Probability Theory

Cheat Sheet & Study materials*

These materials...

Contents

---

*Luka Šimek, 2025./2026.

1 Set structures

1.1 Definitions

Let S be a set and $\mathcal{S} \subseteq \mathcal{P}(S)$ a family of subsets. We define a number of terms: $\mathcal{S}$ is a

(a) **semiring** if

- $A, B \in \mathcal{S} \implies A \cap B \in \mathcal{S}$
- if $A, B \in \mathcal{S}$ then $A \setminus B$ is a finite disjoint union of sets in $\mathcal{S}$

(b) **ring** if

- $A, B \in \mathcal{S} \implies A \cap B \in \mathcal{S}$
- $A, B \in \mathcal{S} \implies A \setminus B \in \mathcal{S}$

(c) **semialgebra** if

- $A, B \in \mathcal{S} \implies A \cap B \in \mathcal{S}$
- $A \in \mathcal{S} \implies A^c$ is a finite disjoint union of sets in $\mathcal{S}$

(d) **algebra** if it is closed under complements and **finite** unions.

(e) **monotone class** if $A_j \uparrow A$ and $A_j \downarrow A$ imply $A \in \mathcal{S}$

(f) **σ-algebra** if...

(g) **π-system** if $A, B \in \mathcal{S} \implies A \cap B \in \mathcal{S}$.

(h) **λ-system** if

- $S \in \mathcal{S}$
- $A \subseteq B,\ A, B \in \mathcal{S} \implies B \setminus A \in \mathcal{S}$
- $A_j \uparrow A \implies A \in \mathcal{S}$

* * *

Clearly, a σ-algebra is all of the above. Also:

- if $\mathcal{S}$ is both a monotone class and an algebra, it is a σ-algebra,
- if $\mathcal{S}$ is an algebra closed with respect to disjoint unions, it is a σ-algebra,
- if $\mathcal{S}$ is both a π-system and λ-system, it is a σ-algebra.

1.2 Equivalent definition of λ-system

A λ-system can also be defined by:

- $S \in \mathcal{S}$
- $A \in \mathcal{S} \implies A^c \in \mathcal{S}$
- if $A_j \in \mathcal{S}$ with $j \in \mathbb{N}$ are disjoint, then $\cup_j A_j \in \mathcal{S}$.

Suppose the first definition and let us prove the other. We simply get complements from $A^c = S \setminus A$ and $S \in \mathcal{S}$. Suppose A_1 and A_2 are disjoint. Then $A_1 \subseteq A_2^c$ so $A_2^c \setminus A_1 \in \mathcal{S}$. This means we have finite disjoint unions because

$$A_2^c \setminus A_1 = A_2^c \cap A_1^c = (A_1 \cup A_2)^c .$$

With finite disjoint unions,

$$\bigcup_{j=1}^{\infty} A_j = \bigcup_{j=1}^{\infty} \left[\bigcup_{k=1}^{j} A_j \right] \in \mathcal{S}$$

as this is a monotone union.

Now let us prove the other direction. If $A \subseteq B$ then $B \setminus A = B \cap A^c = (B^c \cup A)^c$. Then suppose A_j are monotone increasing. We have the disjoint union

$$\bigcup_{j=1}^{\infty} A_j = A_1 \cup \bigcup_{j=2}^{\infty} \left[A_j \setminus \bigcup_{k=1}^{j-1} A_k \right] \in \mathcal{S}$$

as all the set differences are proper.

1.3 Dynkin–Sierpinski π-λ lemma

Let S be a set, $\mathcal{C}$ a π-system and $\mathcal{D}$ a λ-system. Then

$$\mathcal{C} \subseteq \mathcal{D} \implies \sigma(\mathcal{C}) \subseteq \mathcal{D}.$$

The proof is from [3, p. 10]. Clearly we can assume that $\mathcal{D}$ is the smallest possible λ-system, this being denoted $\lambda(\mathcal{D})$. Taking into account that π and λ together make a σ-algebra, the statement becomes:

> If $\mathcal{C}$ is a π-system, then $\lambda(\mathcal{C})$ is also a π-system.

Therefore, we need to prove $A \cap B \in \mathcal{D}$ whenever $A, B \in \mathcal{D}$. For an arbitrary $B \in \mathcal{C}$ define

$$\mathcal{S}_B = \{A \in \mathcal{D} \colon A \cap B \in \mathcal{D}\} .$$

Clearly $\mathcal{S}_B$ is a λ-system containing $\mathcal{C}$ (because $\mathcal{C}$ is itself a π-system). Since $\mathcal{D} = \lambda(\mathcal{C})$, it follows that $\mathcal{D} \subseteq \mathcal{S}_B$. That is, $A \cap B \in \mathcal{D}$ whenever $B \in \mathcal{C}$ and $A \in \mathcal{D}$.

Now for any $A \in \mathcal{D}$ define

$$\mathcal{S}'_A = \{B \in \mathcal{D} \colon A \cap B \in \mathcal{D}\} .$$

Similarly, and also using the previous result, we get $\mathcal{D} \subseteq \mathcal{S}'_A$ and we conclude that $A \cap B \in \mathcal{D}$ whenever $A, B \in \mathcal{D}$.

* * *

Suppose $\mathcal{C} \subseteq \mathcal{D} \subseteq \mathcal{F}$ where $\mathcal{F}$ is a σ-algebra on a probability space. Let $\mathcal{D}$ be the set of all events with a certain property, this property naturally making it a λ-system. Suppose also $\sigma(\mathcal{C}) = \mathcal{F}$. By the lemma, $\mathcal{D} = \mathcal{F}$, that is, all events have the desired property.

1.4 Notable exercises

1.4.1 Any open set of reals is a countable union of intervals

See [1, ex. 1.1.2]. Define

$$\mathcal{S}_d = \left\{ \prod_{j=1}^{d} \langle a_j, b_j] : -\infty \leqslant a < b \leqslant +\infty \right\} \subseteq \mathbb{R}^d.$$

Note that $\mathcal{S}_d$ is a semialgebra (automatically, also a π-system). We claim that $\sigma(\mathcal{S}_d) = \mathcal{B}(\mathbb{R}^d)$. It suffices to prove that every open set in $\mathbb{R}^d$ can be written as the countable union of sets in $\mathcal{S}_d$.

Take $d = 1$ with a simple extension afterwards. Let $A \subseteq \mathbb{R}$ be open, so that for each $x \in A$ there is ε_x such that $B(x, \varepsilon_x) \subseteq A$. Also,

$$A = \bigcup_{x \in A} B(x, \varepsilon_x).$$

Note that $A \cap \mathbb{Q}$ is dense in A. Namely, each x is within its ball, and within the ball is a sequence of rationals converging to x. For each rational y such that $|x - y| < \varepsilon_x/2$ we can clearly choose an ε_y such that $x \in B(y, \varepsilon_y) \subseteq A$. This gives us the countable union[1]

$$A = \bigcup_{y \in A \cap \mathbb{Q}} B(y, \varepsilon_y)$$

* * *

1.4.2 Infinite unions of algebras/σ-algebras

See [1, ex. 1.1.4]. Suppose $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \cdots$ and $\mathcal{F} = \cup_j \mathcal{F}_j$.

