Probability Theory — Lectures 1 & 2

Haruka Eshima

September 2026

Abstract

This is a lecture note that I took for the Information Theory class taught by Prof. Amedeo Roberto Esposito at the Okinawa Institute of Science and Technology, 2026. Any errors are my own.

The questions in green are made by me and are not officially related to the mini test in the class.

Contents

1 Probability spaces and σ-algebras

A probability space is a triple $(\Omega, \mathcal{F}, \mathbb{P})$, consisting of a measurable space $(\Omega, \mathcal{F})$ together with a probability measure $\mathbb{P} : \mathcal{F} \to [0, 1]$. Here Ω is the set of all possible outcomes; e.g. $\Omega = [0, 1]$ or $\Omega = \mathbb{R}$.

Q. *State the definition of a σ-algebra.*

Definition 1.1 (σ-algebra)**.** A collection $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ of subsets of Ω is a *σ-algebra* if

(i) $\varnothing,\ \Omega \in \mathcal{F}$;

(ii) $(A_n)_{n\geq 1} \subseteq \mathcal{F} \implies \displaystyle\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$;

(iii) $A \in \mathcal{F} \implies A^c \in \mathcal{F}$.

Immediate consequences:

$$\begin{array}{rl} \text{(ii}'\text{)} & A_1, \dots, A_n \in \mathcal{F} \implies \bigcup_{i=1}^n A_i \in \mathcal{F}, \\ \text{(iv)} & (A_n)_{n\geq 1} \subseteq \mathcal{F} \implies \bigcap_{n=1}^{\infty} A_n \in \mathcal{F}, \\ \text{(iv}'\text{)} & A_1, \dots, A_n \in \mathcal{F} \implies \bigcap_{i=1}^n A_i \in \mathcal{F}, \\ \text{(v)} & A, B \in \mathcal{F} \implies B \setminus A = B \cap A^c \in \mathcal{F}. \end{array}$$

The countable intersections follow from (ii) and (iii) by De Morgan, $\bigcap_n A_n = \big(\bigcup_n A_n^c\big)^c$; the finite statements follow by padding with $A_n = \varnothing$ (resp. $A_n = \Omega$).

▶ **Remark.** The finite union axiom (ii) and the finite intersection (iv′) are consequences. In (i) it is enough to require $\Omega \in \mathcal{F}$ (or merely $\mathcal{F} \neq \varnothing$), since $\varnothing = \Omega^c$ then follows from (iii).

1.1 Atoms and generated σ-algebras

Q. *State the definition of an atom of a σ-algebra.*

Definition 1.2 (Atom)**.** A set $F \in \mathcal{F}$ with $F \neq \varnothing$ is an *atom* of $\mathcal{F}$ if for every $G \in \mathcal{F}$ with $G \subseteq F$ we have $G = F$ or $G = \varnothing$.

Definition 1.3 (Generated σ-algebra)**.** Given a collection of subsets of Ω, $\mathcal{A} := \{A_i : i \in I\}$, the *$\sigma$-algebra generated by $\mathcal{A}$*, written $\sigma(\mathcal{A})$, is the smallest σ-algebra containing $\mathcal{A}$:

$$\sigma(\mathcal{A}) \;=\; \bigcap \big\{\, \mathcal{G} : \mathcal{G} \text{ is a } \sigma\text{-algebra on } \Omega,\ \mathcal{A} \subseteq \mathcal{G} \,\big\}.$$

Q. *Why is the existence of the σ-algebra generated by $\mathcal{A}$ guaranteed?*

Two facts together give existence:

(a) The family being intersected is non-empty, because $\mathcal{P}(\Omega)$ is a σ-algebra containing $\mathcal{A}$.

(b) An arbitrary intersection of σ-algebras on Ω is again a σ-algebra. (Each axiom is checked "pointwise" across the family: if A_n lies in every $\mathcal{G}$ in the family, so does $\bigcup_n A_n$, etc.)

Hence the intersection above is a σ-algebra, it contains $\mathcal{A}$, and by construction it is contained in every σ-algebra that contains $\mathcal{A}$.

Example 1.4. On $\Omega = \{1, 2, \dots, 6\}$,

$$\mathcal{F}_1 \;=\; \sigma\big(\{\{1\}\}\big) \;=\; \big\{\varnothing,\ \Omega,\ \{1\},\ \{2,3,4,5,6\}\big\}.$$

1.2 The Borel σ-algebra

Q. *State the definition of a Borel set.*

Definition 1.5 (Borel σ-algebra)**.** On $[0, 1]$,

$$\mathcal{B}([0,1]) \;=\; \sigma\Big(\big\{\{0\},\ \{1\}\big\} \;\cup\; \big\{\,]a,b[\ :\ 0 < a < b < 1 \,\big\}\Big).$$

On $\mathbb{R}$,

$$\mathcal{B}(\mathbb{R}) \;=\; \sigma\big(\{\,]a,b[\ :\ a < b \,\}\big) \;=\; \sigma\big(\{\,]-\infty, t[\ :\ t \in \mathbb{R} \,\}\big).$$

Elements of $\mathcal{B}(\cdot)$ are called *Borel sets*.

Two sanity checks on $[0, 1]$:

- $\{x\} \in \mathcal{B}([0,1])$ for every x: for $x \in]0,1[$ write $\{x\} = \bigcap_{n \geq n_0}]x - \frac{1}{n}, x + \frac{1}{n}[$, a countable intersection; and $\{0\}, \{1\}$ are generators.
- $[a, b] \in \mathcal{B}([0,1])$: take $[a,b] = \{a\} \cup]a,b[\cup \{b\}$.

Q. *Is $\mathcal{P}([0,1]) = \mathcal{B}([0,1])$? State a counterexample.*

No: $\mathcal{B}([0,1]) \subsetneq \mathcal{P}([0,1])$. A Vitali set V is the standard counterexample.

Q. *Construct a Vitali set V and prove that V is not Lebesgue measurable.*

Example 1.6 (Vitali set)**.** Define $x \sim y$ on $[0, 1]$ iff $x - y \in \mathbb{Q}$. This is an equivalence relation; by the axiom of choice pick one representative from each class to form $V \subseteq [0, 1]$. Let $(q_k)_{k \geq 1}$ enumerate $\mathbb{Q} \cap [-1, 1]$ and let $V_k := V + q_k$. Then the V_k are pairwise disjoint and

$$[0,1] \;\subseteq\; \bigcup_{k \geq 1} V_k \;\subseteq\; [-1, 2].$$

If V were Lebesgue measurable, then by translation invariance and countable additivity $\lambda\big(\bigcup_k V_k\big) = \sum_k \lambda(V)$, which is 0 if $\lambda(V) = 0$ and $+\infty$ if $\lambda(V) > 0$. Neither is compatible with $1 \leq \lambda\big(\bigcup_k V_k\big) \leq 3$. Hence V is not Lebesgue measurable.

