Probability and Stochastic Processes: Stochastic Convergence

Zhenyu Liao

School of Electronic Information and Communications, HUST

May 16, 2026

Outline

1. Introduction and Motivation
2. Convergence of Sequences of Numbers and Functions
3. Stochastic Convergence
4. Discussions and Implications
5. Laws of Large Numbers
6. Stochastic Small-o and Big-O Notations
7. Delta Method

Today

WHAT?

- **Stochastic Convergence**: convergence of sequences of random variables, a.s. convergence, in r-th mean, in probability, in distribution, etc.

WHY?

- Statistics: model **stochasticity** that cannot be precisely described (in a deterministic fashion)
- Stochastic convergence: allows to understand and **predict** the consequence of the stochasticity

Motivation: convergence of sequences of random variables

Example

Let $\{Y_i\}$ denote a sequence of i.i.d. random variable (RVs) uniformly distributed over the integers $\{0, 1, \ldots, 9\}$, and consider

$$X_n = \sum_{i=1}^{n} Y_i 10^{-i}.$$

Expect that the X_n converges, for $n \to \infty$, to a uniform RV X on $[0, 1]$.This is indeed the case (in some sense) and we write $X_n \to X$.

Example

Let $\{X_i\}$ denote a sequence of i.i.d. RVs with mean μ, and consider the sample mean

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

Expect that as $n \to \infty$, the sample mean converges to the true mean. This is indeed the case (in some sense) and we write $\overline{X}_n \to \mu$.

The "meaning" of stochastic convergence may be quite different according to the cases.

Convergence of sequences of numbers

- The infimum of a set of numbers $A = \{a_1, a_2, \ldots\}$ is the larger number $\underline{a}$ such that $\underline{a} \leq a_i$ for all i. We write $\underline{a} = \inf A$.
- The supremum of a set of numbers $A = \{a_1, a_2, \ldots\}$ is the smallest number $\bar{a}$ such that $\bar{a} \geq a_i$ for all i. We write $\bar{a} = \sup A$.
- Given a sequence of numbers $\{a_n\}$ we define liminf and limsup as

$$\liminf_{n\to\infty} a_n = \lim_{n\to\infty} \inf\{a_n, a_{n+1}, \ldots\}, \quad \limsup_{n\to\infty} a_n = \lim_{n\to\infty} \sup\{a_n, a_{n+1}, \ldots\}$$

- Obviously, for any sequence we have $\liminf a_n \leq \limsup a_n$.
- We say that the sequence $\{a_n\}$ has a limit (i.e., the limit that $\lim_{n\to\infty} a_n$ exists) if $\liminf a_n = \limsup a_n$.

Convergence of (deterministic) functions (1)

- Consider a sequence of functions $f_n : [a,b] \to \mathbb{R}$, for $n = 1,2,3,\ldots$
- Pointwise convergence: if for all $x \in [a,b]$ the sequence of numbers $f_1(x), f_2(x), f_3(x), \ldots$ converges to a number $f(x)$ (we use the short-hand notation $f_n(x) \to f(x)$ as $n \to \infty$ for all $x \in [a,b]$), then we say that $f_n \to f$ *pointwise*.
- Convergence pointwise and uniformly: for all $\epsilon > 0$ there exists $N(\epsilon)$ such that for all $n \geq N(\epsilon)$

$$|f_n(x) - f(x)| \leq \epsilon, \quad \forall \ x \in [a,b]$$

  Notice: the function $(N(\epsilon), \epsilon)$ provides a uniform bound to the convergence absolute error $|f_n(x) - f(x)|$. The bound is called uniform since it is independent of x.

Convergence of (deterministic) functions (2)

- Norm convergence: consider a set of functions V that forms a normed vector space. Let $\|\cdot\| : V \to \mathbb{R}_+$ denote the norm function satisfying the usual norm axioms:
  1. $\|f\| \geq 0$ for all $f \in V$, with equality iff $f = 0$.
  2. $\|af\| = |a| \cdot \|f\|$ for all $a \in \mathbb{R}$.
  3. $\|f + g\| \leq \|f\| + \|g\|$ (triangle inequality).

