &= \mathbb{P}\left\{\bigcap_{m=1}^{\infty}\bigcup_{n=m}^{\infty} A_n\right\} \\ &= \lim_{m\to\infty} \mathbb{P}\left(\bigcup_{n=m}^{\infty} A_n\right) = 0 \quad \text{because } \sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty.\end{aligned}$$

■

Example 22

Consider a sequence of independent events $A_1, A_2, \ldots$ where A_n represents the event that a fair coin flip comes up heads n times in a row. Find $\mathbb{P}\{A_n, \text{i.o.}\}$.

Solution.

Since $\mathbb{P}(A_n) = \frac{1}{2^n}$, it follows that

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) = 1 < \infty.$$

Therefore, by the Borel-Cantelli lemma, $\mathbb{P}\{A_n, \text{i.o.}\} = 0$, which means that the probability that A_n occurs infinitely often is 0. ■

Example 23

Let $X_1, X_2, \ldots$ be independent and identically distributed Uniform$(0,1)$ random variables. Set $Y_n = X_n/n$. Show that $Y_n \xrightarrow{a.s.} 0$.

Proof.

Note that

$$\mathbb{P}\{|Y_n - 0| > \varepsilon\} = \mathbb{P}\{X_n^2 > n^2\varepsilon^2\} \leqslant \frac{\mathbb{E}[X_n^2]}{n^2\varepsilon^2} = \frac{1}{3n^2\varepsilon^2}, \quad \text{by Markov's inequality,}$$

and it follows that

$$\sum_{n=1}^{\infty} \mathbb{P}\{|Y_n - 0| > \varepsilon\} \leqslant \frac{1}{3\varepsilon^2} \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty.$$

Therefore, $\mathbb{P}\{|Y_n - 0| > \varepsilon, \text{i.o.}\} = 0$, and thus $Y_n \xrightarrow{a.s.} 0$. ■

Relation between convergence a.s. and in probability

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Proposition 24

If $X_n \xrightarrow{a.s.} X$, then $X_n \xrightarrow{p} X$.

Proof.

If $X_n \xrightarrow{a.s.} a$, then

$$\mathbb{P}\left(\bigcap_{m=1}^{\infty}\bigcup_{n=m}^{\infty}\{|X_n - X| > \varepsilon\}\right) = 0 \quad \text{for any } \varepsilon > 0,$$

which is equivalent to

$$\lim_{m\to\infty} \mathbb{P}\left(\bigcup_{n=m}^{\infty}\{|X_n - X| > \varepsilon\}\right) = 0.$$

Relation between convergence a.s. and in probability

However,

$$\mathbb{P}\{|X_m - X| > \varepsilon\} \leqslant \mathbb{P}\left(\bigcup_{n=m}^{\infty}\{|X_n - X| > \varepsilon\}\right),$$

and therefore,

$$\lim_{m\to\infty} \mathbb{P}\{|X_m - X| > \varepsilon\} \leqslant \lim_{m\to\infty} \mathbb{P}\left(\bigcup_{n=m}^{\infty}\{|X_n - X| > \varepsilon\}\right) = 0,$$

which implies that $X_n \xrightarrow{p} X$. ■

Remark

The inverse is not correct. That is, if $X_n \xrightarrow{p} X$, then it is not necessarily that $X_n \xrightarrow{a.s.} X$.

Example 25 (Conv. in prob. but not a.s.)

Let $X_n \sim \text{Bernoulli}(1/n)$ be independent. Then $X_n \xrightarrow{p} 0$ but it is not true that $X_n \xrightarrow{a.s.} 0$.

Proof.