If $\mathcal{F}_j$ are algebras, then $\mathcal{F}$ is also an algebra because

$$\bigcup_{j \in N} A_j \in \mathcal{F}_{\max N}, \quad A_j \in \mathcal{F}_j.$$

where $N \subseteq \mathbb{N}$ is finite.

The analogous claim is not true if $\mathcal{F}_j$ are σ-algebras. Let $S = \mathbb{R}$ and

$$\mathcal{F}_j = \sigma\left(\left[\frac{n}{2^j}, \frac{n+1}{2^j} \right\rangle : n \in \mathbb{N} \right), \quad j \in \mathbb{N}.$$

If $\mathcal{F}$ is a σ-algebra,

$$\{0\} = \bigcap_{j \in \mathbb{N}} \left[0, \frac{1}{2^j} \right\rangle \in \mathcal{F},$$

but there is no j such that $\{0\} \in \mathcal{F}_j$, giving us a contradiction.

[1] while different x can give different ε_y, these clearly have a finite supremum

2 Measure and integration

2.1 Products of σ-algebras

2.1.1 Defintions of product σ-algebra

If $(S_\alpha)_{\alpha\in A}$ are measurable spaces with σ-algebras $\mathcal{F}_\alpha$ and $S = \prod_\alpha S_\alpha$, then S has the **product σ-algebra** $\mathcal{F} = \prod_\alpha \mathcal{F}_\alpha = \bigotimes_\alpha \mathcal{F}_\alpha$ defined as

$$\mathcal{F} = \sigma\left(\pi_\alpha^{-1}(E_\alpha)\colon \alpha \in A,\ E_\alpha \in \mathcal{F}_\alpha\right),$$

i.e. **generated by one-dimensional measurable cylinders**.

* * *

If A is countable, then $\mathcal{F}$ is **generated by cuboids** $\prod_\alpha E_\alpha$, $\alpha \in A$ (see [2, p. 23]). To prove it, define

$$\mathcal{F}_1 = \sigma\left(\pi_\alpha^{-1}(E_\alpha)\colon \alpha \in A,\ E_\alpha \in \mathcal{F}_\alpha\right), \quad \mathcal{F}_2 = \sigma\left(\prod_{\alpha\in A} E_\alpha\colon E_\alpha \in \mathcal{F}_\alpha\right).$$

Firstly, $\pi_\alpha^{-1}(E_\alpha) = \prod_\alpha E_\alpha$ where $E_\beta = S_\beta$ for $\beta \neq \alpha$, so $\mathcal{F}_1 \subseteq \mathcal{F}_2$.

Secondly,

$$\prod_{\alpha\in A} E_\alpha = \bigcap_{\alpha\in A} \pi_\alpha^{-1}(E_\alpha) \in \mathcal{F}_1,$$

so that $\mathcal{F}_2 \subseteq \mathcal{F}_1$.

* * *

Instead of taking E_α's from the σ-algebras, we can restrict ourselves to **sets of generators**. Suppose $\sigma(\mathcal{E}_\alpha) = \mathcal{F}_\alpha$, then (the second part again if A is countable, and proven the same way)

$$\bigotimes_{\alpha\in A} \mathcal{F}_\alpha = \sigma\left(\pi_\alpha^{-1}(E_\alpha)\colon \alpha \in A,\ E_\alpha \in \mathcal{E}_\alpha\right) = \sigma\left(\prod_{\alpha\in A} E_\alpha\colon E_\alpha \in \mathcal{E}_\alpha\right).$$

To prove the first part, consider the sets

$$\mathcal{D} = \left\{\pi_\alpha^{-1}(E_\alpha)\colon \alpha \in A,\ E_\alpha \in \mathcal{E}_\alpha\right\}, \quad \mathcal{C}_\alpha = \left\{E \in F_\alpha\colon \pi_\alpha^{-1}(E) \in \sigma(\mathcal{D})\right\}.$$

Then C_α is a σ-algebra containing $\mathcal{E}_\alpha$, which means that $C_\alpha = \mathcal{F}_\alpha$ and $\sigma(\mathcal{D}) = \bigotimes_\alpha \mathcal{F}_\alpha$.

2.1.2 Borel σ-algebra on product metric space

Suppose S_j are separable metric spaces ($j \in \mathbb{N}$) with Borel σ-algebras $\mathcal{B}_j$. Then (see [3, p. 11])

$$\mathcal{B}(S_1 \times S_2 \times \cdots) = \mathcal{B}_1 \otimes \mathcal{B}_2 \otimes \cdots \tag{2.1}$$

Consider the sets

$$\mathcal{C}_1 = \left\{A\colon A \text{ open in } \bigtimes_j S_j\right\}, \quad \mathcal{C}_2 = \left\{S_1 \times \cdots \times B_j \times S_{j+1} \times \cdots : B_j \in \mathcal{B}_j\right\},$$

and $\mathcal{C}$ defined similar to $\mathcal{C}_2$ but where B_j has to be open instead. The claim is then $\sigma(\mathcal{C}_1) = \sigma(\mathcal{C}_2)$.

Clearly $\mathcal{C} \subseteq \mathcal{C}_1, \mathcal{C}_2$. But it is also clear that $\sigma(\mathcal{C}) \supseteq \sigma(\mathcal{C}_2)$, so $\sigma(\mathcal{C}) = \sigma(\mathcal{C}_2)$. This proves that the $\supseteq$ part in (2.1) holds without separability.

If S_j are separable, $\mathcal{C}$ is a topological basis of $\bigtimes_j S_j$, so that $\sigma(\mathcal{C}) \supseteq \sigma(\mathcal{C}_1)$. Then $\sigma(\mathcal{C}) = \sigma(\mathcal{C}_1)$ as well.

2.1.3 Measurability in coordinate functions

See [3, p. 15]. Suppose that $(\Omega, \mathcal{A})$ and $(S_j, \mathcal{S}_j)$ are measurable spaces for $j \in \mathbb{N}$, denoting $S = \bigtimes_j S_j$. Let $f\colon \Omega \to S$ be a function and define f_j as its **coordinate functions** $f_j = \pi_j \circ f$. Then

> f is measurable if and only if all f_j are measurable.

The "only if" part is trivial, because each f_j is a composition of measurable f and π_j. For the "if" part, note that f satisfies the definition of measurability on a generating subset, those being measurable cuboids:

$$f^{-1}(B) = \bigcap_j f_j^{-1}(B_j), \quad B = B_1 \times B_2 \times \cdots, \tag{2.2}$$

where B and all B_j are measurable sets.