  Consider a sequence of functions $f_1, f_2, f_3, \ldots$ in V. We say that $f_n \to f$ in norm if

$$\|f_n - f\| \to 0, \quad \text{as} \quad n \to \infty.$$

Convergence of (deterministic) functions (3)

- Convergence in measure: fix $\epsilon > 0$ and, given two functions h, g defined on $[a, b]$, define the set
$$\mathcal{S}(h, g, \epsilon) = \{x \in [a, b] : |h(x) - g(x)| > \epsilon\}.$$
We say that $f_n \to f$ in measure if, for all $\epsilon > 0$,
$$\int_{\mathcal{S}(f_n, f, \epsilon)} dx = \int 1_{\mathcal{S}(f_n, f, \epsilon)} dx \to 0 \quad \text{as} \quad n \to \infty.$$
- Implications: if $f_n \to f$ pointwise, then $f_n \to f$ in measure, but the converse is not generally true;
- In general, convergence in norm and convergence pointwise do not imply each other.

Modes of stochastic convergence

Definition (Modes of stochastic convergence)

Let $\{X_n\} = \{X_1, X_2, X_3, \ldots\}$ denotes a sequence of RVs defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We say that:

a) $X_n \to X$ almost surely, (written $X_n \xrightarrow{a.s.} X$) if

$$\mathbb{P}\left(\left\{\omega \in \Omega : \boxed{X_n(\omega) \to X(\omega)}\right\}\right) = 1$$

b) $X_n \to X$ in the r-th mean, with $r \geq 1$, (written $X_n \xrightarrow{r} X$) if $\mathbb{E}[|X_n|^r] < \infty$ for all n and

$$\mathbb{E}\left[|X_n - X|^r\right] \to 0, \quad \text{as} \quad n \to \infty$$

c) $X_n \to X$ in probability, (written $X_n \xrightarrow{P} X$) if

$$\mathbb{P}\left(|X_n - X| > \epsilon\right) \to 0, \quad \text{as} \quad n \to \infty, \quad \forall \ \epsilon > 0$$

d) $X_n \to X$ in distribution, (written $X_n \xrightarrow{D} X$) if

$$F_{X_n}(x) \to F_X(x) \quad \forall \ x \in \mathbb{R}$$

(Notice: convergence of cdfs is in the sense for all points of continuity of F_X)

Example: LLN and CLT

- (Strong) law of large numbers (LLN): for a sequence of i.i.d. random variables $X_1, \ldots, X_n$ with the same expectation $\mathbb{E}[X_i] = \mu < \infty$, we have

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to \mu, \tag{1}$$

almost surely as $n \to \infty$. convergence in probability, know as the weak law of LLN

- Central limit theorem (CLT, Lindeberg–Lévy tyep): for a sequence of i.i.d. random variables $X_1, \ldots, X_n$ with the same expectation $\mathbb{E}[x_i] = \mu$ and variance $\operatorname{Var}[x_i] = \sigma^2 < \infty$, we have

$$\sqrt{p}\left(\frac{1}{p}\sum_{i=1}^{p}(x_i - \mu)\right) \to \mathcal{N}(0, \sigma^2), \tag{2}$$

in distribution as $p \to \infty$.

Remarks

- Convergence a.s., also indicated by almost everywhere (a.e.) or with probability 1 (w.p. 1), is akin pointwise convergence of deterministic functions. However, we want to avoid those points $\omega \in \Omega$ belonging to null sets. Hence, instead of requiring that $X_n(\omega) \to X(\omega)$ for all $\omega \in \Omega$, we require the milder condition that the probability ("volume") of the set of ωs for which $X_n(\omega) \to X(\omega)$ has p. 1.
- The most common cases of convergence in the r-th mean are $r = 1$ and $r = 2$. $X_n \xrightarrow{1} X$ is referred to as convergence in mean. $X_n \xrightarrow{2} X$ is referred to as convergence in mean-square.
- Noticing that $\mathbb{P}(|X_n - X| > \epsilon) = \int_{\mathcal{S}(X_n, X, \epsilon)} d\mathbb{P}$, where

$$\mathcal{S}(X_n, X, \epsilon) = \{\omega \in \Omega : |X_n(\omega) - X(\omega)| > \epsilon\}$$

  we recognize that convergence in probability is akin the convergence in measure for deterministic functions.
- Convergence in distribution is also known as "weak convergence", or "convergence in law."

