AI 5090: Stochastic Processes

Lecture 07

Scribe: N. Sri Krishna Chaitanya

28 January 2025

Contents

Notation

We are going to use the following notation throughout the document:

$$A_n \text{ i.o.} = \bigcap_{N=1}^{\infty} \bigcup_{k=N} A_k.$$

This represents the limit supremum of a sequence of events, which will be useful in the discussions ahead.

1 Borel–Cantelli Lemma

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. The Borel–Cantelli Lemma is a fundamental result that describes conditions under which an event occurs infinitely often.

Lemma 1 (Borel–Cantelli Lemma)**.**

1. *Suppose $A_1, A_2, \ldots \in \mathscr{F}$ are such that $\sum_{i=1}^{\infty} \mathbb{P}(A_i) < +\infty$. Then,*

$$\mathbb{P}\left(A_n \textit{ i.o. }\right) = 0.$$

Thus, if the sum of probabilities of events is finite, then almost surely only finitely many of these events occur.

2. *Suppose $A_1, A_2, \ldots \in \mathscr{F}$ are independent and satisfy $\sum_{i=1}^{\infty} \mathbb{P}(A_i) = +\infty$. Then,*

$$\mathbb{P}\left(A_n \textit{ i.o. }\right) = 1.$$

In this case, if the sum of probabilities of independent events diverges, then these events necessarily occur infinitely often with probability 1.

1.1 Proof of the First Borel–Cantelli Lemma

We define the event that infinitely many of the A_n occur as

$$\limsup_{n\to\infty} A_n = \bigcap_{m=1}^{\infty} \bigcup_{n\geq m} A_n.$$

By applying Boole's inequality, we get

$$\mathbb{P}\left(\bigcup_{n\geq m} A_n\right) \leq \sum_{n\geq m} \mathbb{P}(A_n).$$

Since $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, the right-hand side tends to zero as $m \to \infty$, implying

$$\mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = 0.$$

Thus, the probability that infinitely many A_n occur is zero, proving the first part of the lemma.

1.2 Proof of the Second Borel–Cantelli Lemma

Now, assume that the events $\{A_n\}$ are independent and that

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

We want to show that

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

The probability that only finitely many A_n occur is given by

$$\mathbb{P}(A_n^c \text{ for all but finitely many } n).$$

Since the A_n are independent, we compute:

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

Using the standard result that

$$\prod_{n=1}^{\infty}(1 - P(A_n)) = 0 \quad \text{if} \quad \sum_{n=1}^{\infty} P(A_n) = \infty,$$

it follows that

$$P(A_n \text{ i.o.}) = 1.$$

This completes the proof.
For a more detailed proof, we refer the reader to [GS20, Ch. 7, Sec. 7.3].

2 An Alternative Characterisation of Almost-Sure Convergence

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of random variables and X a limiting random variable, all defined with respect to $\mathscr{F}$.

Proposition 1 (Almost Sure Convergence)**.** *The following statements are equivalent:*

1. $X_n \xrightarrow{\text{a.s.}} X$ *(which means X_n converges to X almost surely).*
2. *For every $\varepsilon > 0$,*
$$\mathbb{P}(|X_n - X| \geq \varepsilon \text{ i.o. }) = 0.$$
This means that the probability of the sequence deviating from X by at least ε infinitely often is zero, for every $\varepsilon > 0$.

Proof. To show the equivalence, we use the following logical steps:

$$\begin{aligned}
X_n \xrightarrow{\text{a.s.}} X \quad &\iff \quad \mathbb{P}\left(\bigcap_{q\in\mathbb{Q}_+}\bigcup_{N=1}^{\infty}\bigcap_{n=N}^{\infty}\{|X_n - X| < q\}\right) = 1 \\
&\iff \quad \mathbb{P}\left(\bigcup_{q\in\mathbb{Q}_+}\bigcap_{N=1}^{\infty}\bigcup_{n=N}^{\infty}\{|X_n - X| \geq q\}\right) = 0 \\
&\iff \quad \mathbb{P}\left(\bigcap_{N=1}^{\infty}\bigcup_{n=N}^{\infty}\{|X_n - X| \geq q\}\right) = 0 \quad \forall q \in \mathbb{Q}_+ \\
&\overset{(a)}{\iff} \quad \mathbb{P}\left(\bigcap_{N=1}^{\infty}\bigcup_{n=N}^{\infty}\{|X_n - X| \geq \varepsilon\}\right) = 0 \quad \forall \varepsilon > 0 \\
&\iff \quad \mathbb{P}\left(|X_n - X| \geq \varepsilon \text{ i.o. }\right) = 0 \quad \forall \varepsilon > 0.
\end{aligned}$$

In the above sequence of steps, (a) follows from the fact that the set of rational numbers is dense in the set of real numbers; for a proof of this fact, see [Rud21, Theorem 1.20]. Thus, we have shown that almost sure convergence is equivalent to the given probability condition. □

Consider the following example.