Note that (2.2) indeed holds specifically for cuboid sets B and not generally. Instead of these cuboids, we could have also chosen the simpler generating set of one-dimensional cylinders $\pi_j^{-1}(B_j)$.

2.2 Convergence theorems

We consider whether $\int f_n \to \int f$ if $f \to f_n$ pointwise, where the functions are defined on a σ-finite measure space X. Following [1, §1.4], we present the convergence theorem in order: bounded, Fatou's lemma, monotone, dominated. Note that in all cases pointwise convergence can be replaced with a.e. convergence, which is weaker but without affecting integration.

2.2.1 Bounded convergence theorem

Suppose the f_n have

- **bounded domain**: f_n vanishes on E^c for some E with $\mu(E) < \infty$,
- **bounded range**: $|f_n| \leqslant M$ uniformly,
- $f_n \to f$ **in measure**.

Then $\int f_n \to \int f$.

To **prove** it, we have:

$$\begin{aligned} \left| \int f - f_n \right| &\leqslant \int |f - f_n| \\ &= \int_{\{|f-f_n| \geqslant \varepsilon\}} |f - f_n| + \int_{\{|f-f_n| < \varepsilon\}} |f - f_n| \\ &\leqslant 2M\mu\left(|f - f_n| \geqslant \varepsilon\right) + \varepsilon\mu(E) \to 0, \quad n \to \infty, \ \varepsilon \downarrow 0. \end{aligned}$$

We implicitly used that $|f| \leqslant M$ as well. We obtain this because $|f| \geqslant M + \varepsilon$ implies $|f - f_n| \geqslant \varepsilon$, giving us $\mu\left(|f| > M\right) = 0$ after some simple work.

2.2.2 Fatou's lemma

Suppose $f_n \geqslant 0$, then

$$\liminf_{n\to\infty} \int f_n \geqslant \int \liminf_{n\to\infty} f_n.$$

Recall that

$$\liminf_{n\to\infty} f_n = \sup_m \inf_{n\geqslant m} f_n,$$

thus we define $g_n = \inf_{m\geqslant n} f_m$ so that $g_n \uparrow g := \liminf_n f_n$.

It then suffices to show (note $f_n \geqslant g_n$)

$$\liminf_{n\to\infty} \int g_n \geqslant \int g.$$

Let $E_m \uparrow X$ be measurable sets with $\mu(E_m) < \infty$. For fixed m:

$$(g_n \wedge m) \cdot 1_{E_m} \to (g \wedge m) \cdot 1_{E_m}.$$

Finally,

$$\liminf_{n\to\infty} \int g_n \geqslant \int_{E_m} g_n \wedge m \to \int_{E_m} g \wedge m \to \int g.$$

The inequality holds on a subsequence, the first convergence ($n \to \infty$) is the bounded convergence theorem, and the second ($m \to \infty$) is the below lemma.

* * *

In the proof we used the following **lemma** (see [1, p. 22]): let $E_n \uparrow X$ be measurable sets with $\mu(E_n) < \infty$. Then

$$\int_{E_n} f \wedge n \to \int f.$$

Clearly the left side is increasing and below $\int f$. By definition of the Lebesgue integral, it suffices to show that for every simple nonnegative φ, there exists n with

$$\int \varphi \leqslant \int_{E_n} f \wedge n \leqslant \int f.$$

Let N be such that $\mu(\varphi > N) = 0$ and choose $n \geqslant N$. Then

$$\begin{aligned}\int_{E_n} f \wedge n \geqslant \int_{E_n} \varphi &= \int \varphi - \int_{E_n^c} \varphi \\ &\geqslant \int \varphi - N\mu(E_n^c),\end{aligned}$$

so that

$$\liminf_{n\to\infty} \int_{E_n} f \wedge n \geqslant \int \varphi,$$

since $\mu(E_n^c) \to 0$.

2.2.3 Monotone convergence theorem

Suppose $f_n \geqslant 0$ and $f_n \uparrow f$, then

$$\int f_n \to \int f.$$

Since $f_n \leqslant f$ we have $\limsup_n \int f_n \leqslant \int f$, and by Fatou's lemma we have $\liminf_n \int f_n \geqslant \int f$.

2.2.4 Dominated convergence theorem

Let f_n be such that $|f_n| \leqslant g$ uniformly, for some integrable g. If $f_n \to f$, then

$$\int f_n \to \int f.$$

The condition $|f_n| \leqslant g$ can be written as $g \pm f_n \geqslant 0$. With $g \pm f_n \to f$, we apply Fatou's lemma twice. The $+$ and $-$ parts respectively give $\int f \geqslant \limsup_n \int f_n$ and $\int f \leqslant \liminf_n \int f_n$.

2.3 Measure extension

Carathéodory's theorem (multiple variations) extends measure from family of subsets to whole σ-algebra (unique?); Kolmogorov's theorem extends measures on finite-dimensional subspaces to a single measure on an infinite-dimensional metric probability space.

2.4 Product measure

Suppose $(S, \mathcal{S}, \mu)$ and $(T, \mathcal{T}, \nu)$ are σ-finite measure spaces. We wish to introduce a measure on the product measurable space $(S \times T, \mathcal{S} \otimes \mathcal{T})$.

2.4.1 Sections

For any **measurable set** $E \in \mathcal{S} \otimes \mathcal{T}$, $s \in S$ and $t \in T$ we define the sections

$$E_s = \{t \in T \colon (s,t) \in E\}, \quad E^t = \{s \in S \colon (s,t) \in E\}.$$

For any measurable **function** $f\colon S \times T \to U$ and any $s \in S$ and $t \in T$ we define the sections $f_s\colon T \to U$ and $f^t\colon S \to U$ with

$$f_s(t) = f(s,t), \quad f^t(s) = f(s,t).$$

We claim that the sections are **always measurable** (sets or functions). Consider first the set E_s. Define

$$\mathcal{D} = \{E \in \mathcal{S} \otimes \mathcal{T} \colon E_s \in \mathcal{T} \text{ for all } s \in S\}.$$

Clearly, $\mathcal{D}$ is a λ-system and contains the generating π-system of rectangles $A \times B$ since $(A \times B)_s = B$ if $s \in A$ and $\emptyset$ otherwise. By Dynkin's lemma, $\mathcal{D} = \mathcal{S} \otimes \mathcal{T}$.

To prove the measurability of functions f_s, it is simple to show that $f_s^{-1}(B) = \left(f^{-1}(B)\right)_s \in \mathcal{T}$, for any measurable B.

* * *

Let $f\colon S \times T \to \mathbb{R}_{\geqslant 0}$ be a measurable functions. We claim that the **maps**

$$s \mapsto \int_T f(s,t)\nu(\mathrm{d}t), \quad t \mapsto \int_S f(s,t)\mu(\mathrm{d}s)$$

are measurable.