▶ **Remark.** Strictly, the Vitali construction shows that V is not *Lebesgue* measurable, i.e. $\mathcal{L} \subsetneq \mathcal{P}([0,1])$, where $\mathcal{L}$ is the Lebesgue σ-algebra. Since $\mathcal{B}([0,1]) \subseteq \mathcal{L}$, the conclusion $\mathcal{B}([0,1]) \neq \mathcal{P}([0,1])$ follows *a fortiori*. (A separate cardinality argument gives $|\mathcal{B}([0,1])| = \mathfrak{c} < 2^{\mathfrak{c}} = |\mathcal{P}([0,1])|$, and a separate construction is needed to produce a Lebesgue-measurable set that is not Borel.)

2 Random variables and measurability

Q. *State the definition of a random variable.*

Definition 2.1 (Random variable)**.** A *random variable* on $(\Omega, \mathcal{F})$ is a map $X : \Omega \to \mathbb{R}$ such that

$$\forall B \in \mathcal{B}(\mathbb{R}) : \qquad X^{-1}(B) \ = \ \{\omega \in \Omega : X(\omega) \in B\} \ \in \ \mathcal{F}.$$

Such an X is called $\mathcal{F}$*-measurable.*

Q. *Prove that if $\mathcal{F} = \mathcal{P}(\Omega)$, then any function is measurable.*

Proposition 2.2. *If $\mathcal{F} = \mathcal{P}(\Omega)$, every $X : \Omega \to \mathbb{R}$ is $\mathcal{F}$-measurable.*

Proof. For any $B \in \mathcal{B}(\mathbb{R})$, $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}$ is a subset of Ω, hence lies in $\mathcal{P}(\Omega) = \mathcal{F}$. □

Q. *Prove that if $\mathcal{F} = \{\varnothing, \Omega\}$, only constant functions are measurable.*

Proposition 2.3. *If $\mathcal{F} = \{\varnothing, \Omega\}$, then X is $\mathcal{F}$-measurable iff X is constant.*

Proof. Constants are clearly measurable. Conversely, suppose $X(\omega_1) \neq X(\omega_2)$ for some $\omega_1, \omega_2 \in \Omega$. Choose $B \in \mathcal{B}(\mathbb{R})$ with $X(\omega_1) \in B$ and $X(\omega_2) \notin B$, e.g. $B = \{X(\omega_1)\}$. Then $\omega_1 \in X^{-1}(B)$ and $\omega_2 \notin X^{-1}(B)$, so $X^{-1}(B) \neq \varnothing$ and $X^{-1}(B) \neq \Omega$. Hence $X^{-1}(B) \notin \mathcal{F}$ and X is not measurable. □

Example 2.4. Let $\Omega = \{1, 2, \ldots, 6\}$ and set

$$\mathcal{F}_0 = \{\varnothing, \Omega\}, \qquad \mathcal{F}_1 = \sigma\big(\{1\}\big), \qquad \mathcal{F}_2 = \sigma\big(\{1,3,5\}\big) = \big\{\varnothing, \Omega, \{1,3,5\}, \{2,4,6\}\big\}.$$

Then $X_1(\omega) = \omega$ is measurable with respect to neither $\mathcal{F}_1$ nor $\mathcal{F}_2$ (e.g. $X_1^{-1}(\{2\}) = \{2\}$ lies in neither), whereas $X_2 = \mathbf{1}_{\{1,3,5\}}$ is $\mathcal{F}_2$-measurable.

2.1 Borel measurable functions

Q. *State the definition of a Borel measurable function.*

Definition 2.5. A function $g : \mathbb{R} \to \mathbb{R}$ is *Borel measurable* if $g^{-1}(B) \in \mathcal{B}(\mathbb{R})$ for every $B \in \mathcal{B}(\mathbb{R})$.

Proposition 2.6. *If $g : \mathbb{R} \to \mathbb{R}$ is continuous, then g is Borel measurable.*

Proof sketch. For open $U \subseteq \mathbb{R}$, $g^{-1}(U)$ is open, hence Borel. The collection $\{B \subseteq \mathbb{R} : g^{-1}(B) \in \mathcal{B}(\mathbb{R})\}$ is a σ-algebra (preimages commute with complements and countable unions) and contains all open sets, hence contains $\mathcal{B}(\mathbb{R})$. □

Proposition 2.7. *If X is $\mathcal{F}$-measurable and g is Borel measurable, then $Y = g(X)$ is $\mathcal{F}$-measurable.*

Proof. For $B \in \mathcal{B}(\mathbb{R})$,

$$\{\omega : Y(\omega) \in B\} = \{\omega : g(X(\omega)) \in B\} = \{\omega : X(\omega) \in g^{-1}(B)\} = X^{-1}\big(g^{-1}(B)\big) \in \mathcal{F},$$

since $g^{-1}(B) \in \mathcal{B}(\mathbb{R})$ and X is $\mathcal{F}$-measurable. □

2.2 σ-algebra generated by random variables

Q. *State the definition of the σ-algebra generated by a collection of random variables.*

Definition 2.8. Given a collection of random variables $\mathcal{X} = \{X_i : i \in I\}$, the *$\sigma$-algebra generated by $\mathcal{X}$* is the smallest σ-algebra on Ω making every X_i measurable:

$$\sigma(X_i,\ i \in I) \;=\; \sigma\Big(\big\{\{\omega : X_i(\omega) \in B\} \;:\; B \in \mathcal{B}(\mathbb{R}),\ i \in I\big\}\Big).$$

Q. *Prove that $X_0(\omega) \equiv c$ constant $\implies \sigma(X_0) = \{\varnothing, \Omega\}$.*

Proof. $\{\varnothing, \Omega\} \subseteq \sigma(X_0)$ holds for any σ-algebra. For the reverse inclusion, for every $B \in \mathcal{B}(\mathbb{R})$,

$$X_0^{-1}(B) = \begin{cases} \Omega, & c \in B, \\ \varnothing, & c \notin B. \end{cases}$$

So every generator lies in $\{\varnothing, \Omega\}$, which is itself a σ-algebra; hence $\sigma(X_0) \subseteq \{\varnothing, \Omega\}$. $\square$

Q. *Prove that on $\Omega = \{1, \ldots, 6\}$ with $X(\omega) = \omega$, $\sigma(X) = \mathcal{P}(\Omega)$.*

Proof. Let $A \subseteq \Omega$ be arbitrary. As a finite subset of $\mathbb{R}$, $A \in \mathcal{B}(\mathbb{R})$, and $X^{-1}(A) = A$. Hence $A \in \sigma(X)$. Since A was arbitrary, $\mathcal{P}(\Omega) \subseteq \sigma(X)$; the reverse inclusion is trivial. $\square$

Example 2.9. $\sigma\big(\mathbf{1}_{\{1,3,5\}}\big) = \mathcal{F}_2 = \{\varnothing, \Omega, \{1,3,5\}, \{2,4,6\}\}$.