Example 1. *Let $X_1, X_2, \ldots$ be independent and identically distributed (i.i.d.) random variables following a Bernoulli distribution with a fixed probability $p \in (0, 1)$:*

$$X_i \sim \textit{Ber}(p) \quad \textit{i.i.d.}$$

For each $n \in \mathbb{N}$, define the number of heads in the first n tosses as

$$S_n = \sum_{i=1}^{n} X_i = \#\textit{Heads in first } n \textit{ tosses}. \tag{1}$$

We aim to show that:

$$\frac{S_n}{n} \xrightarrow{\text{a.s.}} p, \tag{2}$$

which means that S_n/n converges almost surely to p.

2.1 Proof using Chebyshev's Inequality

Since $X_1, X_2, \ldots$ are i.i.d. Bernoulli(p), we have:

$$\mathbb{E}[S_n] = \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{E}[X_i] = np.$$

Using variance properties:

$$\mathsf{Var}(S_n) = \sum_{i=1}^{n} \mathsf{Var}(X_i) = \sum_{i=1}^{n} p(1-p) = np(1-p).$$

Now, applying Chebyshev's inequality:

$$\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| > \varepsilon\right) = \mathbb{P}\left(|S_n - np| > n\varepsilon\right). \tag{3}$$

By Chebyshev's inequality:

$$\mathbb{P}\left(|S_n - np| > n\varepsilon\right) \leq \frac{\mathsf{Var}(S_n)}{n^2\varepsilon^2} = \frac{np(1-p)}{n^2\varepsilon^2} = \frac{p(1-p)}{n\varepsilon^2}.$$

Notice that the right-hand side of the above upper bound is not summable (as it exhibits $1/n$ decay). Therefore, the above bounding technique does not suffice to prove almost-sure convergence. We will need a "tighter" upper bound (one that is summable) to show almost-sure convergence.

3 Stronger Bound using Chernoff's Inequality

A stronger bound is given by Chernoff's inequality:

$$\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| > \varepsilon\right) \leq O(e^{-nc}) \text{ for some } c > 0. \tag{4}$$

This shows an exponentially fast decay, strengthening the convergence result. For details on how to derive such a bound, the reader is referred to Question 4 in this document. From (4), we note that

$$\sum_{n\in\mathbb{N}} \mathbb{P}\left(\left|\frac{S_n}{n} - p\right| > \varepsilon\right) < +\infty, \qquad \forall \varepsilon > 0.$$

Hence, from the first part of Borel–Cantelli lemma, it follows that

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

hence proving that $\frac{S_n}{n} \xrightarrow{\text{a.s.}} p$.

4 Example: Moving Rectangles

Suppose that $(\Omega, \mathscr{F}, \mathbb{P}) = ([0,1], \mathcal{B}([0,1]), \text{Uniform})$.
Define a sequence of indicator functions:

$$X_1 = \mathbf{1}_{[0,1]}$$

$$
\begin{aligned}
X_2 &= \mathbf{1}_{[0,\frac{1}{2}]} \\
X_3 &= \mathbf{1}_{[\frac{1}{2},1]} \\
X_4 &= \mathbf{1}_{[0,\frac{1}{4}]} \\
X_5 &= \mathbf{1}_{[\frac{1}{4},\frac{1}{2}]} \\
X_6 &= \mathbf{1}_{[\frac{1}{2},\frac{3}{4}]} \\
X_7 &= \mathbf{1}_{[\frac{3}{4},1]}, \quad \text{and so on.}
\end{aligned}
$$

This sequence represents moving rectangles that shrink and shift over time. Notably, this sequence does not have a pointwise limit or an almost-sure limit.

5 Explanation

The reason why the sequence (X_n) has no pointwise or almost-sure limit is that for any given point $\omega \in [0,1]$, the indicator function $X_n(\omega)$ oscillates indefinitely between 0 and 1 without stabilizing to a fixed value. As n increases, the partitions become finer, and there is no single value to which $X_n(\omega)$ converges for any fixed $\omega \in [0,1]$.
This behavior prevents the sequence from having an almost-sure limit, since for any x, there is no stable long-term value. Moreover, this example illustrates the importance of considering different modes of convergence, which we will explore next.

6 Other Notions of Convergence

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$.

Definition 2 (Convergence in Probability)**.** *A sequence of random variables $\{X_n\}_{n=1}^{\infty}$ converges in probability to X if:*

$$\forall \varepsilon > 0, \qquad \lim_{n\to\infty} \mathbb{P}(\{|X_n - X| > \varepsilon\}) = 0. \tag{5}$$

We write

$$X_n \xrightarrow{\text{p.}} X \tag{6}$$

to denote the fact that $\{X_n\}_{n=1}^{\infty}$ converges in probability to X.