We use a type of Lebesgue's induction. Assume first that μ and ν are finite and that $f = 1_E$ where $E = A \times B$. Then

$$\int_T 1_E(s,t)\nu(\mathrm{d}t) = \nu(E_s) = \nu(B) \cdot 1_A(s).$$

Clearly, $s \mapsto \nu(B) \cdot 1_A(s)$ is measurable. We wish to extend this to **arbitrary indicator function** so we define

$$\mathcal{D} = \{E \in \mathcal{S} \otimes \mathcal{T} : s \mapsto \nu(E_s) \text{ is measurable}\}$$

Again, $\mathcal{D}$ is a λ-system containing the generating π-system of measurable rectangles, so $\mathcal{D} = \mathcal{S} \otimes \mathcal{T}$.

This is then extended to simple functions (linearity of integration) and to general nonnegative measurable functions. For the latter, note that monotone convergence of integrals $\int_T f_n(s,t)\nu(\mathrm{d}t)$ means the pointwise convergence of the maps $s \mapsto \int_T f_n(s,t)\nu(\mathrm{d}t)$, and the class of measurable functions is closed under pointwise limits (see [1, ex. 1.2.7])

Note. The requirement $f \geqslant 0$ serves to insure that the integrals exist. An alternative requirement is $\int |f| \,\mathrm{d}\xi < \infty$ (with ξ product measure).

Note. The assumption of finite measures (important for $\mathcal{D}$ being a λ-system) does not lose on generality because of the following: take $F_n \uparrow T$ with $\nu(F_n) < \infty$. Then the maps $s \mapsto \nu(E_s \cap F_n)$ converge pointwise to $s \mapsto \nu(E_s)$ with pointwise convergence preserving measurability. The family

$$\mathcal{D}' = \{E \in \mathcal{S} \otimes \mathcal{T} : s \mapsto \nu(E_s \cap F_n) \text{ is measurable for all } n\}$$

is certainly a λ-system, ensuring the measurability of all $s \mapsto \nu(E_s \cap F_n)$.

2.4.2 Definitions

We can define a measure ξ on measurable rectangles with

$$\xi(A \times B) = \mu(A)\nu(B), \quad A \in \mathcal{S},\ B \in \mathcal{T}.$$

The function ξ is then extended to a unique measure on $\mathcal{S} \otimes \mathcal{T}$ by Carathéodory-type theorems.

* * *

Alternatively, we can define

$$\xi'(E) = \int_T \mu(E^t)\nu(\mathrm{d}t) = \int_S \nu(E_s)\mu(\mathrm{d}s), \quad E \in \mathcal{S} \otimes \mathcal{T}. \tag{2.3}$$

* * *

Let us **prove** these definitions are equivalent. Formally using Fubini's theorem:

$$\begin{aligned}\xi(E) &= \int_{S\times T} 1_E(u)\xi(\mathrm{d}u) \\ &= \int_S \mu(\mathrm{d}s) \int_T 1_E(s,t)\nu(\mathrm{d}t) \\ &= \int_S \nu(E_s)\mu(\mathrm{d}s) = \xi'(E).\end{aligned} \tag{2.4}$$

Conversely, for $E = A \times B$,

$$\xi'(E) = \int_S \nu(E_s)\mu(\mathrm{d}s) = \int_T \nu(B)1_A(s)\mu(\mathrm{d}s) = \mu(A)\nu(B) = \xi(E). \tag{2.5}$$

$$* \overset{*}{} *$$

We should also **justify** the second equality in (2.3). Define $\xi'(E) = \int_T \mu(E^t)\nu(\mathrm{d}s)$ and $\xi''(E) = \int_S \nu(E_s)\mu(\mathrm{d}t)$. By (2.5) ξ' and ξ'' agree on the generating π-system of measurable rectangles. By a well known lemma, $\xi' = \xi''$. This argument also suffices to prove $\xi = \xi'$ (Fubini's theorem is not necessary, (2.4) serves a heuristic purpose).

2.4.3 Fubini's theorem

Let $f: S \times T \to \mathbb{R}$ be a measurable function such that $\int_{S\times T} |f| \, \mathrm{d}\xi < \infty$. Then

$$\int_{S\times T} f \, \mathrm{d}\xi = \int_S \mu(\mathrm{d}s) \int_T f(s,t)\nu(\mathrm{d}t) = \int_T \nu(\mathrm{d}t) \int_S f(s,t)\nu(\mathrm{d}s). \tag{2.6}$$

Consider first a measurable function $f \geqslant 0$. By previous discussion, the integrals are well-defined and ξ satisfies:

$$\xi(E) = \int_S \mu(\mathrm{d}s) \int_T 1_E(s,t)\nu(\mathrm{d}t), \quad E \in \mathcal{S} \otimes \mathcal{T},$$

where we also know the order of integration can be reversed. Thus (2.6) holds for indicator functions, and is extended to arbitrary nonnegative functions by linearity and monotone convergence. For general integrable f, apply the previous case to f^+ and f^- and subtract. Integrability ensures that the integrands are measurable and the difference is well-defined.

2.5 Notable exercises

2.5.1 Equality of measures agreeing on a generating π-system

Let μ and ν be finite measures on measurable space $(S, \mathcal{S})$. If

$$\mu(C) = \nu(C), \quad C \in \mathcal{C},$$

where $\mathcal{C}$ is a generating π-system, then $\mu = \nu$. This can be extended to σ-finite measures.

To **prove** this, note that

$$\mathcal{D} = \{E \in \mathcal{S} : \mu(E) = \nu(E)\}$$

is a λ-system containing $\mathcal{C}$.

For **σ-finite** measures, note that the previous implies that $\mu = \nu$ when restricted to finite measure spaces $(S \cap E_n, \mathcal{S} \cap E_n)$, where $E_n \uparrow S$ and $\mu(E_n) < \infty$. Then for any $E \in \mathcal{S}$

$$\mu(E) = \lim_{n\to\infty} \mu(E \cap E_n) = \lim_{n\to\infty} \nu(E \cap E_n) = \nu(E).$$

2.5.2 Pointwise convergence preserves measurability

See [1, ex. 1.3.7]. Suppose $f_n \to f$ pointwise with f_n measurable. Then f is measurable because

$$\{f \leqslant a\} = \bigcap_{r\in\mathbb{N}} \bigcup_{k=1}^{\infty} \bigcap_{n \geqslant k} \left\{ f_n \leqslant a + \frac{1}{r} \right\}.$$

2.5.3 Tail-formula for p-th moment

If f is a measurable function, then

$$\|f\|_p = \left(p \int_0^\infty y^{p-1} \mu(|f(x)| \geqslant y)\, \mathrm{d}y \right)^{1/p}.$$

