AI 5090: Stochastic Processes

Lecture 08

Scribe: Aashish Singh & Rohit Sathyajit

04 February 2025

Contents

1 Recap 2
1.1 Almost Sure Convergence and Borel-Cantelli Lemma 2
1.2 Generic Template for Proving Almost Sure Convergence 2
1.2.1 A Recap of Borel-Cantelli Lemma (Part 1) 3
1.2.2 Rate of Decay Requirement 3
1.3 Convergence in Probability 3
1.3.1 Key Relationship Between Convergence Modes 4
1.4 Mean Squared Convergence 4
1.4.1 Key Relationship Between Convergence Modes 4
1.5 Example: Periodic Random Sequence 5

2 Convergence of CDFs – A Subtle Point via an Example 6
2.1 Problem Setup 6
2.2 Guessing the Limit Random Variable 6
2.3 Computation of the CDF of X_n 6
2.3.1 Case 1: n even 6
2.3.2 Case 2: n odd 7
2.4 Pointwise Limit of $F_{X_n}(x)$ 7
2.5 Conclusion: Convergence in Distribution 7

3 Convergence in Distribution 8
3.1 Definition of Convergence in Distribution 8
3.2 Examples and Applications 8
3.3 Conclusion 8

4 Implications Between Types of Convergence and Limit Theorems 9
4.1 Implications Between Different Modes of Convergence 9
4.2 Limit Theorems 10

5 Review: Continuity of Probability Measure 10

6 Proof of Implications 11
6.1 Almost Sure Convergence Implies Convergence in Probability (a.s. $\Rightarrow$ p.) 11
6.2 Mean Square Convergence Implies Convergence in Probability (m.s. $\Rightarrow$ p.) 11
6.3 Convergence in Probability Implies Convergence in Distribution (p. $\Rightarrow$ d.) 12

Topics Covered: Convergence Notions, Proofs of Implications, Problems

1 Recap

In the previous lecture, we studied **almost sure convergence**, the **Borel-Cantelli lemma**, and their connections. We also saw a **generic template** for proving almost sure convergence.

1.1 Almost Sure Convergence and Borel-Cantelli Lemma

Informally, almost sure convergence is equivalent to saying that the bad event does not happen infinitely many times. Formally:

Definition (Almost Sure Convergence)

A sequence of random variables $\{X_n\}_{n=1}^{\infty}$ is said to converge **almost surely** to a random variable X if:

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

or equivalently,

$$\mathbb{P}(\{|X_n - X| \geq \varepsilon\} \text{ i.o.}) = 0 \qquad \forall \varepsilon > 0.$$

Notation:

$$X_n \xrightarrow{\text{a.s.}} X$$

This means that for almost all $\omega \in \Omega$, the sample path $X_n(\omega)$ will eventually stay within a small ε-band around $X(\omega)$, for any arbitrarily small value of $\varepsilon > 0$. This interpretation provides a useful visualization of almost sure convergence.
A key result we discussed states that:

Proposition (Characterization of Almost Sure Convergence)

$$X_n \xrightarrow{\text{a.s.}} X \iff \mathbb{P}(A_n^\varepsilon \text{ i. o.}) = 0, \quad \forall \varepsilon > 0.$$

where $A_n^\varepsilon = \{|X_n - X| \geq \varepsilon\}$ is the **bad event** that X_n deviates from X by at least ε. Almost sure convergence means that such deviations occur only finitely many times.
There are several notations for almost sure convergence which include [GS20]

$$X_n \to X \textit{ almost everywhere}, \text{ or } X_n \xrightarrow{\text{a.e.}} X,$$

$$X_n \to X \textit{ with probability 1}, \text{ or } X_n \to X \text{ w.p.1}.$$

1.2 Generic Template for Proving Almost Sure Convergence

To prove that X_n converges to X almost surely, we use the following structured approach:

1. Fix any $\varepsilon > 0$.
2. Consider the bad event:
$$A_n^\varepsilon = \{|X_n - X| \geq \varepsilon\}.$$
3. Establish that the probability of the bad event summed over all n is **finite**:
$$\sum_{n=1}^{\infty} \mathbb{P}(A_n^\varepsilon) < +\infty.$$
For proving this part, which is the hardest part, we often use **concentration inequalities**. Concentration inequalities tries to get a upper bound on the probability of the deviation (bad event).

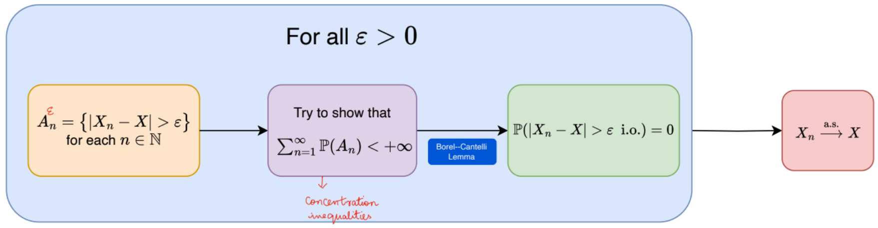

Figure 1: A generic template to prove almost-sure convergence via Borel–Cantelli lemma.