For any $\varepsilon > 0$,

$$\mathbb{P}\{|X_n - 0| > \varepsilon\} = \mathbb{P}\{X_n = 1\} = \frac{1}{n} \to 0 \quad \text{as } n \to \infty.$$

So $X_n \xrightarrow{p} 0$. On the other hand, note that with $A_n = \{|X_n - 0| > \varepsilon\}$,

$$\mathbb{P}(\{|X_n - 0| > \varepsilon, \text{i.o.}\}^c) = \mathbb{P}\{A_n^c, \text{f.o.}\} = \mathbb{P}\left(\bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} A_n^c\right).$$

Examples

For every $m \geqslant 1$, because A_n's are independent,

$$\mathbb{P}\left(\bigcap_{n=m}^{\infty} A_n^c\right) = \prod_{n=m}^{\infty}\left(1 - \frac{1}{n}\right) = 0,$$

and therefore,

$$\mathbb{P}(\{|X_n - 0| > \varepsilon, \text{i.o.}\}^c) \leqslant \sum_{m=1}^{\infty} \mathbb{P}\left(\bigcap_{n=m}^{\infty} A_n^c\right) = 0,$$

which implies that

$$\mathbb{P}\{|X_n - 0| > \varepsilon, \text{i.o.}\} = 1.$$

This proves that X_n does not converge a.s. to 0. ■

Theorem 26

If $X_n \xrightarrow{p} X$, then there exists a sequence $\{n_k\}$ of integers increasing to infinity such that $X_{n_k} \xrightarrow{a.s.} X$.

Briefly stated: convergence in probability implies convergence almost surely along a subsequence.

Proof.

Because $X_n \xrightarrow{p} X$, then for any $\varepsilon_1 > 0$,

$$\lim_{n\to\infty} \mathbb{P}\Big(|X_n - X| > \varepsilon_1\Big) = 0.$$

Take $\varepsilon_1 = 2^{-k}$, then

$$\lim_{n\to\infty} \mathbb{P}\Big(|X_n - X| > \frac{1}{2^k}\Big) = 0, \quad \text{for any } k \geqslant 1.$$

Then, for any $k \geqslant 1$, and for $\varepsilon_{2,k} = 2^{-k}$, there exists $N_k \geqslant 1$ such that

$$\mathbb{P}\Big(|X_n - X| > \frac{1}{2^k}\Big) \leqslant \delta = \frac{1}{2^k}, \quad \text{for any } n \geqslant N_k.$$

Choose $n_k = N_k$ which was defined as above, and let

$$E_k = \{|X_{n_k} - X| > \frac{1}{2^k}\}.$$

Then, we have

$$\sum_{k=1}^{\infty} \mathbb{P}(E_k) \leqslant \sum_{k=1}^{\infty} \frac{1}{2^k} = 1.$$

For any $\varepsilon > 0$, define

$$A_k = \{|X_{n_k} - X| > \varepsilon\}.$$

Then, for $k > \log_2(1/\varepsilon)$, we have $\varepsilon > 2^{-k}$, and it follows that $A_k \subset E_k$.

Therefore,

$$\sum_{k=1}^{\infty} \mathbb{P}(A_k) = \sum_{k=1}^{[\log_2(1/\varepsilon)]} \mathbb{P}(A_k) + \sum_{k=[\log_2(1/\varepsilon)]+1} \mathbb{P}(E_k) \leqslant [\log_2(1/\varepsilon)] + 1 < \infty.$$

By the Borel–Cantelli lemma, we have

$$\mathbb{P}(A_k \; i.o.) = 0,$$

which implies that $X_{n_k} \xrightarrow{X}$ almost surely as $k \to \infty$.

Convergence in L^p and convergence a.s.

- Convergence in L^p does not imply convergence a.s.
- Convergence a.s. does not imply convergence in L^p, either.
- Convergence in probability does not imply convergence in L^p.

Convergence in L^p does not imply convergence a.s.

Consider the probability space $([0,1], \mathscr{B}[0,1], \mathbb{P})$, where $\mathbb{P} = \lambda$ is the Lebesgue measure. Define

$$
\begin{aligned}
&X_1 = \mathbf{1}_{[0,1]},\\
&X_2 = \mathbf{1}_{[0,1/2]}, \quad X_3 = \mathbf{1}_{[1/2,1]},\\
&X_4 = \mathbf{1}_{[0,1/4]}, \quad X_5 = \mathbf{1}_{[1/4,1/2]}, \quad X_6 = \mathbf{1}_{[1/2,3/4]}, \quad X_7 = \mathbf{1}_{[3/4,1]},\\
&X_8 = \mathbf{1}_{[0,1/8]}, \quad \dots
\end{aligned}
$$