It can be derived, using Fubini for nonnegative functions:

$$\begin{aligned} \|f\|_p^p &= \int_{\mathbb{R}} |f(x)|^p \, \mathrm{d}x \\ &= \int_{\mathbb{R}} \int_0^{|f(x)|} p y^{p-1} \, \mathrm{d}y \, \mathrm{d}x \\ &= \int_{\mathbb{R}} p y^{p-1} \int_{\mathbb{R}} 1_{[0,|f(x)|]}(y) \, \mathrm{d}x \, \mathrm{d}y \\ &= p \int_0^\infty y^{p-1} \mu(|f(x)| \geqslant y) \, \mathrm{d}y. \end{aligned}$$

* * *

This holds for all $0 < p < \infty$. In probability, this is written

$$\mathbb{E}(X^p) = p \int_0^\infty x^{p-1} \mathbb{P}(X \geqslant x) \, \mathrm{d}x, \quad X \geqslant 0.$$

* * *

If probability if $X \geqslant 0$ attains **only integer values**, then for $x \in \langle n, n+1]$ we have $\mathbb{P}(X \geqslant x) = \mathbb{P}(X \geqslant n+1)$ giving us

$$\begin{aligned} \mathbb{E}(X^p) &= p \int_0^\infty x^{p-1} \mathbb{P}(X \geqslant x) \, \mathrm{d}x \\ &= p \sum_{n=0}^\infty \int_{\langle n,n+1]} x^{p-1} \mathbb{P}(X \geqslant x) \, \mathrm{d}x \\ &= p \sum_{n=0}^\infty \int_{\langle n,n+1]} x^{p-1} \mathbb{P}(X \geqslant n+1) \, \mathrm{d}x \\ &= p \sum_{n=0}^\infty \mathbb{P}(X \geqslant n+1) \int_{\langle n,n+1]} x^{p-1} \, \mathrm{d}x \\ &= \sum_{n=0}^\infty \left[(n+1)^p - n^p\right] \mathbb{P}(X \geqslant n+1) \\ &= \sum_{n=1}^\infty \left[n^p - (n-1)^p\right] \mathbb{P}(X \geqslant n), \end{aligned}$$

with the exchange of integrals being Fubini for nonnegatives.

* * *

Another **generalization** is for H of the form $H(x) = \int_{-\infty}^{x} h(y)\,\mathrm{d}y$ where $h\colon \mathbb{R} \to \mathbb{R}_{\geqslant 0}$. Then

$$\begin{aligned}\mathbb{E}H(X) &= \mathbb{E}\int_0^X h(y)\,\mathrm{d}y \\ &= \mathbb{E}\int_0^\infty h(y)1_{\{y\leqslant X\}}\,\mathrm{d}y \\ &= \int_0^\infty h(y)\mathbb{P}(X \geqslant y)\,\mathrm{d}y,\end{aligned}$$

again with Fubini for nonnegatives.

2.5.4 ∞-norm as limit of p-norms

We claim that

$$\|f\|_\infty = \lim_{p\to\infty} \|f\|_p\,.$$

where f is a measurable function on a probability space.

Let $\varphi = \sum_{j=1}^n a_j 1_{A_j}$ be a **simple** function on an **arbitrary** measure space, so that $\|\varphi\|_p^p = \sum_j |a_j|^p \mu(A_j)$. Define also $M = \max\{|a_j| : 1 \leqslant j \leqslant n\} = \|\varphi\|_\infty$ and let m be the corresponding index. Clearly $\|\varphi\|_p \leqslant \|\varphi\|_\infty$, but also

$$\|\varphi\|_p \geqslant M\left(\mu(A_m)\right)^{1/p} \to M, \quad p \to \infty,$$

proving the claim for simple functions on arbitrary measure spaces.

Similar ideas are used for general f on a **probability** space. Again $\|f\|_p \leqslant \|f\|_\infty$. Take arbitrary $\varepsilon > 0$ and set $A_\varepsilon = \{|f| > M - \varepsilon\}$. By definition of M, $0 < \mu(A_\varepsilon) < \infty$. Then,

$$\|f\|_p^p \geqslant \int_{A_\varepsilon} |f|^p\,\mathrm{d}\mu \geqslant (M-\varepsilon)^p \mu(A_\varepsilon).$$

Again $\liminf_p \|f\|_p \geqslant M - \varepsilon$, and the claim follows by $\varepsilon \downarrow 0$.

3 Basics of probability

3.1 Inequalities

3.1.1 Jensen

Suppose X is a random variable and $\varphi\colon \mathbb{R} \to \mathbb{R}$ is a **convex** function. Then

$$\varphi(\mathbb{E}X) \leqslant \mathbb{E}(\varphi X).$$

For $c = \mathbb{E}X$, consider an affine function $h(x) = ax + b$ lying under φ ($h \leqslant \varphi$) but with $h(c) = \varphi(c)$ (the graph of h is separating two disjoint convex sets: epigraph of φ and $\langle c, \infty\rangle \times \langle -\infty, c\rangle$). Then

$$\varphi(\mathbb{E}X) = h(\mathbb{E}X) = a\mathbb{E}X + b = \mathbb{E}(aX + b) = \mathbb{E}h(X) \leqslant \mathbb{E}\varphi(X).$$

3.1.2 Hölder

For any measurable functions f, g and real $p, q, r > 0$ such that $1/r = 1/p + 1/q$,

$$\|fg\|_r \leqslant \|f\|_p \|g\|_q .$$

Clearly we can take $r = 1, \|f\|_p = 1, \|g\|_q = 1$ (**unless** $f = 0$ or $g = 0$ which is a trivial edge-case). It is true that

$$uv \leqslant \frac{u^p}{p} + \frac{v^q}{q}, \quad u, v \geqslant 0. \tag{3.1}$$

By setting $u = |f(x)|$ and $v = |g(x)|$ and integrating, we get

$$\int |fg| \leqslant p^{-1} \int |f|^p + q^{-1} \int |g|^q = p^{-1} + q^{-1} = 1.$$

* * *

Note that $1/p + 1/q = 1$ can be written as

$$(p-1)(q-1) = 1$$

or in the more general case $(p/r - 1)(q/r - 1) = 1$ or $(p - r)(q - r) = r^2$.

* * *

The inequality (3.1) can be proven via simple calculus or from (see [3, p. 26]):

$$uv \leqslant \int_0^u x^{p-1}\,\mathrm{d}x + \int_0^v y^{q-1}\,\mathrm{d}y = \frac{u^p}{p} + \frac{v^q}{q}.$$

This is because uv is the area of the rectangle with sides u and v on x- and y- axes respectively, $\int_0^u x^{p-1}\,\mathrm{d}x$ is the area under the graph of x^{p-1} up to u and $\int_0^v y^{q-1}\,\mathrm{d}y$ is the area to the left of said graph up to v. These regions always cover the rectangle with equality when x^{p-1} passes through (u, v), that is $v = u^{p-1}$.