4. Apply **Borel-Cantelli Lemma (Part 1)**, which states that if the above summability condition holds, then

$$\mathbb{P}(A_n^{\varepsilon} \text{ i. o.}) = 0.$$

5. Since this holds for every $\varepsilon > 0$, it follows that X_n converges to X almost surely.

1.2.1 A Recap of Borel-Cantelli Lemma (Part 1)

The **Borel-Cantelli Lemma** (Part 1) provides a powerful tool for proving almost sure convergence.

Definition (Borel-Cantelli Lemma)

Let $A_1, A_2, \ldots \in \mathscr{F}$ be any sequence of events. If

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

then,

$$\mathbb{P}(A_n \text{ i. o.}) = 0.$$

This lemma allows us to conclude that if the probabilities of the bad events decay fast enough, then **infinitely many** of them will not occur, ensuring almost sure convergence.

1.2.2 Rate of Decay Requirement

To satisfy the Borel-Cantelli lemma, the probabilities $\mathbb{P}(A_n^{\varepsilon})$ must decay at a sufficiently fast rate. Specifically, they must follow the following:

$$\mathbb{P}(A_n^{\varepsilon}) \leq \frac{1}{n^{1+\alpha}}, \quad \text{for some } \alpha > 0.$$

This means that the probability must decrease **faster than** $1/n$. Some examples of sufficient decay rates include:

- **Polynomial decay**: $\mathbb{P}(A_n^{\varepsilon}) \leq \frac{1}{n^{1.1}}$
- **Exponential decay**: $\mathbb{P}(A_n^{\varepsilon}) \leq e^{-n}$

However, decay rates **slower than** $1/n$ (such as $1/\sqrt{n}$) are **not** sufficient.

1.3 Convergence in Probability

In the previous lecture, we also reviewed **convergence in probability**, which is a weaker mode of convergence compared to almost sure convergence. We recall the definition of convergence in probability here.

Definition 1 (Convergence in Probability). *A sequence of random variables* $\{X_n\}_{n=1}^{\infty}$ *is said to* **converge in probability** *to* X *if for every* $\varepsilon > 0$,

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

Notation:

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

Informally, the above definition states that the probability of bad events goes to 0 as $n \to \infty$. We also note here that

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

That is, if a sequence admits more than one in-probability limit, then all the limits should agree with probability 1. In other words, the limits can only be differ on sets of zero probability.

1.3.1 Key Relationship Between Convergence Modes

We established the following **implications between different types of convergence**:

Proposition (Implication of Convergence Modes)

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

That is, **almost sure convergence implies convergence in probability**, but the reverse is not necessarily true.

1.4 Mean Squared Convergence

Another mode of convergence we studied is **mean squared convergence**.

Definition (Mean Squared Convergence)

A sequence $\{X_n\}_{n=1}^{\infty}$ **converges in mean square** to X if

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

For this definition to be valid, we require that all the random variables X_n have finite second moments:

$$\mathbb{E}[X_n^2] < +\infty$$

Note

$$X_n \xrightarrow{\text{m.s.}} X, \quad X_n \xrightarrow{\text{m.s.}} Y \quad \Rightarrow \quad \mathbb{P}(X = Y) = 1.$$

1.4.1 Key Relationship Between Convergence Modes

We established the following **implications between different types of convergence**:

Proposition (Implication of Convergence Modes)

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

That is, **mean square convergence implies convergence in probability**, but the reverse is not necessarily true.

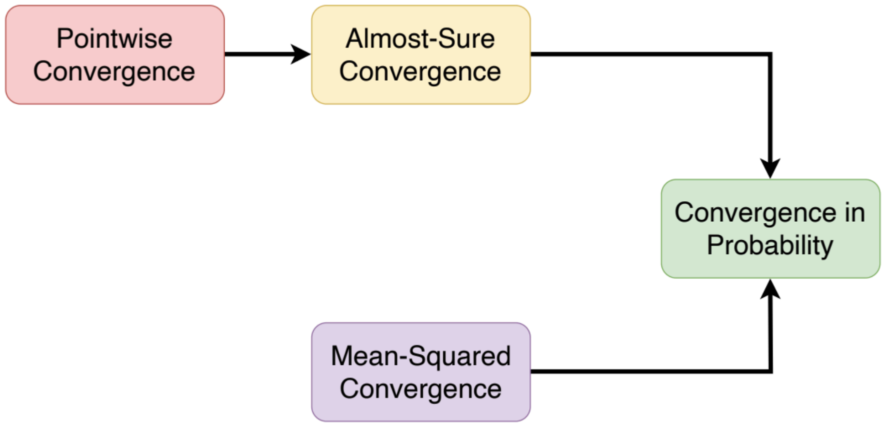

Figure 2: Figure to keep in mind till now

1.5 Example: Periodic Random Sequence

Towards the end of the previous lecture, we discussed an example involving a periodic sequence $\{X_n\}_{n=1}^{\infty}$, which takes values cyclically as:

$$X_n \in \left\{1, \frac{2}{3}, \frac{1}{3}\right\}.$$

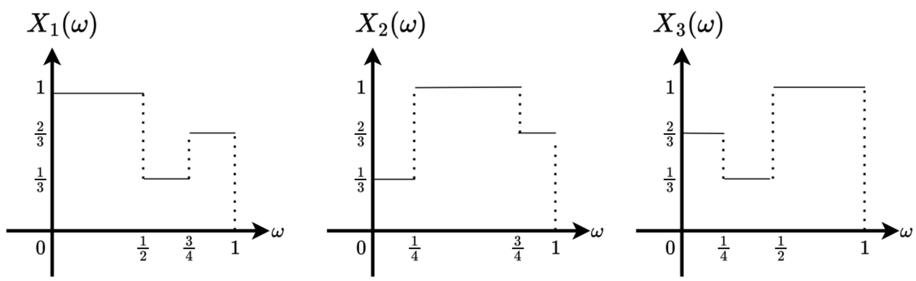

Figure 3: Let $X_n = X_{n+3}$ for all $n \in \mathbb{N}$.

The key observations were:

- The sequence does **not converge pointwise**.
- The sequence does **not converge almost surely** because no single limit exists for all ω.
- Each X_n has **the same PMF** for all n as:

$$\mathbb{P}(X_n = 1) = \frac{1}{2}$$
$$\mathbb{P}(X_n = \frac{1}{3}) = \frac{1}{4}$$
$$\mathbb{P}(X_n = \frac{2}{3}) = \frac{1}{4}$$

Despite lacking almost sure or pointwise convergence, the sequence does exhibit convergence of its cumulative distribution function (CDF). This leads us to the concept of **convergence in distribution**, which we will explore further in this lecture.

Informal Definition (Convergence in Distribution)

A sequence of random variables $\{X_n\}_{n=1}^{\infty}$ is said to **converge in distribution** to X if

$$F_{X_n}(x) \to F_X(x) \quad \text{for all } x \text{ that is a continuity point of } F_X.$$

This definition allows for convergence even in cases where almost sure convergence or convergence in probability fails.

2 Convergence of CDFs – A Subtle Point via an Example

In this section, we analyze a sequence of random variables and their cumulative distribution functions (CDFs). We consider a uniform random variable and examine its convergence properties.

2.1 Problem Setup

Let $U \sim$ Unif$[0, 1]$. For each $n \in \mathbb{N}$, define the sequence of random variables:

$$X_n = \frac{(-1)^n U}{n}.$$

Our goal is to:

- **Guess a limit random variable** for $\{X_n\}_{n=1}^{\infty}$.
- **Identify forms of convergence** to the above limit.
- **Analyze convergence of the sequence of CDFs** $\{F_{X_n}\}_{n=1}^{\infty}$.

2.2 Guessing the Limit Random Variable

Observing the behavior of X_n as $n \to \infty$, we see that:

$$\lim_{n\to\infty} X_n = 0 \quad \text{almost surely.}$$

This suggests that the limit random variable is:

$$X = 0.$$

as numerator is bounded between -1 and 1, and the denominator increases indefinitely.

2.3 Computation of the CDF of X_n

The cumulative distribution function (CDF) of X_n is given by:

$$F_{X_n}(x) = \mathbb{P}(X_n \le x) = \mathbb{P}\left(\frac{(-1)^n U}{n} \le x\right).$$

2.3.1 Case 1: n even

When n is even, $(-1)^n = 1$, so:

$$F_{X_n}(x) = \mathbb{P}\left(\frac{U}{n} \le x\right) = \mathbb{P}(U \le nx).$$

Since $U \sim$ Unif$[0, 1]$, this probability is:

$$F_{X_n}(x) = \begin{cases} 0, & x < 0, \\ nx, & 0 \le x < \frac{1}{n}, \\ 1, & x \ge \frac{1}{n}. \end{cases}$$

2.3.2 Case 2: n odd

When n is odd, $(-1)^n = -1$, so:

$$F_{X_n}(x) = \mathbb{P}\left(-\frac{U}{n} \le x\right) = \mathbb{P}\left(U \ge -nx\right).$$

Rewriting using complement probability,

$$F_{X_n}(x) = 1 - \mathbb{P}(U < -nx).$$

Thus, we obtain:

$$F_{X_n}(x) = \begin{cases} 0, & x < -\frac{1}{n}, \\ 1 + nx, & -\frac{1}{n} \le x < 0, \\ 1, & x \ge 0. \end{cases}$$