Proof. Let $A = \{1,3,5\}$. For every $B \in \mathcal{B}(\mathbb{R})$,

$$\mathbf{1}_A^{-1}(B) = \begin{cases} \varnothing, & 0, 1 \notin B, \\ A, & 1 \in B,\ 0 \notin B, \\ A^c, & 0 \in B,\ 1 \notin B, \\ \Omega, & 0, 1 \in B. \end{cases}$$

Thus every generator of $\sigma(\mathbf{1}_A)$ belongs to $\{\varnothing, \Omega, A, A^c\}$, which is a σ-algebra. Hence

$$\sigma(\mathbf{1}_A) \subseteq \{\varnothing, \Omega, A, A^c\}.$$

Conversely, since $\{0\}, \{1\} \in \mathcal{B}(\mathbb{R})$,

$$A = \mathbf{1}_A^{-1}(\{1\}) \quad \text{and} \quad A^c = \mathbf{1}_A^{-1}(\{0\})$$

belong to $\sigma(\mathbf{1}_A)$. Therefore

$$\sigma\big(\mathbf{1}_{\{1,3,5\}}\big) = \{\varnothing, \Omega, \{1,3,5\}, \{2,4,6\}\} = \mathcal{F}_2.$$

$\square$

3 Probability measures

Q. *State the definition of a probability measure and of a measurable space.*

Definition 3.1 (Measurable space, probability measure)**.** A *measurable space* is a pair $(\Omega, \mathcal{F})$ with $\mathcal{F}$ a σ-algebra on Ω. Given $(\Omega, \mathcal{F})$, a *probability measure* is a map $\mathbb{P} : \mathcal{F} \to [0,1]$ such that

(1) $\mathbb{P}(\Omega) = 1$;

(2) (σ-additivity) if $(A_n)_{n\geq 1} \subseteq \mathcal{F}$ are pairwise disjoint, then $\mathbb{P}\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)$.

▸ **Remark.** Taking $A_n = \varnothing$ for all n in (2) gives $\mathbb{P}(\varnothing) = \sum_{n\geq 1} \mathbb{P}(\varnothing)$, which forces $\mathbb{P}(\varnothing) = 0$ since $\mathbb{P}(\varnothing) \leq 1 < \infty$; finite additivity then follows by padding a finite disjoint family with copies of $\varnothing$.

Q. *State the definition of continuity from below, and prove it.*

Proposition 3.2 (Continuity from below)**.** *If $(A_n)_{n\geq 1} \subseteq \mathcal{F}$ satisfies $A_n \subseteq A_{n+1}$ for all n, then*

$$\mathbb{P}\Big(\bigcup_{n=1}^{\infty} A_n\Big) \;=\; \lim_{n\to\infty} \mathbb{P}(A_n).$$

Proof. Set $A_0 := \varnothing$ and $B_n := A_n \setminus A_{n-1}$ for $n \geq 1$.

Step 1: the B_n are pairwise disjoint. Let $m < \ell$. Then $m \leq \ell - 1$, so $A_m \subseteq A_{\ell-1}$ by monotonicity of the sequence. If $x \in B_m = A_m \setminus A_{m-1}$, then $x \in A_m \subseteq A_{\ell-1}$, so $x \notin A_\ell \setminus A_{\ell-1} = B_\ell$. Hence $B_m \cap B_\ell = \varnothing$.

Step 2: $\bigcup_{n=1}^{N} B_n = A_N$ and $\bigcup_{n\geq 1} B_n = \bigcup_{n\geq 1} A_n$. The first identity is immediate by induction from $A_{N-1} \cup (A_N \setminus A_{N-1}) = A_N$ (using $A_{N-1} \subseteq A_N$); the second follows by letting $N \to \infty$.

Step 3: telescoping. Since $A_{n-1} \subseteq A_n$ we have $\mathbb{P}(B_n) = \mathbb{P}(A_n) - \mathbb{P}(A_{n-1})$, so by σ-additivity

$$\mathbb{P}\Big(\bigcup_{n\geq 1} A_n\Big) = \sum_{n=1}^{\infty} \mathbb{P}(B_n) = \lim_{N\to\infty} \sum_{n=1}^{N} \big[\mathbb{P}(A_n) - \mathbb{P}(A_{n-1})\big] = \lim_{N\to\infty} \big[\mathbb{P}(A_N) - \mathbb{P}(A_0)\big] = \lim_{N\to\infty} \mathbb{P}(A_N). \quad \square$$

4 Distributions and cumulative distribution functions

Q. *State the definition of the distribution of a random variable.*

Definition 4.1 (Distribution / pushforward measure)**.** The *distribution* of a random variable X on $(\Omega, \mathcal{F}, \mathbb{P})$ is the probability measure μ_X on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ given by

$$\mu_X(B) \;=\; \mathbb{P}\big(\{\omega : X(\omega) \in B\}\big) \;=\; \mathbb{P}\big(X^{-1}(B)\big), \qquad B \in \mathcal{B}(\mathbb{R}).$$

This is well defined precisely because $X^{-1}(B) \in \mathcal{F}$ for every Borel B.

Q. *State the definition of a CDF in two ways.*

Definition 4.2 (CDF)**.** Given a random variable X on $(\Omega, \mathcal{F}, \mathbb{P})$, the *cumulative distribution function* of X is $F_X : \mathbb{R} \to [0,1]$,

$$F_X(t) \;=\; \mathbb{P}\big(\{X \le t\}\big) \;=\; \mu_X\big(]-\infty, t]\big).$$

Proposition 4.3 (Properties of a CDF)**.** *F_X is non-decreasing, right-continuous, and satisfies*

$$\lim_{t\to-\infty} F_X(t) = 0, \qquad \lim_{t\to+\infty} F_X(t) = 1, \qquad \lim_{\varepsilon\downarrow 0} F_X(t+\varepsilon) = F_X(t).$$

Remark 4.4. Right-continuity is a consequence of continuity from above (the dual of Proposition 3.2) applied to $A_n = \{X \le t + \frac{1}{n}\}$, whose intersection is $\{X \le t\}$. It is the reason the CDF is defined with $\le$ rather than $<$.