Cauchy convergence

- A sequence of real numbers $\{a_n\}$ is Cauchy convergent if $|a_n - a_m| \to 0$ for $n, m \to \infty$.
- A sequence of real numbers is convergent if and only if it is Cauchy convergent.
- Cauchy convergence has the advantage that we can check convergence even when we do NOT know the limit, just by looking at the difference of terms $|a_n - a_m|$ for large and arbitrary n, m.
- A sequence of RVs $\{X_n\}$ is called a.s. Cauchy convergent if

$$\mathbb{P}\left(\left\{\omega \in \Omega : \boxed{|X_n(\omega) - X_m(\omega)| \to 0}\right\}\right) = 1$$

  and it follows that $\{X_n\}$ is a.s. convergent if and only if it is a.s. Cauchy convergent.

Relations between convergence modes

Example

- Let $X_n = X$ for all n, where X is Bernoulli taking values in $\{0, 1\}$ with equal probability. Clearly, since each X_n has the same cdf (independent of n), we have that $X_n \xrightarrow{D} X$.
- Now, consider $Y = 1 - X$ for all n. Since X and $1 - X$ are identically distributed (NOT independent!) we have that $X_n \xrightarrow{D} Y$ as well.
- However, X_n does not converge to Y in any other way, since $|X_n - Y| = |X_n - 1 + X_n| = 1$ for all n.

Notice: the above example shows that convergence modes do not imply each other in general, with the exception of the generally valid implications summarized by the following theorem.

General implications

Theorem

Let $X_1, X_2, X_3, \ldots$ denote a sequence of RVs defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The following implications hold in general:

$$
\begin{aligned}
&1) \quad (X_n \xrightarrow{a.s.} X) \Rightarrow (X_n \xrightarrow{P} X)\\
&2) \quad (X_n \xrightarrow{r} X) \Rightarrow (X_n \xrightarrow{P} X)\\
&3) \quad (X_n \xrightarrow{P} X) \Rightarrow (X_n \xrightarrow{D} X)
\end{aligned}
$$

and, for $1 \leq s \leq r$,

$$
4) \quad (X_n \xrightarrow{r} X) \Rightarrow (X_n \xrightarrow{s} X)
$$

Notice: no other implications hold in general, but other implications may hold under extra conditions, as we will see later on.

More implications

Theorem

Let $X_1, X_2, X_3, \ldots$ denote a sequence of RVs defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then,

1. $X_n \xrightarrow{D} c$ *for some constant* c*, if and only if* $X_n \xrightarrow{P} c$.
2. $X_n \xrightarrow{D} X$ *and* $X_n - Y_n \xrightarrow{P} 0$*, then* $Y_n \xrightarrow{D} X$.
3. *If* $X_n \xrightarrow{P} X$ *and* $\mathbb{P}(|X_n| \leq C) = 1$ *for all* n *and some constant* C *independent of* n *(uniformly bounded w.p. 1) then* $X_n \xrightarrow{r} X$ *for all* $r \geq 1$.
4. *If* $p_n(\epsilon) = \mathbb{P}(|X_n - X| > \epsilon)$ *satisfies* $\sum_n p_n(\epsilon) < \infty$ *for all* $\epsilon > 0$*, then* $X_n \xrightarrow{a.s.} X$. *(Known as the **Borel–Cantelli Lemma**, commonly used in the proof of a.s. convergence).*

General implication (1)

Lemma

If $X_n \xrightarrow{P} X$, then $X_n \xrightarrow{D} X$. The converse generally fails.