3.1.3 Minkowski

For $p > 1$ (the case $p = \infty$ is easy),

$$\|f+g\|_p \leqslant \|f\|_p + \|g\|_p .$$

Firstly,

$$\int |f+g|^p = \int |f+g|\,|f+g|^{p-1} \leqslant \int |f|\,|f+g|^{p-1} + \int |g|\,|f+g|^{p-1} . \tag{3.2}$$

The integral $\int |f|\,|f+g|^{p-1}$ can be interpreted as the 1-norm of a product, so applying Hölder we get

$$\int |f|\,|f+g|^{p-1} \leqslant \|f\|_p \left(\|f+g\|_p^p\right)^{1/q} .$$

With the analogous inequality for the other part of (3.2) we get

$$\int |f+g|^p \leqslant \|f+g\|_p^{p/q} \left[\|f\|_p + \|g\|_p\right] .$$

Dividing we get the required result ($p - p/q = 1$).

3.1.4 Markov/Chebishev-type inequalities

The variant is from [1, p. 29]. For a function $\varphi\colon \mathbb{R} \to \mathbb{R}_{\geqslant 0}$, A measurable and denoting $m_A = \inf\{\varphi(y) : y \in A\}$,

$$m_A \mathbb{P}(X \in A) \leqslant \mathbb{E}(\varphi(X);\, X \in A) \leqslant \mathbb{E}\varphi(X).$$

The **proof** is from taking expectations in

$$m_A 1_A(X) \leqslant \varphi(X) 1_A(X) \leqslant \varphi(X).$$

* * *

For $X \geqslant 0$, $A = [a, \infty\rangle$ for some $a \geqslant 0$ and $\varphi \equiv 1$ we get Markov's inequality

$$\mathbb{P}(X \geqslant a) \leqslant \frac{\mathbb{E}X}{a}.$$

If X is not nonnegative, we need to take $\varphi(x) = |x|$.

For $A = [a, \infty\rangle$ for some $a \geqslant 0$ and $\varphi(x) = (x - \mathbb{E}X)^2$ we get Chebishev's inequality

$$\mathbb{P}(|X - \mathbb{E}X| \geqslant a) \leqslant \frac{\operatorname{Var} X}{a^2}.$$

The variant with $\varphi(x) = x^2$ is also legitimate:

$$\mathbb{P}(|X| \geqslant a) \leqslant \frac{\mathbb{E}X^2}{a^2}.$$

3.1.5 Tails of the normal distribution

The following is direct (Nourdin, p. 39):

$$\begin{aligned}\left(\frac{1}{x} - \frac{1}{x^3}\right) e^{-x^2/2} &= \int_x^\infty e^{-y^2/2}\left(1 - \frac{3}{y^4}\right) \mathrm{d}y \\ &\leqslant \int_x^\infty e^{-y^2/2}\,\mathrm{d}y \leqslant \frac{1}{x}\int_x^\infty y e^{-y^2/2}\,\mathrm{d}y = \frac{1}{x} e^{-x^2/2}.\end{aligned}$$

3.2 Borel–Cantelli lemma

For events A_j,

$$\sum_{j\geqslant 1}\mathbb{P}(A_j)<\infty \implies \mathbb{P}(A_j \text{ i.o.})=0.$$

Note that

$$\{A_j \text{ i.o.}\}=\left\{\sum_{j\geqslant 1}1_{A_j}=\infty\right\},$$

$$\{A_j \text{ ult.}\}=\left\{\sum_{j\geqslant 1}1_{A_j^c}<\infty\right\}.$$

If $\sum_j \mathbb{P}(A_j)<\infty$ then

$$\sum_{j\geqslant 1}P(A_j)=\sum_{j\geqslant 1}\mathbb{E}\left(1_{A_j}\right)=\mathbb{E}\left(\sum_{j\geqslant 1}1_{A_j}\right)$$

is $<\infty$, so in particular $\mathbb{P}\left(\sum_j 1_{A_j}=\infty\right)=0$.

* * *

The **converse** holds when A_j are **independent**. Note

$$\{A_j \text{ i.o.}\}^c=\left(\bigcap_{n\geqslant 1}\bigcup_{k\geqslant n}A_k\right)^c=\bigcup_{n\geqslant 1}\bigcap_{k\geqslant n}A_k^c.$$

Then

$$\begin{aligned}1&=\mathbb{P}(\text{not } A_j \text{ i.o.})\\&=\lim_{n\to\infty}\mathbb{P}\left(\bigcap_{k\geqslant n}A_k^c\right)=\lim_{n\to\infty}\prod_{k\geqslant n}[1-\mathbb{P}(A_k)]\\&\leqslant\lim_{n\to\infty}\exp\left(-\sum_{k\geqslant n}\mathbb{P}(A_k)\right),\end{aligned}$$

therefore $\lim_{n\to\infty}\sum_{k\geqslant n}\mathbb{P}(A_k)=0$ which is equivalent to $\sum_{n\geqslant 1}\mathbb{P}(A_n)<\infty$.

Note also that we used $1-x\leqslant e^{-x}$ for all x.

3.3 Convergence of random variables

3.3.1 Characterizing weak convergence I (expectation version)

It holds that $X_n\xrightarrow{\text{d}}X$, meaning $F_n(x)\to F(x)$ for all x such that F is continuous at x, **if and only if** $\mathbb{E}g(X_n)\to\mathbb{E}g(X)$ for all **continuous and bounded** functions g.

First **assume weak convergence**. There exist variables Y_n, Y with distributions F_n, F such that $Y_n\xrightarrow{\text{a.s.}}Y$ (see below). Then also $g(Y_n)\xrightarrow{\text{a.s.}}g(Y)$ and

$$\mathbb{E}g(X_n)=\mathbb{E}g(Y_n)\to\mathbb{E}g(Y)=\mathbb{E}g(X)$$

using the **bounded convergence** theorem.

Now the **converse**. We want $\mathbb{E}g(X_n)\to\mathbb{E}g(X)$ where $g=1_{\langle\infty,x]}$, but these functions are not continuous. The idea then is to approximate them by continuous g's.

Fix an $x\in\mathbb{R}$ and for all $r>0$ define

- g_{r+} as 1 up to x, 0 from $x+1/r$ onwards and linear inbetween;
- g_{r-} as 1 up to $x-1/r$, 0 from x onwards and linear inbetween.

Then we have

$$\limsup_{n\to\infty} \mathbb{P}(X_n \leqslant x) \leqslant \lim_{n\to\infty} \mathbb{E}g_{r+}(X_n) = \mathbb{E}g_{r+}(X) \to \mathbb{P}(X \leqslant x).$$

The lattermost convergence holds for all x due to

$$\mathbb{P}(X \leqslant x) \leqslant \mathbb{E}g_{r+}(X) \leqslant P(X \leqslant x + 1/r)$$

and the right-continuity of F.