Remark 1. *The in-probability limit is only specified up to sets of zero probability. That is,*

$$X_n \xrightarrow{\text{p.}} X, \quad X_n \xrightarrow{\text{p.}} Y \quad \implies \quad \mathbb{P}(X = Y) = 1. \tag{7}$$

Definition 3 (Mean-Squared Convergence)**.** *A sequence of random variables $\{X_n\}_{n=1}^{\infty}$ converges to X in mean-squared (m.s.) sense if $\mathbb{E}[X_n^2] < +\infty$ for all $n \in \mathbb{N}$, and*

$$\lim_{n\to\infty} \mathbb{E}[(X_n - X)^2] = 0. \tag{8}$$

We write

$$X_n \xrightarrow{\text{m.s.}} X \tag{9}$$

to denote the fact that $\{X_n\}_{n=1}^{\infty}$ converges in mean-squared sense to X.

Remark 2. *The mean-squared limit is specified uniquely only up to sets of probability 0, i.e., if $X_n \xrightarrow{\text{m.s.}} X$ and $X_n \xrightarrow{\text{m.s.}} Y$, then $\mathbb{P}(X = Y) = 1$.*

Example 2. *Consider the probability space $(\Omega, \mathscr{F}, \mathbb{P}) = ([0,1], \mathcal{B}([0,1]), \text{Unif})$. Suppose we define a sequence of random variables X_n as follows:*

$$X_n(\omega) = \begin{cases} a_n, & \omega \in \left[0, \frac{1}{n}\right], \\ 0, & \textit{otherwise}, \end{cases}$$

for some fixed sequence $\{a_n\}_{n\in\mathbb{N}}$ of real numbers. We analyze different forms of convergence for X_n:

1. *If we take $a_n = n$, then*
$$X_n \xrightarrow{\text{a.s.}} 0.$$
This is because, for each fixed ω, the event $\omega \in [0, 1/n]$ occurs for finitely many n. Hence, beyond some finite n, we have $X_n(\omega) = 0$, ensuring almost-sure convergence to 0.

2. *We examine mean-squared convergence by computing*
$$\mathbb{E}[(X_n - 0)^2] = \mathbb{E}[X_n^2] = \int_0^1 X_n^2(\omega) d\omega.$$
For $a_n = n$,
$$\mathbb{E}[X_n^2] = \int_0^{1/n} n^2 d\omega = n^2 \cdot \frac{1}{n} = n \to \infty,$$
which shows that X_n does not converge in the mean-squared sense. Hence,
$$X_n \xrightarrow{\text{a.s.}} 0, \quad \textit{but} \quad X_n \stackrel{\text{m.s.}}{\not\to} 0.$$
This example illustrates that almost-sure convergence does not necessarily imply mean-squared convergence. In general,
$$X_n \xrightarrow{\text{a.s.}} 0 \quad \not\Rightarrow X_n \xrightarrow{\text{m.s.}} 0, \qquad\qquad X_n \xrightarrow{\text{m.s.}} 0 \quad \not\Rightarrow \quad X_n \xrightarrow{\text{a.s.}} 0.$$

Example 3. *Consider the probability space $(\Omega, \mathscr{F}, \mathbb{P}) = ([0, 1], \mathscr{B}([0, 1]), \textit{Unif})$. Suppose the sequence of random variables satisfies the condition:*
$$X_n = X_{n+3}, \quad \forall n \in \mathbb{N}.$$
This condition implies that X_n is a periodic sequence with period 3. We analyze the different forms of convergence:

1. **Pointwise and Almost-Sure Convergences:**

   Because X_n repeats every three steps, there is no single limit function to which $X_n(\omega)$ converges as $n \to \infty$. Hence, X_n does not converge almost-surely or pointwise.

2. **Mean-Squared and In-Probability Convergences:**

   Because X_n does not settle down to a single function, it does not converge in the mean-squared sense or in probability.

3. **A key observation:**

   Despite the lack of convergence in the previous senses, the probability mass functions (PMFs) and cumulative distribution functions (CDFs) of X_1, X_2, X_3 are identical, thereby implying that there exists a non-trivial limit for the sequence of CDFs $\{F_{X_n}\}_{n\in\mathbb{N}}$. This observation motivates a new notion of convergence known as convergence in distribution (to be covered in the next lecture).

References

[GS20] Geoffrey Grimmett and David Stirzaker. *Probability and Random Processes*. Oxford University Press, 2020.

[Rud21] Walter Rudin. *Principles of mathematical analysis*. 2021.