That is,

$$X_n = \mathbf{1}_{[(j-1)/2^k, j/2^k]},$$

where $k = [\log_2 n]$ and $j = n - 2^k + 1$.
For example, if $n = 9$, then $k = [\log_2 9] = 3$, and $j = n - 2^k + 1 = 2$, then

$$X_9 = \mathbf{1}_{[\frac{1}{2^3}, \frac{2}{2^3}]}.$$

Therefore,

$$\mathbb{E}|X_n|^p = \mathbb{P}(\{\omega : \frac{j-1}{2^k} \leqslant \omega \leqslant \frac{j}{2^k}\}) = \frac{1}{2^k} = \frac{1}{2^{[\log_2 n]}} \leqslant \frac{2}{n} \to 0 \quad \text{as } n \to \infty.$$

Hence, by definition,

$$X_n \xrightarrow{L^p} 0.$$

However, for any $\omega \in [0,1]$, and for any $N \geqslant 1$, there exists $n \geqslant N$ such that

$$X_n(\omega) = 1.$$

Then there exists infinitely many n such that $X_n(\omega) = 1$. Therefore,

$$X_n \not\to 0 \quad \text{almost surely.}$$

Convergence in probability does not implies convergence in L^p

Define

$$Y_n = n^{1/p} X_n.$$

Then, for any $\varepsilon \in (0, 1)$,

$$\mathbb{P}(Y_n > \varepsilon) \leqslant \mathbb{P}(Y_n > 0) = \mathbb{P}(X_n = 1) \leqslant \frac{2}{n} \to 0 \quad \text{as } n \to \infty,$$

which implies that $Y_n \xrightarrow{p} 0$.
However,

$$\mathbb{E}[Y_n^p] = n\,\mathbb{P}(X_n = 1) = \frac{n}{2^{[\log_2 n]}} \geqslant 1.$$

Therefore,

$$Y_n \not\to 0 \quad \text{in } L^p.$$

Convergence a.s. does not imply convergence in L^p

Define

$$U_n = n^{1/p} \mathbf{1}_{[0,1/n]}.$$

Then, as $n \to \infty$, $U_n \to 0$ a.s.
However,

$$\mathbb{E}\,|X_n|^p = n \times \frac{1}{n} = 1.$$

Therefore,

$$X_n \not\to 0 \quad \text{in } L^p.$$

Strong law of large numbers

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Theorem 27

Let $\xi, \xi_1, \ldots, \xi_n$ be a sequence of i.i.d. random variables with common mean μ and variance σ^2. Assume that $\mathbb{E}[\xi^4] < \infty$. Let $X_n = \bar{\xi}_n$ be the sample mean. Then,

$$\mathbb{P}\{\lim_{n\to\infty} X_n = \mu\} = 1.$$

Remark

- The assumption can be relaxed to $\mathbb{E}|\xi| < \infty$.
- The strong law flips the order of limit and probability.
- Strong law $\neq$ deterministic.
- The strong law asserts that there are a finite number of failures ($|X_n - \mu| > \varepsilon$).

Multiple-choice Questions

1. (Multiple Choice) The sequence of random variables $X_1, X_2, X_3, \ldots$ converges almost surely to a constant c if:

   A. For all $\varepsilon > 0$, $\mathbb{P}(|X_n - c| > \varepsilon)$ approaches 0 as n approaches infinity.

   B. For all $\varepsilon > 0$, $\mathbb{P}(|X_n - c| > \varepsilon, \text{i.o.}) = 0$.

   C. For all $\varepsilon > 0$, $\mathbb{P}(|X_n - c| \leqslant \varepsilon, \text{i.o.}) = 1$.

   D. For all $\varepsilon > 0$, $\mathbb{E}(|X_n - c| > \varepsilon)$ approaches 0 as n approaches infinity.

Multiple-choice Questions

2. (True/False) The Strong Law of Large Numbers states that the sample average of a sequence of independent and identically distributed (i.i.d.) random variables with finite mean μ converges almost surely to μ.