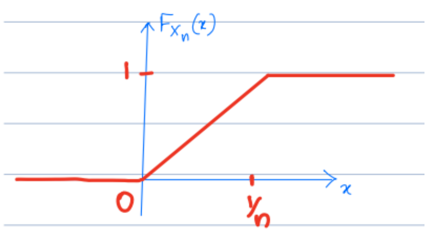

Figure 4: CDF for even case

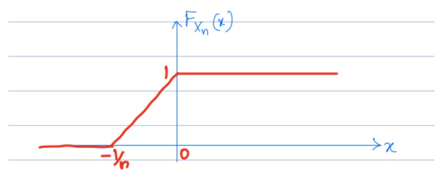

Figure 5: CDF for odd case

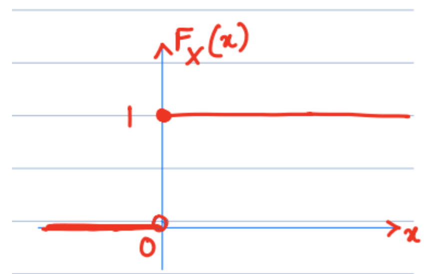

Figure 6: Limit CDF

2.4 Pointwise Limit of $F_{X_n}(x)$

Taking the pointwise limit as $n \to \infty$, we get:

$$\lim_{n\to\infty} F_{X_n}(x) = \begin{cases} 0, & x < 0, \\ 1, & x > 0. \end{cases}$$

At $x = 0$, the limit does not exist in a well-defined manner.
There is convergence of CDF at every x, except 0.

2.5 Conclusion: Convergence in Distribution

Since $F_{X_n}(x) \to F_X(x)$ at all continuity points of F_X, we conclude:

$$X_n \xrightarrow{d} X, \quad X = 0.$$

That is, X_n converges in distribution to $X = 0$, except at $x = 0$.
One of the key observations from our example is that the sequence of CDFs $\{F_{X_n}(x)\}_{n=1}^{\infty}$ does not converge pointwise everywhere due to oscillatory behavior. However, outside the discontinuities of the limiting function, a structured form of convergence can still be seen. This raises an important question—what does it mean for a sequence of random variables to converge if their CDFs exhibit jumps in the limit? Since uniform convergence of CDFs is often too strong to hold in many practical scenarios, probabilists devised a refined notion of convergence that accommodates such jumps while still capturing the limiting behavior effectively. This leads us to the concept of **convergence in distribution**, which we now define and study in detail.

3 Convergence in Distribution

In the previous section, we analyzed the convergence behavior of the cumulative distribution functions (CDFs) associated with a sequence of random variables $\{X_n\}_{n=1}^{\infty}$. We observed that as $n \to \infty$, the CDFs $F_{X_n}(x)$ exhibited jumps at certain points, creating discontinuities. This raised an important question: how should we define the convergence of a sequence of CDFs when the limiting function itself is discontinuous?
To tackle this issue, probabilists introduced a notion of convergence that carefully accounts for these jumps. Instead of requiring pointwise convergence of the CDFs everywhere, they formulated a weaker form of convergence that ensures the limiting behavior aligns with the properties of the final distribution.
Convergence in distribution is also termed as *weak convergence* or *convergence in law*. [GS20].

3.1 Definition of Convergence in Distribution

Definition (Convergence in Distribution)

A sequence of random variables $\{X_n\}_{n=1}^{\infty}$ is said to **converge in distribution** to a random variable X if

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x), \quad \forall x \text{ where } F_X(x) \text{ is continuous.}$$

That is, the sequence of their cumulative distribution functions $\{F_{X_n}(x)\}_{n=1}^{\infty}$ converges to the cumulative distribution function $F_X(x)$ of X at all points where $F_X(x)$ is continuous.
Notation:

$$X_n \xrightarrow{\text{d.}} X$$

This definition effectively bypasses the complications arising from jumps in the limiting CDF. Instead of requiring convergence at every point, it ensures that the convergence holds at all continuity points of $F_X(x)$. This is crucial because probability mass accumulates at discontinuities, making pointwise convergence unsuitable for handling such cases.

3.2 Examples and Applications

To illustrate convergence in distribution, consider the sequence of random variables

$$X_n = \frac{(-1)^n U}{n}, \quad \text{where } U \sim \text{Unif}[0,1].$$

From the previous discussion, we derived that the CDFs $F_{X_n}(x)$ behave differently for even and odd values of n. However, as $n \to \infty$, they converge to the CDF of the limiting random variable $X = 0$, which is:

$$F_X(x) = \begin{cases} 0, & x < 0, \\ 1, & x \geq 0. \end{cases}$$

Since this limiting function is discontinuous at $x = 0$, the definition of convergence in distribution requires that we ignore this point when evaluating the convergence.
Thus, in practical scenarios, convergence in distribution is particularly useful when dealing with limit theorems, such as the central limit theorem (CLT), where the weak convergence of properly scaled sums of random variables to a normal distribution plays a fundamental role.

