Probability and Stochastic Processes: Chapter 5: Stochastic Convergence

Zhenyu Liao

School of Electronic Information and Communications, HUST

December 6, 2022

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.

- The infimum of a set of numbers $A = \{a_1, a_2, \dots\}$ 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 $A = \{a_1, a_2, \dots\}$ 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

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

- 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.

- 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 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)

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.

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 $|X_n - Y| = |X - 1 + X| = 1$ 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:

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

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

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

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. *If $X_n \xrightarrow{D} c$, where c is a constant, then $X_n \xrightarrow{P} c$.*
2. *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 $r \geq 1$.*
3. *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 $F(x)$.

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* $\epsilon > 0$ *(**Borel–Cantelli**).*
3. *If* $X_n \xrightarrow{a.s.} X$, *then* $X_n \xrightarrow{P} X$, *but the converse generally fails.*

Borel Cantelli Lemmas

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

Example

Infinite monkey theorem: endless typing at random will, with probability one, eventually produce every finite text (such as the works of Shakespeare).

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. □

FYI: the concept of sub-sequence is commonly used. We know, for example, that if the sequence $\{X_n\} \to X$ then every sub-sequence of $\{X_n\}$ converges to (the same limit) X; and by Bolzano-Weierstrass Theorem that every bounded sequence contains a convergent sub-sequence.

Some additional results on weak convergence

- **Continuous mapping theorem**: if $X_n \xrightarrow{D} X$ and $g \colon \mathbb{R} \to \mathbb{R}$ is continuous, then $g(X_n) \xrightarrow{D} g(X)$.
- **Slutsky's theorem**: if $X_n \xrightarrow{D} X$ and $Y_n \xrightarrow{P} Y$ for Y being a constant, then
  1. $X_n + Y_n \xrightarrow{D} X + Y$;
  2. $X_n Y_n \xrightarrow{D} XY$;
  3. $X_n / Y_n \xrightarrow{D} X/Y$, provided that $Y \neq 0$.
- The following statement are equivalent (i.e., there is an "if and only if" relationship between them):
  1. $X_n \xrightarrow{D} X$;
  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 $g(x) = f(x) I_{\{x \in [a,b]\}}$ where $f(x)$ is continuous in $[a,b]$ and a,b are point of continuity of the cdf of X.

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. In fact, $\frac{1}{n} S_n \xrightarrow{D} \mu$ implies $\frac{1}{n} S_n \xrightarrow{P} \mu$ (i.e., weak convergence implies convergence in probability, if the corresponding limit is a constant).

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.

We skip this proof

1) We can find explicitly a sub-sequence that converges almost surely. Let $n_i = i^2$, then using Chebyshev's inequality we have

$$\mathbb{P}\left(\frac{1}{i^2}\left|S_{i^2} - i^2\mu\right| > \epsilon\right) \leq \frac{\operatorname{Var}(S_{i^2})}{i^4\epsilon^2} = \frac{\operatorname{Var}(X_1)}{i^2\epsilon^2}$$

Summing over i and using Statement 2 of Lemma 9, we have that $i^{-2}S_{i^2} \xrightarrow{a.s.} \mu$.
2) Now we need to fill in the gaps, i.e., for any i consider the integers n in between i^2 and $(i+1)^2$. We start by assuming that the X_n's are non-negative, such that S_n is monotonically increasing. In this case

$$S_{i^2} \leq S_n \leq S_{(i+1)^2}$$

which implies

$$\frac{1}{(i+1)^2}S_{i^2} \leq \frac{1}{n}S_n \leq \frac{1}{i^2}S_{(i+1)^2}$$

Letting $i \to \infty$ and the fact that $i^2/(i+1)^2 \to 1$, we deduce that $\frac{1}{n} S_n \xrightarrow{a.s.} \mu$.
3) For general X_n (not necessarily non-negative), we write $X_n(\omega) = X_n^+(\omega) - X_n^-(\omega)$ where the positive and negative parts of X_n are defined as

$$X_n^+(\omega) = \max\{0, X_n(\omega)\}, \quad X_n^-(\omega) = -\min\{0, X_n(\omega)\}$$

and notice that $X_n = X_n^+ - X_n^-$, $\mathbb{E}[X_n] = \mathbb{E}[X_n^+] - \mathbb{E}[X_n^-]$ and since $X_n^+ \leq |X_n|$ and $X_n^- \leq |X_n|$, then also $\mathbb{E}[|X_n^+|^2] < \infty$ and $\mathbb{E}[|X_n^-|^2] < \infty$.
Then, convergence can be proved individually for $\frac{1}{n}\sum_{i=1}^n X_i^+$ and for $\frac{1}{n}\sum_{i=1}^n X_i^-$ and using the a.s. convergence of the sum, we get the result.

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.

Thank you! Q & A?