   A. True

   B. False

Multiple-choice Questions

3. (Multiple Choice) Which of the following best describes the difference between the Weak Law of Large Numbers and the Strong Law of Large Numbers?

   A. The Weak Law of Large Numbers applies to sequences of i.i.d. random variables, while the Strong Law of Large Numbers applies to sequences of dependent random variables.

   B. The Weak Law of Large Numbers pertains to convergence in probability, while the Strong Law of Large Numbers pertains to almost sure convergence.

   C. The Weak Law of Large Numbers pertains to almost sure convergence, while the Strong Law of Large Numbers pertains to convergence in probability.

   D. The Weak Law of Large Numbers applies to sequences of random variables with finite mean, while the Strong Law of Large Numbers applies to sequences of random variables with finite variance.

Multiple-choice Questions

4. (True/False) Almost sure convergence implies convergence in probability, but convergence in probability does not imply almost sure convergence.

   A. True

   B. False

Multiple-choice Questions

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

5. (Multiple Choice) Given a sequence of i.i.d. random variables $X_1, X_2, X_3, \ldots$ with finite mean μ, under which condition does the Strong Law of Large Numbers hold?

   A. The random variables have finite variance.

   B. The random variables have finite kurtosis.

   C. The random variables have a finite second moment.

   D. None of the above is necessary.

Convergence in distribution

Definition of Convergence in Distribution

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Definition 28

- A sequence of random variables $X_1, X_2, ..., X_n$ converges in distribution (or in law) to a random variable X if the cumulative distribution function (CDF) of X_n converges to the CDF of X at all points where $F(x)$ is continuous.
- Formally, X_n converges to X in distribution if $\lim_{n\to\infty} F_n(x) = F(x)$ for all x at which $F(x)$ is continuous.

Convergence in Distribution vs. Other Convergence

- Convergence in probability and almost sure convergence are stronger than convergence in distribution. That is, if a sequence of random variables converges in probability or almost surely, it will also converge in distribution.
- However, the converse is not true: convergence in distribution does not imply convergence in probability or almost surely.

Central Limit Theorem

- The Central Limit Theorem (CLT) is a fundamental theorem in probability theory and statistics which states that the sum of a large number of independent and identically distributed (i.i.d.) random variables, each with finite mean and variance, will have a distribution that is approximately normal.
- The CLT is an example of convergence in distribution: as the number of random variables increases, the distribution of the sum (properly normalized) converges to a normal distribution.

Theorem 29

If $X_n \xrightarrow{p} X$, then $X_n \xrightarrow{d} X$.

Proof.

For any $a < b$,

$$\begin{aligned} F(a) = \mathbb{P}\{X \leqslant a\} &= \mathbb{P}(\{X \leqslant a\} \cap \{X_n \leqslant b\}) + \mathbb{P}(\{X \leqslant a\} \cap \{X_n > b\}) \\ &\leqslant \mathbb{P}\{X_n \leqslant b\} + \mathbb{P}(\{X \leqslant a\} \cap \{X_n > b\}), \end{aligned}$$

and the second term can be bounded by

$$\mathbb{P}(\{X \leqslant a\} \cap \{X_n > b\}) \leqslant \mathbb{P}\{|X_n - X| > b - a\} \to 0.$$

Therefore, $F(a) \leqslant \liminf_{n\to\infty} \mathbb{P}\{X_n \leqslant b\}$. Similarly, $\limsup_{n\to\infty} \mathbb{P}\{X_n \leqslant b\} \leqslant F(a')$ if $a' > b$. If b is a continuity point of F, then as $a \uparrow b$ and $a' \downarrow b$, $F(b) = \lim_{n\to\infty} \mathbb{P}\{X_n \leqslant b\}$. ■

Theorem 30

For any constant a,

$$X_n \xrightarrow{p} a \quad \Longleftrightarrow \quad X_n \xrightarrow{d} a.$$

Proof.