Proof. Suppose $X_n \xrightarrow{P} X$ and write

$$F_n(x) = \mathbb{P}(X_n \leq x), \quad \text{and} \quad F(x) = \mathbb{P}(X \leq x)$$

For $\epsilon \geq 0$, we can write

$$\begin{aligned} F_n(x) &= \mathbb{P}(X_n \leq x, X \leq x+\epsilon) + \mathbb{P}(X_n \leq x, X > x+\epsilon) \\ &\leq F(x+\epsilon) + \mathbb{P}(|X_n - X| > \epsilon), \\ F(x-\epsilon) &= \mathbb{P}(X \leq x-\epsilon, X_n \leq x) + \mathbb{P}(X \leq x-\epsilon, X_n > x) \\ &\leq F_n(x) + \mathbb{P}(|X_n - X| > \epsilon). \end{aligned}$$

Thus we have

$$F(x-\epsilon) - \mathbb{P}(|X_n - X| > \epsilon) \leq F_n(x) \leq F(x+\epsilon) + \mathbb{P}(|X_n - X| > \epsilon)$$

which implies, for $n \to \infty$,

$$F(x-\epsilon) \leq \liminf_{n\to\infty} F_n(x) \leq \limsup_{n\to\infty} F_n(x) \leq F(x+\epsilon)$$

Since ϵ is arbitrary, this implies convergence (limit exists) of $F_n(x)$ to $F(x)$ for any point of continuity x of $F(x)$.

General implications (3) and (4)

Lemma

If $X_n \xrightarrow{1} X$, *then* $X_n \xrightarrow{P} X$. *Furthermore, if* $X_n \xrightarrow{r} X$ *then* $X_n \xrightarrow{s} X$ *for* $1 \leq s < r$.

Proof: Using Markov inequality we have, for all $\epsilon > 0$,

$$\mathbb{P}(|X_n - X| > \epsilon) \leq \frac{\mathbb{E}[|X_n - X|]}{\epsilon}$$

Using Lyapunov inequality, we have that for $1 \leq s \leq r$,

$$\mathbb{E}[|X_n - X|^s]^{1/s} \leq \mathbb{E}[|X_n - X|^r]^{1/r}.$$

General Implication (2)

Lemma

Define the set $A_n(\epsilon) = \{|X_n - X| > \epsilon\}$ *and* $B_m(\epsilon) = \bigcup_{n \geq m} A_n(\epsilon)$. *Then,*

1. $X_n \xrightarrow{a.s.} X$ *if and only if, for all* $\epsilon > 0$, $\mathbb{P}(B_m(\epsilon)) \to 0$ *as* $m \to \infty$.
2. $X_n \xrightarrow{a.s.} X$ *if* $\sum_n \mathbb{P}(A_n(\epsilon)) < \infty$ *for all* $\epsilon > 0$.
3. *If* $X_n \xrightarrow{a.s.} X$, *then* $X_n \xrightarrow{P} X$, *but the converse generally fails.*

Borel Cantelli Lemmas: as proof ingredient

- Consider a sequence of events $A_1, A_2, A_3, \ldots$ in a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$.
- We define the event $\{A_n \text{ i.o.}\}$ (read: event that infinitely many of the A_n's occur, or, A_n occurs infinitely often) as

$$\{A_n \text{ i.o.}\} = \limsup_{n\to\infty} A_n = \bigcap_n \bigcup_{m \geq n} A_m$$

Theorem

1. *If* $\sum_n \mathbb{P}(A_n) < \infty$*, then* $\mathbb{P}(A_n \textit{ i.o.}) = 0$.
2. *If* $\sum_n \mathbb{P}(A_n) = \infty$ *and the* A_n*'s are independent, then* $\mathbb{P}(A_n \textit{ i.o.}) = 1$.

Proof.

