AI 5090: Stochastic Processes

Lecture 11

Scribe: Abhishek Kumar , Siddharth Saha

11 February 2025

Contents

1 Some Reverse Implications

We saw several notions of convergence of sequence of random variables (pointwise, almost-sure, mean-squared, in-probability, and in-distribution), and also implications between some of these notions. In particular, we saw the following implications.

- Pointwise convergence implies almost-sure convergence.
- Almost-sure convergence implies convergence in probability.
- Mean-squared convergence implies convergence in probability.
- Convergence in probability implies convergence in distribution.

Other implications do not hold in general. However, some implications in reverse direction hold under stronger conditions, as outlined below.

1.1 Convergence in Distribution $\Rightarrow$ Convergence in Probability?

Special Case : When we have convergence in distribution of a random variable to a constant random variable, we have a special case where there is convergence in probability to the same constant random variable as well.

Proposition 1. *Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Let $\{X_n\}_{n=1}^{\infty}$ be defined with respect to $\mathscr{F}$. For any $c \in \mathbb{R}$,*

$$X_n \xrightarrow{\text{d.}} c \quad \Rightarrow \quad X_n \xrightarrow{\text{p.}} c.$$

Proof. We start from RHS.

$$X_n \xrightarrow{\text{p.}} c. \quad \Rightarrow \quad \forall \varepsilon > 0, \quad \lim_{n\to\infty} \mathbb{P}(|X_n - X| \geq \varepsilon) = 0.$$

Then, for every $\varepsilon > 0$, we have

$$\begin{aligned}\mathbb{P}(|X_n - X| > \varepsilon) &= \mathbb{P}(\{X_n - X > \varepsilon\} \cup \{X_n - X < -\varepsilon\}) \\ &= \mathbb{P}(X_n - X > \varepsilon) + \mathbb{P}(X_n - X < -\varepsilon).\end{aligned}$$

All we need to prove is convergence to a constant random variable. Hence, take X as a constant random variable:

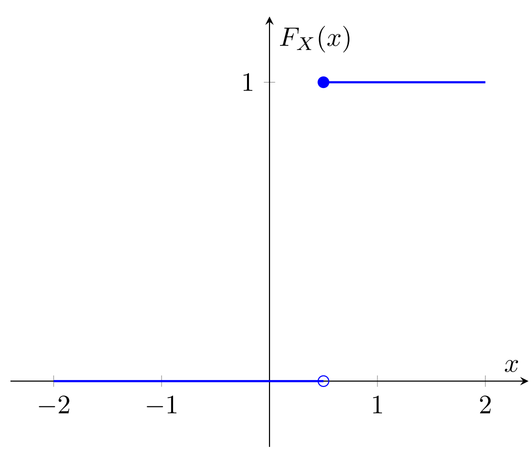
$$X(\omega) = c, \quad \forall \omega \in \Omega, \quad \text{where } c \in \mathbb{R} \text{ is a constant.}$$

Then, we have

$$\begin{aligned}\mathbb{P}(|X_n - X| > \varepsilon) &= \mathbb{P}(X_n > c + \varepsilon) + \mathbb{P}(X_n - c < -\varepsilon) \\ &= 1 - \mathbb{P}(X_n \leq c + \varepsilon) + \mathbb{P}(X_n - c < -\varepsilon) \\ &\leq 1 - \mathbb{P}(X_n \leq c + \varepsilon) + \mathbb{P}(X_n - c \leq -\varepsilon) \\ &\leq \mathbb{P}(|X_n - X| > \varepsilon) \leq 1 - F_{X_n}(c + \varepsilon) + F_{X_n}(c - \varepsilon),\end{aligned}$$

where the penultimate line above follows noting from the axioms of probability measure that

$$\mathbb{P}(X_n - c < -\varepsilon) \leq \mathbb{P}(X_n - c < -\varepsilon) + \mathbb{P}(X_n - c = -\varepsilon) = \mathbb{P}(X_n - c \leq \varepsilon).$$



Note that the CDF of X exhibits a jump at the point $c \in \mathbb{R}$; in particular, $F_X(c-\varepsilon) = 0$ and $F_X(c+\varepsilon) = 0$. Therefore, it follows that

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) &\leq 1 - \lim_{n\to\infty} F_{X_n}(c+\varepsilon) + \lim_{n\to\infty} F_{X_n}(c-\varepsilon) \\ &= 1 - F_X(c+\varepsilon) + F_X(c-\varepsilon) \\ &\text{[using convergence in distribution]} \\ &= 1 - 1 + 0 \\ &= 0,\end{aligned}$$

thus establishing that $\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$ for every $\varepsilon > 0$, or equivalently $X_n \xrightarrow{\text{p.}} c$. □

1.2 Convergence in Probability ⇒ Implies Mean-Squared Convergence?

We now outline one special case under which convergence in probability implies mean-squared convergence.