We only prove the $\Longleftarrow$ part. For any $\varepsilon > 0$,

$$\begin{aligned}
\mathbb{P}\{|X_n - a| \geqslant \varepsilon\} &= \mathbb{P}\{X_n \geqslant a + \varepsilon\} + \mathbb{P}\{X_n \leqslant a - \varepsilon\} \\
&\leqslant \mathbb{P}\{X_n > a + \frac{\varepsilon}{2}\} + \mathbb{P}\{X_n \leqslant a - \varepsilon\} \\
&= 1 - F_n(a + \frac{\varepsilon}{2}) + F_n(a - \varepsilon),
\end{aligned}$$

where

$$F_n(a + \frac{\varepsilon}{2}) \to F(a + \frac{\varepsilon}{2}) = 1, \quad F_n(a - \varepsilon) \to F(a - \varepsilon) = 0,$$

and thus $\mathbb{P}\{|X_n - a| \geqslant \varepsilon\} \to 0$.

Classical Central Limit Theorems

Introduction

- The law of large numbers:

$$\bar{\xi}_n - \mu = \frac{1}{n}\sum_{i=1}^{n}(\xi_i - \mu) \to 0.$$

- How close is the convergence?
- We know that

$$\mathrm{Var}(\bar{\xi}_n - \mu) = \frac{\sigma^2}{n},$$

where $\sigma^2 = \mathrm{Var}(\xi)$.

- Consider the normalized variable:

$$Z_n = \frac{\sqrt{n}(\bar{\xi}_n - \mu)}{\sigma} \to ?$$

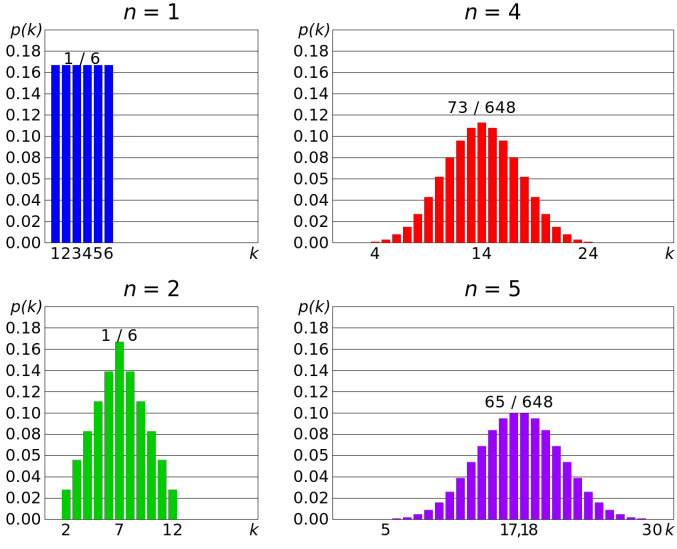


Figure: Pictorial illustration of the Central Limit Theorem. Suppose we throw a die and record the face $\sum_{i=1}^{n} \xi_i$.

CLT: i.i.d. random variables

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Theorem 31

Let $\xi, \xi_1, \xi_2, \ldots$ be a sequence of i.i.d. random variables with $\mu = \mathbb{E}[\xi]$ and $\sigma^2 = \mathrm{Var}(\xi)$. Let

$$Z_n = \frac{\sqrt{n}(\bar{\xi}_n - \mu)}{\sigma}.$$

The cdf of Z_n is converging pointwise to the cdf of $N(0,1)$, in other words, for every $x \in \mathbb{R}$,

$$\mathbb{P}\{Z_n \leqslant x\} \to \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt, \quad \text{as } n \to \infty.$$

Remark

If $\xi_i \sim N(\mu, \sigma^2)$, then Z_n is exactly a $N(0,1)$ random variable.

Normal approximation

Let $S_n = \xi_1 + \cdots + \xi_n$. If n is large, then the probability $\mathbb{P}\{S_n \leqslant s\}$ can be approximated by treating S_n as if it were normal:

(i) Calculate the mean $n\mu$ and the variance $n\sigma^2$ of S_n.