3.3 Conclusion

Convergence in distribution provides a powerful and flexible way to describe the limiting behavior of random variables, even in cases where other modes of convergence fail. By focusing on the continuity points of the limiting CDF, this concept allows for meaningful analysis of random variables that exhibit jumps or accumulate probability mass at specific locations.

4 Implications Between Types of Convergence and Limit Theorems

We have introduced multiple notions of convergence: pointwise, almost-sure, in-probability, and in-distribution. Each of these forms of convergence provides a different level of insight into the limiting behavior of a sequence of random variables. A key question that arises is how these different notions of convergence relate to one another.

4.1 Implications Between Different Modes of Convergence

The relationships between different types of convergence can be summarized as follows:

$$A_{\lim} = \left\{ \omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega) \right\}$$

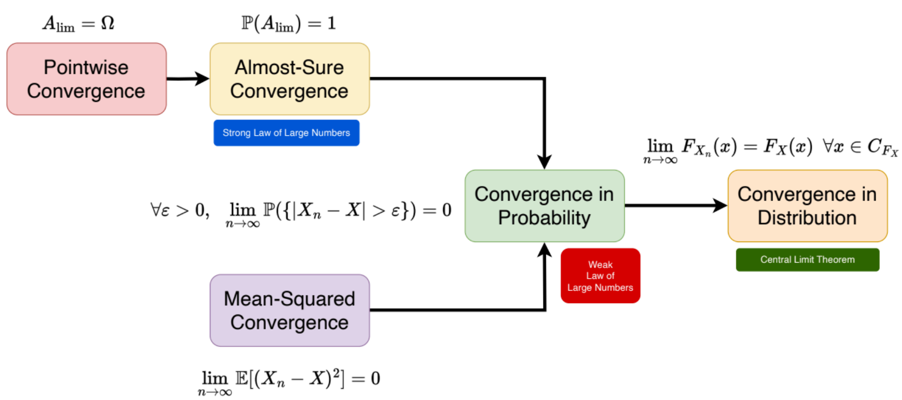

Figure 7: Convergence - Full Picture

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

Theorem

The following implications hold:

$$\left.\begin{array}{c}(X_n \xrightarrow{\text{a.s.}} X) \\ (X_n \xrightarrow{r} X)\end{array}\right. \Rightarrow (X_n \xrightarrow{p.} X) \Rightarrow (X_n \xrightarrow{d.} X)$$

for any $r \geq 1$. Also, if $r > s \geq 1$, then

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

No other implications hold **in general**.[GS20]

Here, convergence in r^{th} mean is defined as following:

Definition (Convergence in r^{th} mean)

Let $X, X_1, X_2, \dots$ be random variables on some probability space $(\Omega, \mathscr{F}, \mathbb{P})$.
We say $X_n \to X$ in r^{th} mean, where $r \geq 1$, written $X_n \xrightarrow{r} X$, if $\mathbb{E}[|X_n|^r] < \infty$ for all n and

$$\mathbb{E}(|X_n - X|^r) \to 0 \quad \text{as } n \to \infty.$$

Of most use are the values $r = 1$ and $r = 2$, in which cases we write respectively [GS20]

$$X_n \xrightarrow{1} X, \text{ or } X_n \to X \textit{ in mean}, \text{ or l.i.m. } X_n = X,$$

$$X_n \xrightarrow{2} X, \text{ or } X_n \to X \textit{ in mean square}, \text{ or } X_n \xrightarrow{\text{m.s.}} X.$$

These implications define a hierarchy, meaning that stronger forms of convergence automatically imply weaker forms. However, the reverse is generally not true:

- Convergence in distribution **does not** imply convergence in probability.
- Convergence in probability **does not** necessarily imply almost sure convergence.
- Convergence in probability **does not** necessarily imply mean squared convergence.

To conclude stronger results in the reverse direction, additional conditions such as boundedness or moment constraints are required.

4.2 Limit Theorems

- **Strong Law of Large Numbers (SLLN)** is a statement about **Almost-Sure Convergence**.
- **Weak Law of Large Numbers (WLLN)** is a statement about **Convergence in probability**.
- **Central Limit Theorem (CLT)** is a statement about **Convergence in Distribution**.

5 Review: Continuity of Probability Measure

Before proceeding to the proofs of convergence implications, we begin by formally stating a fundamental property of probability measures, known as the **continuity of probability**. This section serves as a brief overview of the concept.
Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a fixed probability space.