Proposition 2. *Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Let $\{X_n\}_{n=1}^{\infty}$ be defined with respect to $\mathscr{F}$. Furthermore, let X, Y be defined with respect to $\mathscr{F}$. Suppose that the following conditions hold.*

1. *$E[X_n^2] < +\infty$ for all $n \in \mathbb{N}$, i.e., all random variables in the sequence have finite second moment.*
2. *$\mathbb{P}(|X_n| \leq Y) = 1$ for all n, with $E[Y^2] < +\infty$, i.e., all in $\{X_n\}_{n=1}^{\infty}$ are dominated by Y with finite second moment.*

Then,

$$X_n \xrightarrow{p.} X \quad \Rightarrow \quad X_n \xrightarrow{m.s.} X.$$

Remark 1. *Some remarks are in order.*

1. *Recall that convergence in probability implies that for every $\varepsilon > 0$,*

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

This implies that for large n , the probability that X_n deviates from X more than ε goes to 0. However, with small probability, X_n can potentially deviate from X by a large amount. Hence, merely convergence in probability does not imply convergence in the mean-squared error between X_n and X.

2. *In the above point, we also note the unboundedness of each X_n, thereby necessitating us to impose condition of domination.*
3. *Conditioning on 2nd moments of random variables comes from the requirement of having convergence of mean-squared error.*
4. *The random variable Y may be a constant random variable.*

2 Limit Theorems

Limit Theorems provide an operational meaning to quantities such as expected values as the limit of certain averages, hence providing the interpretation of expected value as the "mean" of a random variable.

2.1 Tools

2.1.1 Characteristic Function

Definition 3 (Characteristic Function)**.** *Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Let X be any random variiable defined with respect to $\mathscr{F}$. The characteristic function of X is a mapping $C_X : \mathbb{R} \to \mathbb{C}$, defined as*

$$C_X(s) = \mathbb{E}[e^{\mathrm{j}sX}] = \mathbb{E}[\cos(sX)] + \mathrm{j}\,\mathbb{E}[\sin(sX)], \qquad s \in \mathbb{R}.$$

Remark 2. *Some remarks are in order.*

1. *The characteristic function is always finite in magnitude, i.e.,*

$$|C_X(s)| \le 1, \qquad \forall s \in \mathbb{R}.$$

2. *Each CDF has a unique characteristic function associated with it. That is, the characteristic function uniquely characterize its CDF. Thus there is bijective map between a CDF and a characteristic function (one-one + onto). We can also work with characteristic function instead of working with CDF.*

2.1.2 Taylor Series for Characteristic Function

Characteristic functions admit a Taylor's series expansion about the point $s = 0$ under some regularity conditions, as outlined in the below result.

Proposition 4. *Suppose that $\mathbb{E}[|X|^k] < +\infty$ for some $k \in \mathbb{N}$, i.e., the kth absolute moment of X is finite for some $k \in \mathbb{N}$. Then,*

$$C_X(s) = \sum_{m=0}^{k} \frac{\mathbb{E}[X^m]\,(sj)^m}{m!} + o(s^k), \qquad \forall s \in \mathbb{R}.$$

Remark 3. *Some remarks on the proposition are in order.*

1. *If kth absolute moment of X is finite, then all absolute moments of X up to order k are finite. This bounds all terms on the right-hand side of the Taylor series expansion for C_X.*
2. *$o(s^k)$ term decays faster than s^k.*
3. *For a Gaussian random variable X, all absolute moments are finite, and hence we can expand $C_X(s)$ upto any desired moment.*

2.1.3 Convergence in Distribution and Convergence of Characteristic Functions

The below result shows that convergence in distribution is equivalent to pointwise convergence of the associated characteristic functions.