1) Let $C = \{\omega : X_n(\omega) \to X(\omega)\}$ and define

$$A(\epsilon) = \{\omega : |X_n(\omega) - X(\omega)| > \epsilon \text{ infinitely often}\} = \bigcap_m \bigcup_{n \geq m} A_n(\epsilon) = \bigcap_m B_m(\epsilon)$$

Now, $X_n(\omega) \to X(\omega)$ if and only if $\omega \notin A(\epsilon)$ for all $\epsilon > 0$. Hence, a.s. convergence (i.e., $\mathbb{P}(C) = 1$) implies $\mathbb{P}(A(\epsilon)) = 0$. Using the continuity of the probability measure, we have

$$\lim_{m\to\infty} \mathbb{P}(B_m(\epsilon)) = \mathbb{P}(\lim_{m\to\infty} B_m(\epsilon)) = \mathbb{P}\left(\bigcap_m B_m(\epsilon)\right) = \mathbb{P}(A(\epsilon)) = 0$$

2) From the definition of $B_m(\epsilon)$ and the union bound we have

$$\mathbb{P}(B_m(\epsilon)) \leq \sum_{n=m}^{\infty} \mathbb{P}(A_n(\epsilon))$$

so $\mathbb{P}(B_m(\epsilon)) \to 0$ if $\sum_{n=1}^{\infty} \mathbb{P}(A_n(\epsilon)) < \infty$.
3) Since $A_m(\epsilon) \subseteq B_m(\epsilon)$ then statement 1) implies that

$$\mathbb{P}(|X_m - X| > \epsilon) = \mathbb{P}(A_m(\epsilon)) \leq \mathbb{P}(B_m(\epsilon)) \to 0$$

which yields convergence in probability. □

A.s. convergence of sub-sequences

Theorem

If $X_n \xrightarrow{P} X$, then there exists a non-random increasing sequence of integers $n_1, n_2, \ldots$, such that the sub-sequence $\{X_{n_i} : i = 1, 2, 3, \ldots\}$ converges to X almost surely, i.e., $X_{n_i} \xrightarrow{a.s.} X$ as $i \to \infty$.

Proof.

Since $X_n \xrightarrow{P} X$, then $\mathbb{P}(|X_n - X| > \epsilon) \to 0$ for all $\epsilon > 0$. Then, pick the sequence $\{n_i\}$ such that

$$\mathbb{P}(|X_{n_i} - X| > i^{-1}) \leq i^{-2}$$

For any $\epsilon > 0$ we have

$$\sum_{i > \epsilon^{-1}} \mathbb{P}(|X_{n_i} - X| > \epsilon) \leq \sum_{i > \epsilon^{-1}} \mathbb{P}(|X_{n_i} - X| > i^{-1}) \leq \sum_{i > \epsilon^{-1}} \frac{1}{i^2} < \infty$$

Then, the result follows from the Borel–Cantelli Lemma. □

Notice: the different modes of convergence majorly concern with the "speed"/rate of convergence; consider the example of $\mathbb{P}(|X_n - X| > \epsilon) \leq n^{-1}$ (convergence in probability) versus $\mathbb{P}(|X_n - X| > \epsilon) \leq n^{-2}$ (almost sure convergence).

Some additional results on weak convergence

Theorem (Continuous mapping theorem)

For $g\colon \mathbb{R} \to \mathbb{R}$ is continuous (or continuous at every point of a set C such that $\mathbb{P}(X \in C) = 1$), we have

1. *if $X_n \xrightarrow{D} X$ then $g(X_n) \xrightarrow{D} g(X)$;*
2. *if $X_n \xrightarrow{P} X$ then $g(X_n) \xrightarrow{P} g(X)$;*
3. *if $X_n \xrightarrow{a.s.} X$ then $g(X_n) \xrightarrow{a.s.} g(X)$;*

Theorem (Slutsky's theorem)

If $X_n \xrightarrow{D} X$ and $Y_n \xrightarrow{P} c$ for some constant c, then

1. *$X_n + Y_n \xrightarrow{D} X + c$;*
2. *$X_n Y_n \xrightarrow{D} cX$;*
3. *$X_n / Y_n \xrightarrow{D} X/c$, provided that $c \neq 0$.*

Lemma (Portmanteau)

The following statement are equivalent (i.e., there is an "if and only if" relationship between them):