The $\liminf$ side is obtained using g_{r-}:

$$\liminf_{n\to\infty} \mathbb{P}(X_n \leqslant x) \geqslant \lim_{n\to\infty} \mathbb{E}g_{r-}(X_n) = \mathbb{E}g_{r-}(X) \to \mathbb{P}(X \leqslant x),$$

with the lattermost convergence being true for $x \notin \operatorname{discont} F$, since only then does

$$\mathbb{P}(X \leqslant x - 1/r) \leqslant \mathbb{E}g_{r-}(X) \leqslant \mathbb{P}(X \leqslant x)$$

yield the desired sandwiching.

$*$ $*$ $*$

Earlier we used the following **lemma**: if $F_n \xrightarrow{\text{w}} F$ there exist variables with those distributions such that $Y_n \xrightarrow{\text{a.s.}} Y_n$. The common probability space is defined as usual

$$\Omega = [0,1]\,, \quad \mathcal{F} = \text{Borel sets}, \quad \mathbb{P} = \text{Lebesgue measure on } [0,1]\,,$$

and the variables by

$$Y_n(\omega) = \sup\{x \in \mathbb{R} \colon F_n(x) < \omega\}\,, \quad \omega \in [0,1]\,.$$

For proof that $Y_n \xrightarrow{\text{a.s.}} Y$ see [1, p. 118].

3.3.2 Characterizing weak convergence II (portmanteau theorem)

3.3.3 Characterizing convergence in probability

3.3.4 Continuous mapping theorems

Assume $X_n \xrightarrow{\text{a.s.}} X$ with g a measurable function such that $\mathbb{P}(X \in \operatorname{discont} g) = 0$. **Then** $g(X_n) \xrightarrow{\text{a.s.}} g(X)$. This simply follows from

$$\{g(X_n) \to g(X)\} \supseteq \{X_n \to X\} \cap \{X \notin \operatorname{discont} g\}$$

$*$ $*$ $*$

Assume $X_n \xrightarrow{\text{d}} X$ with g a measurable function such that $\mathbb{P}(X \in \operatorname{discont} g) = 0$. **Then** $g(X_n) \xrightarrow{\text{d}} g(X)$. In particular and if g is bounded, $\mathbb{E}g(X_n) \to \mathbb{E}g(X)$.

Again let $Y_n \xrightarrow{\text{a.s.}} Y$. For any bounded continous f is $f \circ g$ also bounded and $\operatorname{discont}(f \circ g) \subseteq \operatorname{discont} g$. Thus $f(g(Y_n)) \xrightarrow{\text{a.s.}} f(g(Y))$ and $\mathbb{E} f(g(Y_n)) \to \mathbb{E} f(g(Y))$ follows by the bounded convergence theorem.

$$* \stackrel{*}{} *$$

Assume $X_n \xrightarrow{\mathbb{P}} X$ with g a measurable function such that $\mathbb{P}(X \in \operatorname{discont} g) = 0$. **Then** $g(X_n) \xrightarrow{\mathbb{P}} g(X)$.

See [3, p. 103]. Fix a subsequence $N' \subseteq \mathbb{N}$. The convergence $X_n \xrightarrow{\mathbb{P}} X$ implies the convergence $X_n \xrightarrow{\text{a.s.}} X$ along some $N'' \subseteq N'$. Then also $g(X_n) \xrightarrow{\text{a.s.}} g(X)$ along N'' and $g(X_n) \xrightarrow{\text{a.s.}} g(X)$ due to the uniqueness of the a.s.-limit. Since N' was arbitrary, it follows that $g(X_n) \xrightarrow{\mathbb{P}} g(X)$.

3.3.5 a.s. implies $\mathbb{P}$

For any $\varepsilon > 0$,

$$\begin{aligned} \{X_n \not\to X\} &\supseteq \{|X_n - X| \geqslant \varepsilon \text{ i.o.}\} \\ &\implies \mathbb{P}\left(|X_n - X| \geqslant \varepsilon \text{ i.o.}\right) = 0 \\ &\implies \sum_{n=1}^{\infty} \mathbb{P}(|X_n - X| \geqslant \varepsilon) < \infty \\ &\implies \lim_{n\to\infty} \mathbb{P}(|X_n - X| \geqslant \varepsilon) = 0. \end{aligned}$$

3.3.6 L^p implies $\mathbb{P}$

We use a Markov-type inequality

$$\mathbb{P}(|X_n - X| \geqslant \varepsilon) \leqslant \frac{\mathbb{E}\,|X_n - X|^p}{\varepsilon^p} \to 0.$$

$$* \stackrel{*}{} *$$

The **converse** will hold if $X_n - X$ is bounded by some $M > 0$, which is specifically true if both X and all X_n are bounded. Then

$$\mathbb{E}\,|X_n - X|^p = \mathbb{E}\left(|X_n - X|^p \wedge M^p\right) \leqslant \varepsilon \mathbb{P}(|X_n - X| < \varepsilon) + M^p \mathbb{P}(|X_n - X| \geqslant \varepsilon)$$

which goes to 0 as $\varepsilon \downarrow 0$.

3.3.7 $\mathbb{P}$ implies d

3.3.8 Cauchyness in probability and in L^p

3.4 Uniform integrability

3.4.1 Definition and characterization

See [3, p. 106]. A family $(X_j)_{j\in J}$ of random variables is uniformly integrable if

$$\lim_{r\to\infty} \sup_{j\in J} \mathbb{E}\left(|X_j|\,;\,|X_j| > r\right) = 0.$$

If it's a sequence $(X_n)_{n\in\mathbb{N}}$ of **integrable** variables, then

$$\lim_{r\to\infty} \limsup_{n\to\infty} \mathbb{E}\left(|X_n|\,;\,|X_n|>r\right)=0$$

suffices, because

$$\lim_{r\to\infty}\left[\limsup_{n\to\infty} \mathbb{E}\left(|X_n|\,;\,|X_n|>r\right)-\sup_{n\in\mathbb{N}} \mathbb{E}\left(|X_n|\,;\,|X_n|>r\right)\right]=0,$$

since both parts themselves tend to 0.

$$* \overset{*}{} *$$

For uniform integrability, boundedness in L^p for some $p>1$ (meaning $\sup_j \mathbb{E}\,|X_j|^p<\infty$) **suffices**. We use the Markov-type inequality: $\mathbb{P}(|X_j|\geqslant x)\leqslant x^{-p}\mathbb{E}\,|X_j|^p$ to get

$$\begin{aligned}\mathbb{E}(|X_j|\,;\,|X_j|>r) &= \int_0^\infty \mathbb{P}(|X_j|\cdot 1_{\{|X_j|>r\}}\geqslant x)\,\mathrm{d}x=\int_r^\infty \mathbb{P}(|X_j|\geqslant x)\,\mathrm{d}x\\ &\leqslant \mathbb{E}\,|X_j|^p\int_r^\infty x^{-p}\,\mathrm{d}x=\frac{1}{p-1}r^{1-p}\mathbb{E}\,|X_j|^p\,.\end{aligned}$$

for all $j\in J$ and $r>0$, though [3] says something different.