Definition (Probability of Limit on Increasing Sequence of Sets)

Let $\{A_n\}_{n\in\mathbb{N}}$ be a non-decreasing sequence of sets, i.e.,

$$A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots, \quad \text{with } A_n \in \mathscr{F} \text{ for each } n \in \mathbb{N}.$$

Then, $\lim_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n$, and

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

Definition (Probability of Limit on Decreasing Sequence of Sets)

Let $\{A_n\}_{n\in\mathbb{N}}$ be a non-increasing sequence of sets, i.e.,

$$A_1 \supseteq A_2 \supseteq A_3 \supseteq \cdots, \quad \text{with } A_n \in \mathscr{F} \text{ for each } n \in \mathbb{N}.$$

Then the limit $\lim_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} A_n$, and

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

That is, **the probability of the limit set is equal to the limit of the probabilities**. While we do not discuss the notions of lim inf and lim sup of sets here, we note that a limit of a sequence of sets exists when the lim sup and lim inf sets coincide.

6 Proof of Implications

6.1 Almost Sure Convergence Implies Convergence in Probability (a.s. $\Rightarrow$ p.)

Suppose we are given a sequence of random variables $X_1, X_2, \ldots$ that converges almost surely to a random variable X. Then,

$$X_n \xrightarrow{\text{a.s.}} X \Rightarrow \forall \varepsilon > 0,\ \mathbb{P}\left(\{|X_n - X| > \varepsilon\} \text{ i.o.}\right) = 0$$

(This follows from the Borel–Cantelli Lemma, discussed in Section 1.2.)

Let us define $A_n = \{|X_n - X| > \varepsilon\}$. Then, A_n i.o. means that for every $n \in \mathbb{N}$, there exists $k \geq n$ such that A_k occurs. This is equivalent to:

$$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k.$$

We can now rewrite the implication in terms of these sets:

$$\begin{aligned}
X_n \xrightarrow{\text{a.s.}} X &\Rightarrow \forall \varepsilon > 0,\ \mathbb{P}\left(\{|X_n - X| > \varepsilon\} \text{ i.o.}\right) = 0 \\
&\Rightarrow \forall \varepsilon > 0,\ \mathbb{P}\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} \{|X_k - X| > \varepsilon\}\right) = 0 \\
&\Rightarrow \forall \varepsilon > 0,\ \mathbb{P}\left(\bigcap_{n=1}^{\infty} B_n\right) = 0 \\
&\quad \text{(where } B_n = \textstyle\bigcup_{k=n}^{\infty} \{|X_k - X| > \varepsilon\}) \\
&\Rightarrow \forall \varepsilon > 0,\ \lim_{n\to\infty} \mathbb{P}(B_n) = 0 \\
&\quad \text{(This follows from the continuity of probability for decreasing or non-increasing sequences, since } B_1 \supseteq B_2 \supseteq B_3 \supseteq \cdots) \\
&\Rightarrow \forall \varepsilon > 0,\ \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{k=n}^{\infty} \{|X_k - X| > \varepsilon\}\right) = 0 \\
&\overset{(1)}{\Rightarrow} \forall \varepsilon > 0,\ \lim_{n\to\infty} \mathbb{P}\left(|X_k - X| > \varepsilon\right) = 0.
\end{aligned}$$

The final line is precisely the definition of convergence in probability. Hence,

$$X_n \xrightarrow{\text{a.s.}} X \Rightarrow X_n \xrightarrow{\text{p.}} X.$$

It is important to note that $B_n = \bigcup_{k=n}^{\infty} \{|X_k - X| > \varepsilon\}$. Since $\{|X_k - X| > \varepsilon\} \subseteq B_n$ for all $k \geq n$ it follows that,

$$0 \leq \mathbb{P}(|X_k - X| > \varepsilon) \leq \mathbb{P}(B_n), \quad \text{for all } k \geq n \text{ (since the probability of a set is always less than or equal to the probability of the union of sets that contain it)}$$

But $\mathbb{P}(B_n) = 0$, and hence $\mathbb{P}(|X_k - X| > \varepsilon) = 0$ for all $k \geq n$. In particular, this gives $\mathbb{P}(|X_n - X| > \varepsilon) = 0$, which is what was used in implication (1).

6.2 Mean Square Convergence Implies Convergence in Probability (m.s. $\Rightarrow$ p.)

$$\begin{aligned}
X_n \xrightarrow{\text{m.s.}} X &\Rightarrow \lim_{n\to\infty} \mathbb{E}\left[(X_n - X)^2\right] = 0 && \text{(By definition of convergence in mean square)} \\
&\Rightarrow \mathbb{P}(|X_n - X| > \varepsilon) = \mathbb{P}((X_n - X)^2 > \varepsilon^2) \leq \frac{\mathbb{E}[(X_n - X)^2]}{\varepsilon^2} && \text{(By Chebyshev's inequality)} \\
&\Rightarrow \lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) \leq \frac{1}{\varepsilon^2} \cdot \lim_{n\to\infty} \mathbb{E}[(X_n - X)^2] && \text{(Taking limits on both sides)} \\
&\Rightarrow \lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) \leq 0 && \text{(By definition of mean square convergence; RHS} \to 0) \\
&\Rightarrow \lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0 && \text{(Since probability is non-negative)}
\end{aligned}$$

$$\Rightarrow X_n \xrightarrow{\text{p.}} X \qquad \text{(By definition of convergence in probability)}$$

The convergence of X_n to X in mean square, i.e., $X_n \xrightarrow{\text{m.s.}} X$, implies that X_n also converges to X in probability, i.e., $X_n \xrightarrow{\text{p.}} X$.