1. $X_n \xrightarrow{D} X$ ($F_{X_n}(x) \to F_X(x) \quad \forall \ x \in \mathbb{R}$ *for all continuity points of the cdf* F_X);
2. $\lim_{n\to\infty} \mathbb{E}[g(X_n)] = \mathbb{E}[g(X)]$ *for all bounded continuous functions* g.
3. $\lim_{n\to\infty} \mathbb{E}[g(X_n)] = \mathbb{E}[g(X)]$ *for all functions* g *of the form* $g(x) = f(x)1_{\{x\in[a,b]\}}$ *where* $f(x)$ *is continuous in* $[a,b]$ *and* a, b *are point of continuity of the cdf of* X.

Markov's inequality

Theorem (Markov's inequality)

A sequence of random variables $X_1, \ldots, X_n$ (that is uniformly tight) satisfies, for $\mathbb{E}[|X_n|^p] = O(1)$ for some $p > 0$,

$$\mathbb{P}(|X_n| > M) \leq \frac{\mathbb{E}[|X_n|^p]}{M^p}. \tag{3}$$

Remark: apply Markov on the r.v. $(X_n - \mathbb{E}[X_n])^2$,

$$\mathbb{P}(|X_n - \mathbb{E}[X_n]| > a) = \mathbb{P}(|X_n - \mathbb{E}[X_n]|^2 > a^2) \overset{\textit{Markov}}{\leq} \frac{\mathbb{E}[(X_n - \mathbb{E}[X_n])^2]}{a^2} = \frac{\mathrm{Var}[X_n]}{a^2},$$

known as the **Chebyshev's inequality**.

Convergence results for the sum of two RVs

Using Markov, Chebyshev, Hölder, Minkowski, and Lyapunov inequalities, we can prove the following statements:

- if $X_n \to X$ and $Y_n \to Y$, where convergence is *a.s.*, *r*-th mean or P, then

$$X_n + Y_n \to X + Y$$

  where convergence is of the same type (respectively, *a.s.*, r or P).

- One **important observation**: if $X_n \xrightarrow{D} X$ and $Y_n \xrightarrow{D} Y$, it is NOT generally true that $X_n + Y_n \xrightarrow{D} X + Y$.

- General problem: given a sequence of RVs $\{X_n\}$ with partial sum $S_n = \sum_{i=1}^n X_i$, two sequences of numbers $\{a_n\}$ and $\{b_n\}$ and a RV S, under what conditions the following convergence occurs?

$$\frac{S_n}{b_n} - a_n \to S, \quad \text{for } \; n \to \infty$$

  and in what sense?
- For example, by using the characteristics function and its uniqueness properties, we have already established:

$$\frac{1}{n} S_n \xrightarrow{D} \mu, \quad \frac{S_n - n\mu}{\sqrt{n}\sigma} \xrightarrow{D} \mathcal{N}(0,1)$$

  for $\{X_n\}$ i.i.d. with mean μ and variance σ^2.

- Restricting to the case of i.i.d. sequences of RVs $\{X_n\}$ with $\mathbb{E}[X_1] = \mu$ (so we assume that the mean exists),
  1. if $\frac{1}{n} S_n \xrightarrow{P} \mu$ we say that the sequence obeys the weak law of large numbers (WLLN);
  2. while if $\frac{1}{n} S_n \xrightarrow{a.s.} \mu$ we say that the sequence obeys the strong law of large numbers (SLLN).
- We already know that if $\{X_n\}$ is an i.i.d. sequence with $\mathbb{E}[X_1] = \mu$, then it obeys the WLLN.

Sufficient condition for the SLLN

Theorem

Let $\{X_n\}$ denote an i.i.d. sequence with $\mathbb{E}[X_1^2] < \infty$ and $\mathbb{E}[X_1] = \mu$. Then,

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to \mu, \quad \textit{for} \;\; n \to \infty$$

almost surely and in mean-square sense.

Proof.