5 Discrete, continuous, and mixed random variables

Q. *State the definition of a discrete random variable.*

Definition 5.1 (Discrete random variable)**.** X is *discrete* if there is a finite or countable set $D \subseteq \mathbb{R}$ with $\mathbb{P}(X \in D) = 1$.

For a discrete X there is a probability mass function

$$p_x = \mathbb{P}(\{X = x\}), \qquad 0 \le p_x \le 1, \qquad \sum_{x\in D} p_x = 1, \qquad F_X(t) = \sum_{x \le t} p_x.$$

The CDF is a step function, jumping by p_{x_i} at each atom x_i.

Example 5.2 (Bernoulli)**.** $X \sim \mathrm{Bern}(p)$ means

$$X = \begin{cases} 1 & \text{w.p. } p, \\ 0 & \text{w.p. } 1-p. \end{cases}$$

Q. *State the definition of a continuous random variable.*

Definition 5.3 (Continuous random variable)**.** X is *(absolutely) continuous* if

$$\mathbb{P}(X \in B) = 0 \quad \text{whenever} \quad \lambda(B) = 0,$$

where λ is Lebesgue measure.

By the Radon–Nikodym theorem this is equivalent to the existence of a density $p_X : \mathbb{R} \to [0,\infty)$ with

$$\int_{\mathbb{R}} p_X(x)\,dx = 1, \qquad F_X(t) = \int_{-\infty}^{t} p_X(x)\,dx.$$

Usually $F_X'(t) = p_X(t)$, but not always — only at Lebesgue-almost every t, and only where F_X is differentiable.

Example 5.4. $p_X(t) = \mathbf{1}_{[0,1]}(t)$ is the uniform density on $[0,1]$; F_X is not differentiable at $t = 0$ or $t = 1$, so $F_X' = p_X$ fails at those two points.

5.1 A random variable that is neither discrete nor continuous

Q. *Construct a random variable that is neither continuous nor discrete.*

Example 5.5. Let $X \sim \text{Bern}(p)$ with $0 < p < 1$, let $Y \sim N(0,1)$, let $X \perp\!\!\!\perp Y$, and set $Z = X \cdot Y$.

Z is not continuous. We have $\lambda(\{0\}) = 0$, but

$$\mathbb{P}\big(Z \in \{0\}\big) = \mathbb{P}\big(\{X=0\} \cup (\{X=1\} \cap \{Y=0\})\big) \;\geq\; \mathbb{P}(X=0) \;=\; 1-p \;>\; 0.$$

(In fact $\mathbb{P}(Z=0) = 1-p$ exactly, since $\mathbb{P}(Y=0)=0$.)

Z is not discrete. Let $D \subseteq \mathbb{R}$ be any countable set. Then

$$\mathbb{P}(Z \in D) \;\leq\; \mathbb{P}(X=0) + \mathbb{P}(X=1,\; Y \in D) \;=\; (1-p) + p \cdot 0 \;=\; 1-p \;<\; 1,$$

using $\mathbb{P}(Y \in D) = 0$ for countable D because Y is continuous. So no countable set carries full mass.

Hence Z has neither a pmf nor a pdf; only a CDF exists.

Q. *Derive the CDF of Z.*

Derivation. Write Φ for the standard normal CDF.

For $t \geq 0$: the event $\{Z \leq t\}$ contains all of $\{X=0\}$ (since then $Z = 0 \leq t$), and otherwise requires $X=1$ and $Y \leq t$:

$$\begin{aligned}
\mathbb{P}(Z \leq t) &= \mathbb{P}\big(\{X=0\} \;\cup\; \{X=1,\; Y \leq t\}\big) \\
&= (1-p) + \mathbb{P}\big(\{X=1\} \cap \{Y \leq t\}\big) && \text{(disjoint; additivity)} \\
&= (1-p) + \mathbb{P}(X=1)\,\mathbb{P}(Y \leq t) && (X \perp\!\!\!\perp Y) \\
&= (1-p) + p\,\Phi(t).
\end{aligned}$$

For $t < 0$: the event $\{X=0\}$ contributes nothing, so

$$\mathbb{P}(Z \leq t) = \mathbb{P}\big(\{X=1,\; Y \leq t\}\big) = \mathbb{P}(X=1)\,\mathbb{P}(Y \leq t) = p\,\Phi(t).$$

□

So

$$F_Z(t) = p\,\Phi(t) + (1-p)\,\mathbf{1}_{\{t \geq 0\}},$$

a continuous curve with a single jump at $t=0$ of height $1-p$:

$$F_Z(0^-) = \tfrac{p}{2}, \qquad F_Z(0) = 1 - \tfrac{p}{2}.$$

6 Independence

Q. *State the condition for two events A_1 and A_2 to be independent.*

Definition 6.1. $A_1 \perp\!\!\!\perp A_2$ iff $\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2)$.

Q. *Prove $A_1 \perp\!\!\!\perp A_2 \implies A_1^c \perp\!\!\!\perp A_2$.*

Proof.

$$\mathbb{P}(A_1^c \cap A_2) = \mathbb{P}(A_2 \setminus A_1) = \mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_2) - \mathbb{P}(A_1)\mathbb{P}(A_2) = \mathbb{P}(A_1^c)\mathbb{P}(A_2). \qquad \square$$

Q. *Prove $A_1 \perp\!\!\!\perp A_2 \implies A_1^c \perp\!\!\!\perp A_2^c$.*

Proof.

$$\begin{aligned}\mathbb{P}(A_1^c \cap A_2^c) &= \mathbb{P}\big((A_1 \cup A_2)^c\big) = 1 - \mathbb{P}(A_1 \cup A_2) \\ &= 1 - \mathbb{P}(A_1) - \mathbb{P}(A_2) + \mathbb{P}(A_1 \cap A_2) = 1 - \mathbb{P}(A_1) - \mathbb{P}(A_2) + \mathbb{P}(A_1)\mathbb{P}(A_2) \\ &= \big(1 - \mathbb{P}(A_1)\big)\big(1 - \mathbb{P}(A_2)\big) = \mathbb{P}(A_1^c)\mathbb{P}(A_2^c).\end{aligned} \qquad \square$$

Q. *State the condition for two random variables to be independent.*

Definition 6.2 (Independent random variables)**.** $X_1 \perp\!\!\!\perp X_2$ iff

$$\forall\, t_1, t_2 \in \mathbb{R}: \qquad \mathbb{P}(X_1 \le t_1,\ X_2 \le t_2) = \mathbb{P}(X_1 \le t_1)\,\mathbb{P}(X_2 \le t_2).$$