Definition(Chebyshev's Inequality)

We used Chebyshev's inequality in the form

$$\mathbb{P}(|Y| \geq \varepsilon) \leq \frac{\mathbb{E}[Y^2]}{\varepsilon^2},$$

which holds for any random variable Y with $\mathbb{E}[Y^2] < \infty$. In our case, we applied it to the random variable $Y = X_n - X$. A detailed proof or discussion is omitted, as it is not central to our objective.

6.3 Convergence in Probability Implies Convergence in Distribution (p. $\Rightarrow$ d.)

We aim to prove convergence in distribution, i.e.,

$$X_n \xrightarrow{\text{d.}} X,$$

which means that for all $x \in C_{F_X}$, where C_{F_X} is the set of continuity points of the cumulative distribution function F_X, we have

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x).$$

Here, F_X denotes the CDF of the limiting random variable X, and C_{F_X} refers to the set of all points $x \in \mathbb{R}$ where F_X is continuous, i.e., where it is both left-continuous and right-continuous (which for CDFs, happens at every continuity point by definition).
We start by picking an arbitrary $x \in C_{F_X}$ and $\varepsilon > 0$.

$$\begin{aligned}
F_{X_n}(x) &= \mathbb{P}(X_n \leq x) && \text{(By definition of CDF)}\\
&= \mathbb{P}(X_n \leq x,\, X \leq x+\varepsilon) + \mathbb{P}(X_n \leq x,\, X > x+\varepsilon) && \text{(from point 1 below)}\\
&\leq \mathbb{P}(X \leq x+\varepsilon) + \mathbb{P}(X_n - X < -\varepsilon) && \text{(from points 2 and 3 below)}\\
&\leq F_X(x+\varepsilon) + \mathbb{P}(X_n - X < -\varepsilon) && \text{(By definition of CDF)}\\
&\leq F_X(x+\varepsilon) + \mathbb{P}(|X_n - X| > \varepsilon) && \text{(from point 3)}
\end{aligned}$$

For now, we assume the points mentioned to be true. We elaborate on them in a further section below.

Therefore,

$$\begin{aligned}
& F_{X_n}(x) \leq F_X(x+\varepsilon) + \mathbb{P}(|X_n - X| > \varepsilon)\\
\Rightarrow & \lim_{n\to\infty} F_{X_n}(x) \leq \lim_{n\to\infty} F_X(x+\varepsilon) + \lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) && \text{(Taking limits on both sides)}\\
\Rightarrow & \lim_{n\to\infty} F_{X_n}(x) \leq \lim_{n\to\infty} F_X(x+\varepsilon) + 0 && \text{(By definition of convergence in probability)}\\
\Rightarrow & \lim_{n\to\infty} F_{X_n}(x) \leq \lim_{n\to\infty} F_X(x+\varepsilon) && \text{(1)}
\end{aligned}$$

Similarly,

$$\begin{aligned}
F_X(x-\varepsilon) &= \mathbb{P}(X \leq x-\varepsilon) && \text{(By definition of CDF)}\\
&= \mathbb{P}(X \leq x-\varepsilon, X_n \leq x) + \mathbb{P}(X \leq x-\varepsilon, X_n > x) && \text{(from point 1)}\\
&\leq \mathbb{P}(X_n \leq x) + \mathbb{P}(X_n - X > \varepsilon) && \text{(from points 2 and 3)}\\
&\leq F_{X_n}(x) + \mathbb{P}(|X_n - X| > \varepsilon) && \text{(By definition of CDF)}\\
&\leq F_{X_n}(x) + \mathbb{P}(|X_n - X| > \varepsilon) && \text{(from point 3)}
\end{aligned}$$

Therefore,

$$F_X(x-\varepsilon) \leq F_{X_n}(x) + \mathbb{P}(|X_n - X| > \varepsilon)$$

$$
\begin{aligned}
&\Rightarrow \lim_{n\to\infty} F_X(x-\varepsilon) \leq \lim_{n\to\infty} \left(F_{X_n}(x) + \mathbb{P}(|X_n - X| > \varepsilon)\right) && \text{(Taking limits on both sides)}\\
&\Rightarrow \lim_{n\to\infty} F_X(x-\varepsilon) \leq \lim_{n\to\infty} F_{X_n}(x) + \lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon)\\
&\Rightarrow \lim_{n\to\infty} F_X(x-\varepsilon) \leq \lim_{n\to\infty} F_{X_n}(x) + 0 && \text{(By definition of convergence in probability)}\\
&\Rightarrow \lim_{n\to\infty} F_X(x-\varepsilon) \leq \lim_{n\to\infty} F_{X_n}(x) && (2)
\end{aligned}
$$