(ii) Calculate the normalized value

$$z = \frac{s - n\mu}{\sqrt{n}\sigma}.$$

(iii) Use the approximation:

$$\mathbb{P}\{S_n \leqslant s\} \approx \Phi(z).$$

Example 32

We load on a plane 100 packages whose weights are independent random variables that are **uniformly distributed** between 5 and 50 pounds. What is the probability that the total weight will exceed 3000 pounds? It is not easy to calculate the cdf of the total weight and the desired probability, but an approximate answer can be quickly obtained using the central limit theorem.

Solution.

Let ξ_i be the weight of the ith package for every $i \geqslant 1$, then it follows that $\xi_i \sim \text{Uniform}(5, 50)$. Let $S_n = \xi_1 + \cdots + \xi_n$ be the sum of the weights of n packages. We want to calculate $\mathbb{P}\{S_{100} > 3000\}$.

Solution (Cont'd).

Now, since

$$\mu = \frac{5+50}{2} = 27.5, \quad \sigma^2 = \frac{(50-5)^2}{12} = 168.75,$$

we have

$$z = \frac{3000 - (100)(27.5)}{\sqrt{(100)(168.75)}} = 1.92.$$

Therefore, by the CLT,

$$\mathbb{P}\{S_{100} > 3000\} = 1 - \mathbb{P}\{S_{100} \leqslant 3000\} \approx 1 - \Phi(1.92) \approx 0.274.$$

■

Example 33

We poll n voters and record the fraction M_n of those polled who are in favor of a particular candidate. If p is the fraction of the entire voter population that supports this candidate, then

$$M_n = \frac{1}{n}(\xi_1 + \cdots + \xi_n),$$

where $\{\xi_i\}$ are independent Bernoulli random variables with parameter p. Then, $\mathbb{E}[\xi_i] = p$ and $\mathrm{Var}(\xi_i) = p(1-p)$, and it follows that

$$Z_n = \frac{\sqrt{n}(M_n - \mu)}{\sigma} = \frac{\sqrt{n}(M_n - p)}{\sqrt{p(1-p)}}.$$

Therefore, by the CLT,

$$\mathbb{P}\{M_n - p \geqslant \varepsilon\} \leqslant 1 - \Phi(2\sqrt{n}\varepsilon).$$

de Moivre
棣莫弗
1733

Laplace
拉普拉斯
1812

Lyapunov
李亚普诺夫
1901

Pólya
波利亚
1920

Time

- 1733: de Moivre used the normal distribution to approximate the distribution of the number of heads resulting from many tosses of a fair coin.
- 1812: Laplace approximated the binomial distribution with the normal distribution.
- 1901: Lyapunov gave a rigorous proof of the central limit theorem.
- 1920: Pólya referred to the theorem as ”central” due to its importance in probability theory.

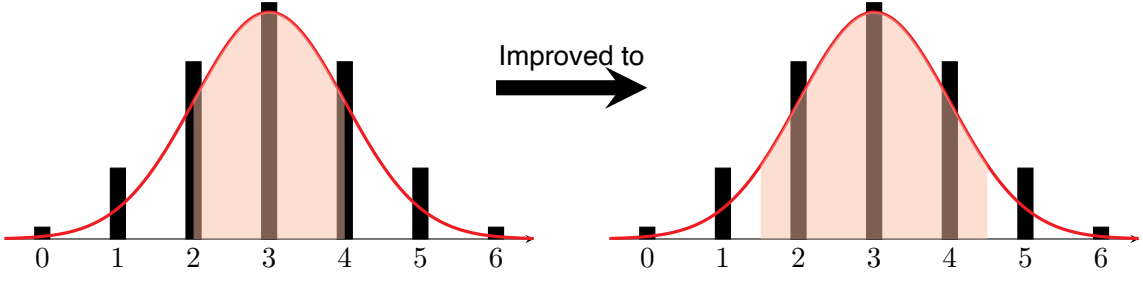
de Moivre-Laplace approximation to the binomial

- If $S_n \sim \text{Binomial}(n, p)$, then it can be viewed as

$$S_n = \xi_1 + \cdots + \xi_n, \quad \xi_i \sim \text{Bernoulli}(p).$$

- Suppose that we want to calculate

$$\mathbb{P}\{k \leqslant S_n \leqslant \ell\} \leqslant \mathbb{P}\{k - \frac{1}{2} \leqslant S_n \leqslant \ell + \frac{1}{2}\} \leqslant \mathbb{P}\{k - 1 < S_n < \ell + 1\}$$