Proposition 6.3 (Equivalent form)**.** *Definition 6.2 is equivalent to*

$$\forall\, B_1, B_2 \in \mathcal{B}(\mathbb{R}): \qquad \mathbb{P}(X_1 \in B_1,\ X_2 \in B_2) = \mathbb{P}(X_1 \in B_1)\,\mathbb{P}(X_2 \in B_2).$$

Proof sketch. "$\Leftarrow$" is immediate with $B_i =]-\infty, t_i]$. "$\Rightarrow$" follows from Dynkin's π–λ theorem: the half-lines $\{\,]-\infty, t]\,\}$ form a π-system generating $\mathcal{B}(\mathbb{R})$, and the class of B's for which the product rule holds (with the other coordinate fixed) is a λ-system. $\square$

Q. *X, Y are independent random variables. State a sufficient condition on functions applied to X, Y that preserves independence.*

Proposition 6.4. *Let $f_1, f_2 : \mathbb{R} \to \mathbb{R}$ be Borel measurable and let $X \perp\!\!\!\perp Y$. Then $f_1(X) \perp\!\!\!\perp f_2(Y)$.*

Proof. By Proposition 2.7, $f_1(X)$ and $f_2(Y)$ are random variables. For $t_1, t_2 \in \mathbb{R}$ put $B_i = f_i^{-1}\big(]-\infty, t_i]\big)$, which are Borel sets since f_i is Borel measurable. Then

$$\begin{aligned}\mathbb{P}\big(f_1(X) \le t_1,\ f_2(Y) \le t_2\big) &= \mathbb{P}\big(X \in B_1,\ Y \in B_2\big) \\ &= \mathbb{P}(X \in B_1)\,\mathbb{P}(Y \in B_2) && \text{(Prop. 6.3)} \\ &= \mathbb{P}\big(f_1(X) \le t_1\big)\,\mathbb{P}\big(f_2(Y) \le t_2\big).\end{aligned} \qquad \square$$

▶ **Remark.** For independence to be meaningful, X and Y must be defined on the *same* probability space $(\Omega, \mathcal{F}, \mathbb{P})$; $\sigma(X)$ and $\sigma(Y)$ are then sub-σ-algebras of the common $\mathcal{F}$.

Marginalization

$$f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x, y)\, dy, \qquad p_X(x) = \sum_y p_{X,Y}(x, y).$$

7 Expectation

Let X be a random variable and $g : \mathbb{R} \to \mathbb{R}$ Borel measurable with $\mathbb{E}\big[|g(X)|\big] < \infty$. Then

$$\mathbb{E}\big[g(X)\big] = \begin{cases} \displaystyle\sum_x p_x\, g(x), & X \text{ discrete}, \\ \displaystyle\int_{\mathbb{R}} f_X(x)\, g(x)\, dx, & X \text{ continuous}. \end{cases}$$

Basic properties

- $\mathbb{P}(A) = \mathbb{E}[\mathbf{1}_A]$.
- (Linearity) $\forall c \in \mathbb{R}$: $\mathbb{E}[cX + Y] = c\,\mathbb{E}[X] + \mathbb{E}[Y]$.
- (Non-negativity) $X \geq 0 \implies \mathbb{E}[X] \geq 0$.
- $X \geq 0$ and $\mathbb{E}[X] = 0 \implies X = 0$ a.s., i.e.

$$\mathbb{P}\big(\{\omega \in \Omega : X(\omega) = 0\}\big) = 1.$$

- (Monotonicity) $X \geq Y \implies \mathbb{E}[X] \geq \mathbb{E}[Y]$, where $X \geq Y$ means either $X(\omega) \geq Y(\omega)$ for all $\omega \in \Omega$, or $X \geq Y$ a.s., i.e. $\mathbb{P}\big(\{\omega : X(\omega) < Y(\omega)\}\big) = 0$.

7.1 Jensen's inequality

Q. *Prove Jensen's inequality.*

Theorem 7.1 (Jensen)**.** *Let $f : \mathbb{R} \to \mathbb{R}$ be convex and let X be integrable with $\mathbb{E}[f(X)]$ well defined. Then*

$$f\big(\mathbb{E}[X]\big) \;\leq\; \mathbb{E}\big[f(X)\big].$$

Proof. **Step 1: the supporting line.** Fix $x, y \in \mathbb{R}$ and $t \in (0, 1)$. Convexity gives

$$f\big(tx + (1-t)y\big) \leq t f(x) + (1-t) f(y),$$

i.e., subtracting $f(y)$ from both sides,

$$f\big(y + t(x-y)\big) - f(y) \;\leq\; t\big(f(x) - f(y)\big).$$

Dividing by $t > 0$,

$$\frac{f\big(y + t(x-y)\big) - f(y)}{t} \;\leq\; f(x) - f(y).$$

Writing $h = t(x - y)$ (for $x \neq y$), the left side equals

$$(x-y) \cdot \frac{f(y+h) - f(y)}{h} \xrightarrow[t \downarrow 0]{} (x-y)\, f'_+(y),$$

where f'_+ is the right derivative, which exists for every convex f because the difference quotient is monotone in h. Therefore

$$\boxed{f(x) \;\geq\; f(y) + f'_+(y)\,(x-y)}$$

for all x, y.

Step 2: take expectations. Substituting $x = X$ and taking expectations (valid by monotonicity and linearity),

$$\mathbb{E}\big[f(X)\big] \;\geq\; f(y) + f'_+(y)\big(\mathbb{E}[X] - y\big) \qquad \text{for every } y.$$

Choosing $y = \mathbb{E}[X]$ makes the last term vanish, giving $\mathbb{E}[f(X)] \geq f\big(\mathbb{E}[X]\big)$. $\square$

▶ **Remark.** A convex function need not be differentiable everywhere (e.g. $f(u) = |u|$ at 0). The right derivative f'_+ always exists, and any subgradient works.

7.2 Variance and covariance

$$\mathrm{Var}(X) = \mathbb{E}\big[(X - \mathbb{E}[X])^2\big] = \mathbb{E}[X^2] - \big(\mathbb{E}[X]\big)^2, \qquad \mathrm{Cov}(X,Y) = \mathbb{E}\big[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\big].$$

Q. *State an example with* $\mathrm{Cov}(X,Y) = 0$ *but* $X \not\perp\!\!\!\perp Y$.

Example 7.2. $\mathrm{Cov}(X,Y) = 0$ does not imply $X \perp\!\!\!\perp Y$. Let

$$X = \begin{cases} -1 & \text{w.p. } \frac{1}{3} \\ 0 & \text{w.p. } \frac{1}{3} \\ 1 & \text{w.p. } \frac{1}{3} \end{cases} \qquad \text{and} \qquad Y = X^2.$$

Then $\mathbb{E}[X] = 0$ and $\mathbb{E}[XY] = \mathbb{E}[X^3] = \frac{1}{3}(-1) + \frac{1}{3}(0) + \frac{1}{3}(1) = 0$, so $\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0$. But X and Y are not independent:

$$\mathbb{P}(X = 0,\ Y = 0) = \tfrac{1}{3} \quad \neq \quad \mathbb{P}(X = 0)\,\mathbb{P}(Y = 0) = \tfrac{1}{3} \cdot \tfrac{1}{3} = \tfrac{1}{9}.$$

Indeed Y is a deterministic function of X.