Proposition 5. *Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Let $\{X_n\}_{n=1}^{\infty}$ and X be defined with respect to $\mathscr{F}$. Then,*

$$X_n \xrightarrow{d} X \quad \Longleftrightarrow \quad C_{X_n}(s) \xrightarrow{n\to\infty} C_X(s) \quad \forall s \in \mathbb{R}.$$

2.2 Weak Law of Large Numbers

In this section, we state and prove the weak law of large numbers, a key result in probability theory.

Theorem 6. *Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Let $\{X_n\}_{n=1}^{\infty}$ be defined with respect to $\mathscr{F}$. Let $\{X_n\}_{n=1}^{\infty}$ be i.i.d with $\mathbb{E}[|X_1|] < +\infty$. Further, let $\mathbb{E}[X_1] = \mu$. For each $n \in \mathbb{N}$, let $S_n = \sum_{k=1}^{n} X_k$. Then,*

$$\frac{S_n}{n} \xrightarrow{\text{p.}} \mu.$$

More formally,

$$\forall \varepsilon > 0, \qquad \lim_{n\to\infty} \mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) = 0.$$

2.2.1 Proof Under Finite Variance Assumption

In this section, we present a proof for the weak law of large numbers under the assumption that $\text{Var}(X_1) < +\infty$. From Jensen's inequality, we also note that

$$+\infty > \mathbb{E}[|X|] \geq |\mathbb{E}[X]|.$$

Then, for any $\varepsilon > 0$, we have

$$\begin{aligned}
\mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) &\leq \frac{\mathbb{E}[(\frac{S_n}{n} - \mu)^2]}{\varepsilon^2} \qquad \text{[Chebyshev's inequality]} \\
&\leq \frac{\mathbb{E}\left[\left(\frac{S_n}{n} - \mathbb{E}\left[\frac{S_n}{n}\right]\right)^2\right]}{\varepsilon^2} \\
&\leq \frac{\text{Var}(\frac{S_n}{n})}{\varepsilon^2} \\
&= \frac{\text{Var}(X_1)}{n\varepsilon^2}.
\end{aligned}$$

Taking limits as $n \to \infty$ on both sides of the above relation, we get

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) = 0.$$

Because this holds for every choice of $\varepsilon > 0$, we see that $\frac{S_n}{n} \xrightarrow{\text{p.}} \mu$. This completes the desired proof.

2.2.2 Proof Under Only Finite First Moment Assumption

In this section, we present a more general proof for the weak law of large numbers, one that relies on only the first moment being finite. The overall idea in the proof will be to use the result of Section 2.1.3 to prove convergence in distribution, and subsequently invoke the result of Section 1.1 to reach the end goal.
For any $s \in \mathbb{R}$, we note that

$$\begin{aligned}
C_{\frac{S_n}{n}}(s) &= \mathbb{E}[e^{\mathrm{j}s\frac{S_n}{n}}] \\
&= \mathbb{E}[e^{\frac{s\mathrm{j}}{n}\sum_{k=1}^{n} X_k}] \\
&= \prod_{k=1}^{n} \mathbb{E}[e^{\mathrm{j}\frac{s}{n}X_k}] \qquad \text{[using independence of } X_1, \ldots, X_n\text{]} \\
`&= \prod_{k=1}^{n} \mathbb{E}[e^{\mathrm{j}\frac{s}{n}X_1}] \qquad \text{[using the fact that } X_1, \ldots, X_n \text{ are identically distributed]} \\
&= \left(C_{X_1}\left(\frac{s}{n}\right)\right)^n \\
&= \left(\sum_{m=0}^{1} \frac{\mathbb{E}[(X_1)^m]\,\mathrm{j}^m\,(\frac{s}{n})^m}{m!} + o\left(\frac{s}{n}\right)\right)^n \qquad \text{[applying Taylor series expansion upto first moment term]} \\
&= \left(1 + \frac{\mu\mathrm{j}s}{n} + o\left(\frac{s}{n}\right)\right)^n.
\end{aligned}$$

Taking limits as $n \to \infty$ on both sides of the above relation, we get

$$\lim_{n\to\infty} C_{\frac{S_n}{n}}(s) = e^{\mathrm{j}s\mu} = C_\mu(s).$$

Because the above relation holds for every choice of $s \in \mathbb{R}$, we get that $\frac{S_n}{n} \xrightarrow{\text{d.}} \mu$, and thereby in turn $\frac{S_n}{n} \xrightarrow{\text{p.}} \mu$.