Theorem 34 (De Moivre-Laplace approximation)

If $S_n \sim \text{Binomial}(n, p)$, where n is large and $k \leqslant \ell$ are nonnegative integers, then

$$\mathbb{P}\{k \leqslant S_n \leqslant \ell\} \approx \Phi\left(\frac{\ell + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k - \frac{1}{2} - np}{\sqrt{np(1-p)}}\right).$$

Example 35

Let $S_n \sim \text{Binomial}(36, 0.5)$. Find $\mathbb{P}\{S_n \leqslant 21\}$.

Solution.

An exact calculation yields: $\mathbb{P}\{S_n \leqslant 21\} = \sum_{k=0}^{21} \binom{36}{k} (0.5)^{36} = 0.8785075$. The central limit theorem, without the above refinement, yields:

$$\mathbb{P}\{S_n \leqslant 21\} \approx \Phi\left(\frac{21 - (36)(0.5)}{\sqrt{(36)(0.5)(1-0.5)}}\right) = \Phi(1) \approx 0.8413.$$

However, using the refinement,

$$\mathbb{P}\{S_n \leqslant 21\} \approx \Phi\left(\frac{21.5 - (36)(0.5)}{\sqrt{(36)(0.5)(1-0.5)}}\right) = \Phi(1.17) \approx 0.8789995.$$

■

Multiple-choice Questions

1. (Multiple Choice) Consider a sequence of independent and identically distributed (i.i.d.) random variables $X_1, X_2, ..., X_n$ with mean μ and variance σ^2. The Central Limit Theorem (CLT) best applies to:

   A. $\dfrac{X_1 + X_2 + ... + X_n}{n}$

   B. $\sqrt{n}\left(\dfrac{X_1 + X_2 + ... + X_n}{n} - \mu\right)$

   C. $\sqrt{n}(X_1 - \mu)$

   D. None of the above

Multiple-choice Questions

2. (True/False) Convergence in distribution implies that the sequence of random variables converges to the limit in the sense of having the same expected values and variances.

   A. True

   B. False

Multiple-choice Questions

3. (Multiple Choice) Which of the following is true about the Central Limit Theorem (CLT)?

   A. It states that the sum of a large number of independent and identically distributed random variables, each with finite mean μ and variance σ^2, will be approximately normally distributed $N(n\mu, n\sigma^2)$.

   B. It applies only to normally distributed random variables.

   C. It states that the mean of a large number of independent and identically distributed random variables, each with finite mean μ and variance σ^2, will be approximately normally distributed $N(\mu, \sigma^2)$.

   D. It requires the random variables to be independent but not identically distributed.

Multiple-choice Questions

4. (True/False) The Central Limit Theorem ensures that, with a sufficiently large sample size, the sampling distribution of the sample mean is approximately normally distributed, regardless of the shape of the population distribution.

   A. True

   B. False

Multiple-choice Questions

5. (Multiple Choice) Convergence in distribution of a sequence of random variables $\{X_n\}$ to a random variable X is defined by:

   A. $P(X_n \leqslant x) \longrightarrow P(X \leqslant x)$ as $n \longrightarrow \infty$ for every number x at which $F(x)$ is continuous.

   B. $E(X_n) \longrightarrow E(X)$ and $\mathrm{Var}(X_n) \longrightarrow \mathrm{Var}(X)$ as $n \longrightarrow \infty$.

   C. $P(|X_n - X| > \varepsilon) \longrightarrow 0$ as $n \longrightarrow \infty$.

   D. $P(|X_n - X| > \varepsilon, \text{i.o.}) = 0$.

Further reading

[1] Sheldon M. Ross (谢尔登・M. 罗斯).

A first course in probability (概率论基础教程): Chapter 8.

10th edition (原书第十版), 机械工业出版社

[2] Stanley H. Chan.

Introduction to Probability for Data Science: Chapter 6.

Michigan Publishing