Q. *State* $\mathrm{Var}(X+Y)$.

$$\mathrm{Var}(cX) = c^2\,\mathrm{Var}(X), \qquad \mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X,Y).$$

8 Markov and Chebyshev inequalities

Q. *State and prove Chebyshev's inequality.*

Theorem 8.1 (Generalized Markov / Chebyshev)**.** *Let* $\psi : \mathbb{R} \to [0,\infty)$ *be non-decreasing, let* $t \in \mathbb{R}$ *with* $\psi(t) > 0$*, and suppose* $\mathbb{E}[\psi(X)] < \infty$*. Then*

$$\mathbb{P}\big(\{X \geq t\}\big) \;\leq\; \frac{\mathbb{E}\big[\psi(X)\big]}{\psi(t)}.$$

Proof. Since ψ is non-decreasing, $X \geq t \implies \psi(X) \geq \psi(t)$, i.e. $\{X \geq t\} \subseteq \{\psi(X) \geq \psi(t)\}$, hence pointwise

$$\mathbf{1}_{\{X\geq t\}} \;\leq\; \mathbf{1}_{\{\psi(X)\geq\psi(t)\}} \;\leq\; \frac{\psi(X)}{\psi(t)}\,\mathbf{1}_{\{\psi(X)\geq\psi(t)\}} \;\leq\; \frac{\psi(X)}{\psi(t)},$$

where the second inequality uses $\psi(X)/\psi(t) \geq 1$ on the event, and the third uses $\psi \geq 0$. Taking expectations and using monotonicity,

$$\mathbb{P}(X \geq t) = \mathbb{E}\big[\mathbf{1}_{\{X\geq t\}}\big] \leq \frac{\mathbb{E}[\psi(X)]}{\psi(t)}. \qquad \square$$

▶ **Remark.** ψ non-decreasing gives only the inclusion $\{X \geq t\} \subseteq \{\psi(X) \geq \psi(t)\}$, not equality. (Equality would need ψ strictly increasing. Counterexample: $\psi \equiv 1$ constant makes the right-hand event all of Ω.)

Corollary 8.2 (Chebyshev)**.** *For any random variable X with $\mathrm{Var}(X) < \infty$ and any $t > 0$,*

$$\mathbb{P}\big(|X - \mathbb{E}[X]| \geq t\big) \;\leq\; \frac{\mathrm{Var}(X)}{t^2}.$$

Proof. Apply Theorem 8.1 to the non-negative variable $|X - \mathbb{E}[X]|$ with $\psi(u) = (u^+)^2$, so that $\mathbb{E}[\psi(|X - \mathbb{E}[X]|)] = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathrm{Var}(X)$ and $\psi(t) = t^2 > 0$. $\square$

▶ **Remark.** Markov's inequality $\mathbb{P}(X \geq t) \leq \mathbb{E}[X]/t$ $(t > 0)$ if $X \geq 0$ a.s.

9 Laws of large numbers and the CLT

Throughout, $S_n = X_1 + X_2 + \cdots + X_n$, and the goal is $\dfrac{S_n}{n} \longrightarrow \mathbb{E}[X]$ in a sense to be specified.

Q. *State and prove the weak law of large numbers.*

Theorem 9.1 (WLLN)**.** *Let $X_1, X_2, \ldots$ be i.i.d. with $\mathbb{E}[X^2] < \infty$. Then for every $\varepsilon > 0$,*

$$\mathbb{P}\Big(\Big\{\omega \in \Omega : \Big|\tfrac{S_n}{n} - \mathbb{E}[X]\Big| \geq \varepsilon\Big\}\Big) \xrightarrow[n\to\infty]{} 0,$$

that is, $S_n/n \xrightarrow{\mathbb{P}} \mathbb{E}[X]$.

Proof. Note $\mathbb{E}[S_n] = n\mathbb{E}[X]$ and, by independence, $\mathrm{Var}(S_n) = n\,\mathrm{Var}(X)$. Then

$$\begin{aligned}\mathbb{P}\Big(\Big|\tfrac{S_n}{n} - \mathbb{E}[X]\Big| \geq \varepsilon\Big) &= \mathbb{P}\big(\big|S_n - n\mathbb{E}[X]\big| \geq n\varepsilon\big) \\ &\leq \frac{\mathbb{E}\big[\big(S_n - \mathbb{E}[S_n]\big)^2\big]}{n^2\varepsilon^2} = \frac{\mathrm{Var}(S_n)}{n^2\varepsilon^2} = \frac{n\,\mathrm{Var}(X)}{n^2\varepsilon^2} = \frac{\mathrm{Var}(X)}{n\varepsilon^2} \xrightarrow[n\to\infty]{} 0. \qquad \square\end{aligned}$$

Remark 9.2. The hypothesis $\mathbb{E}[X^2] < \infty$ is what makes the Chebyshev step available. The WLLN in fact holds assuming only $\mathbb{E}|X| < \infty$, but the proof is different.

Q. *State the strong law of large numbers.*

Theorem 9.3 (SLLN)**.** *Let $X_1, X_2, \ldots$ be i.i.d. with $\mathbb{E}|X| < \infty$. Then*

$$\mathbb{P}\Big(\Big\{\omega : \lim_{n\to\infty} \tfrac{S_n}{n}(\omega) = \mathbb{E}[X]\Big\}\Big) = 1, \qquad \textit{equivalently} \qquad \mathbb{P}\Big(\Big\{\omega : \lim_{n\to\infty} \tfrac{S_n}{n}(\omega) \neq \mathbb{E}[X]\Big\}\Big) = 0.$$

Remark 9.4. The SLLN needs only a first moment, unlike the proof of the WLLN above. The "$\neq$" formulation is understood to include the case where the limit fails to exist.