In order to show m.s. convergence, we write:

$$\mathbb{E}\left[\left|\frac{1}{n}S_n - \mu\right|^2\right] = \frac{1}{n^2}\mathbb{E}\left[\left|\sum_{i=1}^{n} X_i - n\mu\right|^2\right] = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}(X_i) \to 0$$

In order to show a.s. convergence we have to work a bit harder. □

Discussion

- The conditions in the theorem above are both necessary and sufficient for the convergence in mean square.
- For almost sure convergence, the condition $\mathbb{E}[|X_1|] < \infty$ is necessary and sufficient, but the proof is considerably more involved.
- There exist sequences that satisfy the WLLN but NOT the SLLN.

Example: t-statistic

T-statistic

Let $X_1, \ldots, X_n$ be i.i.d. RVs with $\mathbb{E}[X_i] = 0$ and $\mathbb{E}[X_i^2] = \sigma^2 < \infty$. Then, the so-called t-statistic

$$\sqrt{n} \cdot \frac{\frac{1}{n}\sum_{i=1}^n X_i}{\sqrt{\frac{1}{n-1}\sum_{i=1}^n \left(X_i - \frac{1}{n}\sum_{i=1}^n X_i\right)^2}} \tag{4}$$

is standard normal, with $\frac{1}{n}\sum_{i=1}^n X_i$ the sample mean and $\frac{1}{n-1}\sum_{i=1}^n \left(X_i - \frac{1}{n}\sum_{i=1}^n X_i\right)^2$ the sample variance.

Proof.

- WLLN and continuous-mapping theorem: $\frac{1}{n-1}\sum_{i=1}^n \left(X_i - \frac{1}{n}\sum_{i=1}^n X_i\right)^2 \xrightarrow{P} 1 \cdot ((\mathbb{E}[X_i])^2 - \mathbb{E}[X_i^2]) = \mathrm{Var}[X_i] = \sigma^2$
- continuous-mapping theorem: $\sqrt{(\cdot)} \xrightarrow{P} \sigma$
- by CLT, $\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n X_i\right) \xrightarrow{D} \mathcal{N}(0, \sigma^2)$
- Slutsky: conclude the proof

□

Stochastic small-o and big-O notations

Definition (Small-o and big-O notations)

We say

- $X_n = o_P(1)$ or simply $X_n = o(1)$ if the sequence $X_n \xrightarrow{P} 0$;
- $X_n = O_P(1)$ or simply $X_n = O(1)$ if the sequence X_n is bounded in probability.

Rules of calculus

- $o(1) + o(1) = o(1)$
- $o(1) + O(1) = O(1)$
- $O(1)o(1) = o(1)$
- $(1 + o(1))^{-1} = O(1)$
- etc.

"Proof" of t-statistics

For $\sqrt{n} \cdot \frac{\frac{1}{n}\sum_{i=1}^n X_i}{\sqrt{\frac{1}{n-1}\sum_{i=1}^n \left(X_i - \frac{1}{n}\sum_{i=1}^n X_i\right)^2}}$ with $\mathbb{E}[X_i] = 0$ and $\mathbb{E}[X_i^2] = \sigma^2 < \infty$, we write

- $\frac{1}{n-1}\sum_{i=1}^n \left(X_i - \frac{1}{n}\sum_{i=1}^n X_i\right)^2 = \sigma^2 + o(1)$ so $\sqrt{\cdot} = \sigma + o(1)$
- $\frac{1}{n}\sum_{i=1}^n X_i = 0 + o(1)$ but useless! In fact, by CLT $\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n X_i\right) \xrightarrow{D} \mathcal{N}(0, \sigma^2)$

Delta method

Question:

- Given an estimator T_n for some parameter θ
- our quantity of interest is $\phi(\theta)$ for some known function $\phi(\cdot)$
- a natural estimator is $\phi(T_n)$
- but what do we know about $\phi(T_n)$ from T_n?