2.3 Strong Law of Large Numbers

Theorem 7. *Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Let $\{X_n\}_{n=1}^{\infty}$ be defined with respect to $\mathscr{F}$. Let $\{X_n\}_{n=1}^{\infty}$ be i.i.d with $\mathbb{E}[|X_1|] < +\infty$. Further assume $\mathbb{E}[X_1] = \mu$. For each $n \in \mathbb{N}$, let S_n be defined as before. Then,*

$$\frac{S_n}{n} \xrightarrow{\text{a.s.}} \mu.$$

More formally,

$$\mathbb{P}\left(\left\{\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right\} \text{ i.o.}\right) = 0 \qquad \forall \varepsilon > 0.$$

This is essentially the weak law of large numbers, except that instead of the sequence converging in probability, the convergence stated here is almost-sure convergence. It is important to note that this theorem, is only a hundred years old, as opposed to the Weak Law of Large Numbers which was proved in 1771. We have retained the weak law for mostly historical reasons.

It is the strong law that gives the operational meaning to expectation as an "average".

Another point to note is that the strong law only works when the numerator is a sum. We cannot take other operations such as max, or min for the numerator, as otherwise the strong law does not hold.

2.3.1 Proof of SSLN under Finite Fourth Moment Assumption

Note: We assume that the 4th moment is finite to keep this proof brief. The proof without this assumption is much more involved.
Suppose that $\mathbb{E}[X_1^4] < +\infty$. Assume, without loss of generality, that $\mathbb{E}[X_1] = 0$. We want to show

$$\frac{S_n}{n} \xrightarrow{\text{a.s.}} 0.$$

Fix an arbitrary $\varepsilon > 0$. Then,

$$\begin{aligned}
\mathbb{P}\left(\left|\frac{S_n}{n}\right| > \varepsilon\right) &\leq \frac{\mathbb{E}[(S_n)^4]}{n^4\varepsilon^4} \\
&= \frac{\mathbb{E}[(X_1 + \cdots + X_n)^4]}{n^4\varepsilon^4} \\
&= \frac{n\mathbb{E}[X_1^4] + 6\binom{n}{2}(\text{Var}(X_1))^2}{n^4} \\
&\leq \frac{\mathbb{E}[X_1^4]}{n^3} + 3\text{Var}(X_1)\frac{1}{n^2} = O\left(\frac{1}{n^3}\right) + O\left(\frac{1}{n^2}\right).
\end{aligned}$$

Because the right-hand side of the above relation is summable, by the Borel–Cantelli lemma, we have

$$\mathbb{P}\left(\left\{\left|\frac{S_n}{n}\right| > \varepsilon\right\} \text{ i.o.}\right) = 0.$$

Because the choice of $\varepsilon > 0$ is arbitrary, we conclude that $\frac{S_n}{n} \xrightarrow{\text{a.s.}} 0$.

2.4 Central Limit Theorem

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Let $\{X_n\}_{n=1}^{\infty}$ be defined w.r.t. $\mathscr{F}$.

Theorem 8. *Let $\{X_n\}_{n=1}^{\infty}$ be i.i.d. with* Var$(X_1) < +\infty$. *Let*

$$S_n = \sum_{i=1}^{n} X_i.$$

Then,

$$\frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathsf{Var}(S_n)}} \xrightarrow{\text{d.}} X, \quad X \sim \mathcal{N}(0,1).$$

More formally,

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathsf{Var}(S_n)}} \leq x\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dt, \quad \forall x \in \mathbb{R}.$$

The Central Limit Theorem (CLT) extends beyond the assumption of independent and identically distributed (i.i.d.) random variables. In a broader sense, the CLT states that the deviation of the sum S_n from its expected value, $S_n - n\mathbb{E}[X_i]$, grows at a sublinear rate. This aligns with the Strong Law of Large Numbers, which ensures that the sample mean converges to the true mean. More precisely, this deviation scales as $O(\sqrt{n})$.
The key insight of the CLT is that when we normalize this deviation by $\sqrt{n}$, the resulting distribution converges to a standard Gaussian distribution. This fundamental result will be revisited in Lecture 12.
.