Q. *State the Central Limit Theorem.*

Theorem 9.5 (CLT)**.** *Let $X_1, X_2, \ldots$ be i.i.d. with finite variance $\sigma^2 = \mathrm{Var}(X) < \infty$ (and $\sigma^2 > 0$). Then*

$$\frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathrm{Var}(S_n)}} = \frac{S_n - n\,\mathbb{E}[X]}{\sqrt{n}\,\sqrt{\mathrm{Var}(X)}} \xrightarrow{d} N(0,1).$$

10 Modes of convergence

Q. *State the definitions of $X_n \to X$ (all modes).*

Definition 10.1 ((0) Convergence in distribution)**.** $X_n \xrightarrow{d} X$ iff for every $t \in \mathbb{R}$ such that $\mathbb{P}(X = t) = 0$ (i.e. every continuity point of F_X),

$$\mathbb{P}\big(\{\omega : X_n(\omega) \le t\}\big) \longrightarrow \mathbb{P}\big(\{\omega : X(\omega) \le t\}\big) \qquad \text{as } n \to \infty.$$

Fix t; never look at X_n and X on the same ω.

Definition 10.2 ((1) Convergence in probability)**.** $X_n \xrightarrow{\mathbb{P}} X$ iff for every $\varepsilon > 0$,

$$\mathbb{P}\big(\{\omega : |X_n(\omega) - X(\omega)| > \varepsilon\}\big) \longrightarrow 0 \qquad \text{as } n \to \infty.$$

Fix n and examine different outcomes $X_n(\omega)$, $\omega \in \Omega$.

Definition 10.3 ((2) Almost sure convergence)**.** $X_n \to X$ a.s. iff

$$\mathbb{P}\Big(\Big\{\omega : \forall\varepsilon > 0,\ \exists N(\varepsilon,\omega) \text{ s.t. } \forall n \ge N,\ |X_n(\omega) - X(\omega)| < \varepsilon\Big\}\Big) = 1,$$

equivalently

$$\mathbb{P}\Big(\Big\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\Big\}\Big) = 1.$$

Fix ω and examine the sequence $X_1(\omega), X_2(\omega), X_3(\omega), \ldots$

Definition 10.4 ((3) $X_n = X$ eventually, a.s.)**.**

$$\mathbb{P}\Big(\Big\{\omega : \exists N(\omega) \text{ s.t. } \forall n \ge N,\ X_n(\omega) = X(\omega)\Big\}\Big) = 1.$$

Fix ω; then $X_n(\omega) = X(\omega)$ for all large enough n.

Remark 10.5 (Hierarchy)**.** (3) $\implies$ (2) $\implies$ (1) $\implies$ (0), and none of the implications reverses. Section 11 gives a counterexample at each level. The one partial converse: if the limit is a *constant* c, then $X_n \xrightarrow{d} c \implies X_n \xrightarrow{\mathbb{P}} c$.

10.1 Restating almost sure convergence via "infinitely often"

For a sequence of events (A_n), write

$$\{A_n \text{ i.o.}\} := \limsup_{n\to\infty} A_n = \bigcap_{N=1}^{\infty} \bigcup_{n\geq N} A_n,$$

the set of ω lying in infinitely many A_n. Concretely, fixing ω,

$$X_n(\omega) = 1 \text{ for infinitely many } n \iff \forall N \in \mathbb{N},\ \exists n \geq N \text{ s.t. } X_n(\omega) = 1.$$

Q. *Restate almost sure convergence.*

Proposition 10.6. *$X_n \to X$ almost surely if and only if*

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

Proof. Write $A_n^\varepsilon = \{|X_n - X| > \varepsilon\}$. Then

$$\Big\{\omega : \lim_n X_n(\omega) = X(\omega)\Big\}^c = \bigcup_{\varepsilon>0} \{A_n^\varepsilon \text{ i.o.}\} = \bigcup_{k=1}^{\infty} \{A_n^{1/k} \text{ i.o.}\},$$

the last equality because $\{A_n^\varepsilon \text{ i.o.}\}$ increases as $\varepsilon \downarrow 0$, so a countable union over $\varepsilon = 1/k$ suffices. A countable union of null sets is null, and conversely each term is contained in the union; so the union has probability 0 iff every term does. □

Q. *State the second Borel–Cantelli lemma.*

Lemma 10.7 (Borel–Cantelli I)**.** *If $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, then $\mathbb{P}(A_n \text{ i.o.}) = 0$.*

Lemma 10.8 (Borel–Cantelli II)**.** *If the events $(A_n)_{n\geq 1}$ are independent and $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$, then $\mathbb{P}(A_n \text{ i.o.}) = 1$.*

Remark 10.9. Independence is essential in BC II. Without it the conclusion fails: take $A_n = A$ for all n with $0 < \mathbb{P}(A) < 1$; then $\sum_n \mathbb{P}(A_n) = \infty$ but $\mathbb{P}(A_n \text{ i.o.}) = \mathbb{P}(A) < 1$.

11 Counterexamples separating the modes

11.1 (0) vs (1): converges in distribution but not in probability

Q. *Construct an example with $X_n \xrightarrow{d} X$ but $X_n \overset{\mathbb{P}}{\not\to} X$.*

Example 11.1. Let $X, X_1, X_2, \ldots \overset{\text{iid}}{\sim} \text{Bern}(\frac{1}{2})$.

$X_n \xrightarrow{d} X$**.** Each X_n has exactly the same law as X, so $F_{X_n}(t) = F_X(t)$ for every $t \in \mathbb{R}$; convergence at continuity points is immediate.

$X_n \overset{\mathbb{P}}{\not\to} X$**.** Since $|X_n - X| \in \{0, 1\}$, for every $\varepsilon \in (0, 1)$,

$$\mathbb{P}\big(\{\omega : |X_n(\omega) - X(\omega)| > \varepsilon\}\big) = \mathbb{P}\big(\{\omega : X_n(\omega) \neq X(\omega)\}\big) = \tfrac{1}{2} \not\to 0.$$

Remark 11.2. Convergence in distribution constrains only the marginal law of each X_n; it never inspects X_n and X at the same ω. Convergence in probability does. So any counterexample here is built by keeping the marginals right and wrecking the coupling.