Answer:

- by continuous mapping theorem, if $T_n \xrightarrow{P} \theta$ and ϕ is continuous, then $\phi(T_n) \xrightarrow{P} \phi(\theta)$
- of greater interest is **limiting distribution**: if $\sqrt{n}(T_n - \theta)$ converges weakly (in distribution) to some limiting distribution, what about $\sqrt{n}\,(\phi(T_n) - \phi(\theta))$?
- if $\phi(\cdot)$ is differentiable, then **YES**! Informally, by Taylor-expansion,

$$\sqrt{n}\,(\phi(T_n) - \phi(\theta)) \approx \phi'(\theta)\sqrt{n}(T_n - \theta). \tag{5}$$

- in particular, if $\sqrt{n}(T_n - \theta)$ is **asymptotically normal** $\sqrt{n}(T_n - \theta) \xrightarrow{D} \mathcal{N}(0, \sigma^2)$, we expect that

$$\sqrt{n}\,(\phi(T_n) - \phi(\theta)) \approx \mathcal{N}(0, \phi'(\theta)^2\sigma^2). \tag{6}$$

Multivariate delta method

Theorem (Multivariate delta method)

Let $\phi\colon \mathbb{R}^k \to \mathbb{R}^m$ be a map defined on (a possibly subset of) $\mathbb{R}^k$ and differentiable at $\theta \in \mathbb{R}^k$. Let T_n be random vectors taking values in the domain of ϕ. If $r_n(T_n - \theta) \xrightarrow{D} T$ for numbers $r_n \to \infty$, then

$$r_n \left(\phi(T_n) - \phi(\theta)\right) \xrightarrow{D} \phi'(T). \tag{7}$$

Moreover, the difference between $r_n \left(\phi(T_n) - \phi(\theta)\right) - \phi' \left(r_n(T_n - \theta)\right) \xrightarrow{P} 0$.

Example: sample variance

Sample variance

For n observations $X_1, \ldots, X_n$, the sample variance is defined as

$$S^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2 = \phi(\overline{X}, \overline{X^2}), \quad \overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i, \tag{8}$$

for the function $\phi(x, y) = y - x^2$. Then, for X_i drawn from some distribution with first to fourth moments m_1, m_2, m_3, m_4, we have, by multivariate CLT that

$$\sqrt{n} \left(\begin{bmatrix} \overline{X} \\ \overline{X^2} \end{bmatrix} - \begin{bmatrix} m_1 \\ m_2 \end{bmatrix} \right) \xrightarrow{D} \mathcal{N} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} m_2 - m_1^2 & m_3 - m_1 m_2 \\ m_3 - m_1 m_2 & m_4 - m_2^2 \end{bmatrix} \right) \equiv \begin{bmatrix} T_1 \\ T_2 \end{bmatrix}. \tag{9}$$

Since ϕ is differentiable at $\binom{m_1}{m_2}$, with derivative $\phi'(x, y) = \binom{-2x}{1}$. Then, it follows from multivariate delta method that

$$\sqrt{n} \left(\phi(\overline{X}, \overline{X^2}) - \phi(m_1, m_2) \right) \xrightarrow{D} -2 m_1 T_1 + T_2. \tag{10}$$

In the case of $m_1 = 0$, we get $\boxed{\sqrt{n}(S^2 - m_2) \xrightarrow{D} \mathcal{N}(0, m_4 - m_2^2)}$.

Exercises

Exercise 1: convergence in probability does not imply a.s. convergence

Consider the independent sequence of RVs $X_1, X_2, \ldots,$ defined by

$$X_n = \begin{cases} 1 & \text{with prob. } n^{-1} \\ 0 & \text{with prob. } 1 - n^{-1} \end{cases}$$

Show that $X_n \xrightarrow{P} 0$ but $\{X_n\}$ does not converge almost surely to 0.

Exercise 2: convergence in probability does not imply r-th mean

Consider the sequence of RVs $X_1, X_2, \ldots,$ defined by

$$X_n = \begin{cases} n^3 & \text{with prob. } n^{-2} \\ 0 & \text{with prob. } 1 - n^{-2} \end{cases}$$

Show that $X_n \xrightarrow{P} 0$ but $\{X_n\}$ does not converge to 0 in the mean.

Thank you! Q & A?