Combining equations (1) and (2)

$$
\begin{aligned}
&\lim_{n\to\infty} F_X(x-\varepsilon) \leq \lim_{n\to\infty} F_{X_n}(x) \leq \lim_{n\to\infty} F_X(x+\varepsilon)\\
\Rightarrow &\lim_{\varepsilon\downarrow 0}\lim_{n\to\infty} F_X(x-\varepsilon) \leq \lim_{\varepsilon\downarrow 0}\lim_{n\to\infty} F_{X_n}(x) \leq \lim_{\varepsilon\downarrow 0}\lim_{n\to\infty} F_X(x+\varepsilon) && \text{(applying limits as } \varepsilon \downarrow 0)\\
\Rightarrow &\lim_{n\to\infty} F_X(x) \leq \lim_{n\to\infty} F_{X_n}(x) \leq \lim_{n\to\infty} F_X(x) && \text{(the middle term does not have } \varepsilon\text{, hence it will not be affected)}\\
\Rightarrow &F_X(x) \leq \lim_{n\to\infty} F_{X_n}(x) \leq F_X(x). && \text{(The leftmost and rightmost terms do not contain } n)
\end{aligned}
$$

Thus, we have: $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$. This is the definition of convergence in distribution. Therefore, we have proved that

$$
X_n \xrightarrow{\text{p.}} X \quad \Rightarrow \quad F_{X_n}(x) \to F_X(x), \quad \forall x \in C_{F_X}.
$$

Explanation for steps used in proof of implication

The following points were used in intermediate steps to arrive at the convergence.
Assume that A and B are two events.

1. We have

$$
\begin{aligned}
\mathbb{P}(A) &= \mathbb{P}(A, B \cup B^c) = \mathbb{P}(A, B) + \mathbb{P}(A, B^c)\\
\Rightarrow \mathbb{P}(X_n \leq x) &= \mathbb{P}(X_n \leq x,\, \{X \leq x+\varepsilon\} \cup \{X > x+\varepsilon\})\\
&= \mathbb{P}(X_n \leq x,\, X \leq x+\varepsilon) + \mathbb{P}(X_n \leq x,\, X > x+\varepsilon) \quad \text{(Splitting by } X)
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
\mathbb{P}(X \leq x-\varepsilon) &= \mathbb{P}(X \leq x-\varepsilon,\, \{X_n \leq x\} \cup \{X_n > x\})\\
&= \mathbb{P}(X \leq x-\varepsilon,\, X_n \leq x) + \mathbb{P}(X_n \leq x-\varepsilon,\, X > x) \quad \text{(Splitting by } X_n)
\end{aligned}
$$

2. We have

$$
\begin{aligned}
\mathbb{P}(A, B) = \mathbb{P}(A \cap B) \leq \mathbb{P}(A) \quad &\text{and} \quad \mathbb{P}(A, B) \leq \mathbb{P}(B)\\
\Rightarrow \mathbb{P}(X_n \leq x,\, X \leq x+\varepsilon) \leq \mathbb{P}(X \leq x+\varepsilon) \quad &\text{and}\\
\mathbb{P}(X \leq x-\varepsilon,\, X_n \leq x) \leq \mathbb{P}(X_n \leq x)&
\end{aligned}
$$

3. If A implies B, then $A \subseteq B$, and thus $\mathbb{P}(A) \leq \mathbb{P}(B)$. This implies the following for equation (1):

$$
\{X_n \leq x, X > x+\varepsilon\} = \{X_n \leq x, -X < -x-\varepsilon\} \subseteq \{X_n - X < -\varepsilon\}
$$

$$
\begin{aligned}
\Rightarrow \mathbb{P}(X_n \leq x, X > x+\varepsilon) \leq \mathbb{P}(X_n - X < -\varepsilon) \quad \text{and}\\
\{X_n - X < -\varepsilon\} \subseteq \{|X_n - X| > \varepsilon\} \Rightarrow \mathbb{P}(X_n - X < -\varepsilon) \leq \mathbb{P}(|X_n - X| > \varepsilon)
\end{aligned}
$$

Similarly, for equation (2):

$$
\{X \leq x-\varepsilon, X_n > x\} = \{-X \geq -x+\varepsilon, X_n > x\} \subseteq \{X_n - X > \varepsilon\}
$$

$$
\begin{aligned}
\Rightarrow \mathbb{P}(X \leq x-\varepsilon, X_n \geq x) \leq \mathbb{P}(X_n - X > \varepsilon) \quad \text{and}\\
\{X_n - X > \varepsilon\} \subseteq \{|X_n - X| > \varepsilon\} \Rightarrow \mathbb{P}(X_n - X > \varepsilon) \leq \mathbb{P}(|X_n - X| > \varepsilon)
\end{aligned}
$$

References

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