Proposition 11.3 (Constant limits are special)**.** *For a constant c, $X_n \xrightarrow{d} c \implies X_n \xrightarrow{\mathbb{P}} c$.*

Proof. For $\varepsilon > 0$,

$$\mathbb{P}\big(|X_n - c| > \varepsilon\big) = \mathbb{P}(X_n > c+\varepsilon) + \mathbb{P}(X_n < c-\varepsilon) \le \big[1 - \mathbb{P}(X_n \le c+\varepsilon)\big] + \mathbb{P}(X_n \le c-\varepsilon).$$

The degenerate limit $X \equiv c$ satisfies $\mathbb{P}(X = c+\varepsilon) = \mathbb{P}(X = c-\varepsilon) = 0$, so $c \pm \varepsilon$ are continuity points of F_X. Since $X_n \xrightarrow{d} c$,

$$\mathbb{P}(X_n \le c+\varepsilon) \to 1, \qquad \mathbb{P}(X_n \le c-\varepsilon) \to 0,$$

so the right-hand side tends to $(1-1)+0 = 0$. Hence $\mathbb{P}(|X_n - c| > \varepsilon) \to 0$. $\square$

11.2 (1) vs (2): converges in probability but not almost surely

Q. *Construct an example with $X_n \xrightarrow{\mathbb{P}} X$ but $X_n \not\to X$ a.s.*

Example 11.4. Let $X_1, X_2, \ldots$ be *independent* with

$$X_n = \begin{cases} 1 & \text{w.p. } \frac{1}{n}, \\ 0 & \text{w.p. } 1 - \frac{1}{n}, \end{cases} \qquad \text{and} \qquad X = 0 \text{ a.s.}$$

$X_n \xrightarrow{\mathbb{P}} X$**.** For every $\varepsilon > 0$,

$$\mathbb{P}\big(\{\omega : |X_n(\omega) - X(\omega)| > \varepsilon\}\big) \;\le\; \tfrac{1}{n} \;\longrightarrow\; 0 \qquad \text{as } n \to \infty,$$

so $X_n \xrightarrow{\mathbb{P}} X$.

$X_n \not\to X$ **a.s.** By Proposition 10.6, a.s. convergence would require $\mathbb{P}(|X_n - X| > \varepsilon \text{ i.o.}) = 0$ for every $\varepsilon > 0$. Take $\varepsilon = \frac{1}{2}$. Since $X_n(\omega) \in \{0, 1\}$ and $X = 0$ a.s.,

$$\big\{|X_n - X| > \tfrac{1}{2} \text{ i.o.}\big\} = \big\{X_n = 1 \text{ i.o.}\big\} \quad \text{a.s.}$$

Q. *Rewrite $\{|X_n - X| \ne 0$ i.o.$\}$ as a set operation.*

$$\{X_n = 1 \text{ i.o.}\} \;=\; \bigcap_{N=1}^{\infty} \bigcup_{n \ge N} \{X_n = 1\}.$$

Now define $A_n := \{\omega : X_n(\omega) = 1\}$. The A_n are independent (because the X_n are), and

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

By Borel–Cantelli II,

$$\mathbb{P}(A_n \text{ i.o.}) = \mathbb{P}\Big(\bigcap_{N=1}^{\infty} \bigcup_{n \ge N} A_n\Big) = \mathbb{P}\Big(\bigcap_{N=1}^{\infty} \bigcup_{n \ge N} \{\omega : X_n(\omega) = 1\}\Big) = 1.$$

Therefore

$$\mathbb{P}\Big(\big\{\omega : |X_n(\omega) - X(\omega)| > \tfrac{1}{2} \text{ i.o.}\big\}\Big) = 1 \neq 0 \qquad \Longrightarrow \qquad X_n \not\to X \text{ a.s.}$$

▶ **Remark.** $\mathbb{P}(A_n \text{ i.o.})$ is not $\mathbb{P}\big(\bigcup_{n\geq 1}\{X_n = 1\}\big)$. The latter is the event "at least one n", not "infinitely many". The correct expression is the $\bigcap_N \bigcup_{n\geq N}$ above.

▶ **Remark.** The handwritten line "the necessary condition for a.s. convergence is. . ." asserts $\mathbb{P}(|X_n - X| > \frac{1}{2} \text{ i.o.}) = 0$ and then proves it equals 1. Phrased as an assertion this contradicts the next page. It should be stated as the condition a.s. convergence *would* require, and then shown to fail — as above. (The condition is in fact necessary *and* sufficient, by Proposition 10.6.)

11.3 (2) vs (3): converges a.s. but not eventually equal

Q. *Construct an example with $X_n \to X$ a.s. but not $X_n = X$ eventually a.s.*

Example 11.5. $X_n = \frac{1}{n}$ (deterministic) and $X = 0$ a.s.

$X_n \to X$ **a.s.** Given $\varepsilon > 0$, every $n > 1/\varepsilon$ satisfies $|X_n - X| = \frac{1}{n} < \varepsilon$. Hence

$$\mathbb{P}\Big(\Big\{\omega : \forall\varepsilon > 0,\ \forall n > \big\lfloor\tfrac{1}{\varepsilon}\big\rfloor + 1,\ |X_n(\omega) - X(\omega)| = \tfrac{1}{n} < \varepsilon\Big\}\Big) = 1 \quad\Longrightarrow\quad X_n \to X \text{ a.s.}$$

But not eventually equal. For *every* n, $|X_n - X| = \frac{1}{n} \neq 0$, so

$$\mathbb{P}\big(\{\omega : \forall n,\ X_n(\omega) \neq X(\omega)\}\big) = 1,$$

so no $N(\omega)$ as in Definition (3) can exist.

11.4 (3): an example where $X_n = X$ eventually, a.s.

Q. *Construct an example with $X_n \to X$ a.s. eventually.*

Example 11.6. Take the probability space $\big([0,1],\ \mathcal{B}([0,1]),\ \lambda\big)$ with λ Lebesgue measure, and set

$$X_n(\omega) = \begin{cases} 1, & \omega \in \big[0, \frac{1}{n}\big), \\ 0, & \omega \in \big[\frac{1}{n}, 1\big], \end{cases} \qquad X = 0 \text{ a.s.}$$

For any $\omega \neq 0$: once $n \geq 1/\omega$ we have $\omega \geq \frac{1}{n}$, hence $X_n(\omega) = 0 = X(\omega)$. So with $N(\omega) = \big\lfloor\frac{1}{\omega}\big\rfloor + 1$,

$$\mathbb{P}\Big(\Big\{\omega : \omega \neq 0,\ \forall n > \big\lfloor\tfrac{1}{\omega}\big\rfloor + 1,\ X_n(\omega) = X(\omega) = 0\Big\}\Big) = 1,$$

since $\mathbb{P}(\{\omega : \omega = 0\}) = \lambda(\{0\}) = 0$.

Remark 11.7. The exclusion $\omega \neq 0$ matters: $X_n(0) = 1$ for every n, so the convergence genuinely fails at $\omega = 0$. It is exactly because $\lambda(\{0\}) = 0$ that the statement holds "almost surely" rather than "for all ω", and it is why $\lfloor 1/\omega\rfloor$ — undefined at $\omega = 0$ — causes no trouble.