$$* \overset{*}{} *$$

Now the main **characterization** of uniform integrability: $(X_j)_j$ are uniformly integrable if and only if

$$\sup_{j\in J}\mathbb{E}\,|X_j|<\infty \quad\text{and}\quad \lim_{\mathbb{P}(A)\to 0}\sup_{j\in J}\mathbb{E}(|X_j|\,;\,A)=0.$$

Only if. Set arbitrary $\varepsilon>0$ and choose r_0 so that $\sup_j \mathbb{E}(|X_j|\,;\,|X_j|>r)<\varepsilon$ for $r\geqslant r_0$.

$$\begin{aligned}\mathbb{E}\,|X_j| &= \mathbb{E}(|X_j|\,;\,|X_j|\leqslant r_0)+\mathbb{E}(|X_j|\,;\,|X_j|>r_0)\\ &\leqslant r_0+\mathbb{E}(|X_j|\,;\,|X_j|>r_0)\leqslant r_0+\varepsilon\end{aligned}$$

This proves the first part.

More generally for any $r>0$,

$$\begin{aligned}\mathbb{E}(|X_j|\,;\,A) &= \mathbb{E}(|X_j|\,;\,|X_j|\leqslant r, A)+\mathbb{E}(|X_j|\,;\,|X_j|>r, A)\\ &\leqslant r\mathbb{P}(A)+\mathbb{E}(|X_j|\,;\,|X_j|>r)\to 0\end{aligned}$$

as $\mathbb{P}(A)\to 0$ and $r\to\infty$ and uniformly in j. This proves the second part.

If. Define $A_{j,r}=\{|X_j|>r\}$. Then the probabilities of these events are bounded uniformly in j by Markov's inequality:

$$\sup_j \mathbb{P}(A_{j,r})\leqslant r^{-1}\sup_j \mathbb{E}\,|X_j|\,.$$

For any $\varepsilon>0$ and large enough r, the assumption implies

$$\sup_{j,k\in J}\mathbb{E}(|X_j|\,;\,A_{k,r})<\varepsilon,$$

which in particular gives the desired uniform integrability by taking only $k=r$.

$$* \overset{*}{} *$$

A similar technique can be used to prove the following: if X is integrable, then

$$\lim_{\mathbb{P}(A)\to 0} \mathbb{E}(|X|\,;\,A) = 0.$$

We know that

$$\lim_{r\to\infty} \mathbb{E}(|X|\,;\,|X|>r) = 0$$

because $X \cdot 1_{\{X>r\}} \to 0$ and dominated convergence. The "rest of A" is not a problem since

$$\mathbb{E}(|X|\,;\,A) \leqslant r\mathbb{P}(A) + \mathbb{E}(|X|\,;\,|X|>r).$$

3.4.2 Convergence of means

For random variables $X, X_1, X_2, \ldots$ with $X_n \xrightarrow{\mathrm{d}} X$ we have

$$\mathbb{E}X_n \to \mathbb{E}X < \infty \iff X_n \text{ are uniformly integrable.}$$

$\boxed{\Leftarrow}$ To get $\mathbb{E}X < \infty$, we switch to **bounded** functions. For all $r > 0$,

$$\liminf_{n\to\infty} \mathbb{E}X_n \geqslant \lim_{n\to\infty} \mathbb{E}\left(X_n \wedge r\right) = \mathbb{E}\left(X \wedge r\right)$$

so with $r \to \infty$ we have $\mathbb{E}X \leqslant \liminf_n \mathbb{E}X_n < \infty$.

The other part follows from

$$|\mathbb{E}X_n - \mathbb{E}X| \leqslant |\mathbb{E}X_n - \mathbb{E}\left(X_n \wedge r\right)| + |\mathbb{E}\left(X_n \wedge r\right) - \mathbb{E}\left(X \wedge r\right)| + |\mathbb{E}\left(X \wedge r\right) - \mathbb{E}X|$$

for all $r > 0$, wherein we first let $n \to \infty$ and then $r \to \infty$, the uniformity being key.

$\boxed{\Rightarrow}$ If X is **continuous**, then $\mathbb{E}(X_n;\, X_n > r) \to \mathbb{E}(X;\, X > r)$ by the continuous mapping theorem.

More generally, we need to replace $x \mapsto x \cdot 1(x > r)$ with a $\geqslant$ continuous function; [3] offers

$$g(x) = x - [x \wedge (r-x)_+]$$

which is more nicely written as $g(x) = 0$ if $x \leqslant r/2$; $2x - r$ if $r/2 < x \leqslant r$ and x otherwise. Then

$$\mathbb{E}\left(X_n;\, X_n > r\right) \leqslant \mathbb{E}g(X_n) \xrightarrow{n} \mathbb{E}g(X) \xrightarrow{r} 0$$

with, again, obvious uniformity in n.

3.5 Independence and 0-1 laws

3.6 Notable exercises

3.6.1 Asymptotics of tail-expectations

See [1, ex. 1.6.14]. Let $X \geqslant 0$ (without assuming $E(1/X) < \infty$). Then

$$\lim_{y\to\infty} y\mathbb{E}\left(\frac{1}{X};\, X > y\right) = 0, \quad \lim_{y\downarrow 0} y\mathbb{E}\left(\frac{1}{X};\, X > y\right) = 0.$$

For the **first** part,

$$\mathbb{E}\left(\frac{y}{X};\, X > y\right) \leqslant \mathbb{E}(1;\, X > y) = \mathbb{P}(X > y) \to 0.$$

For the **second** part, denoting $n = 1/y$ and replacing X with $1/X$, the limit becomes $\lim_{n\to\infty} \frac{1}{n}\mathbb{E}(X; X < n)$ and we have

$$\frac{1}{n}\mathbb{E}(X;\, X \leqslant n) = \frac{1}{n}\int_0^n x \,\mathrm{d}F = \frac{1}{n}\int_0^{\sqrt{n}} x \,\mathrm{d}F + \frac{1}{n}\int_{\sqrt{n}}^n x \,\mathrm{d}F \to 0,$$

because

$$\int_0^{\sqrt{n}} x \,\mathrm{d}F \leqslant \sqrt{n}, \qquad \int_{\sqrt{n}}^n x \,\mathrm{d}F \leqslant n \left[F(n) - F(\sqrt{n})\right].$$

3.6.2 For $p < q$, L^q-convergence implies L^p-convergence

See [1, ex. 1.6.11]. Let $0 < p < q$ and $\mathbb{E}\,|X|^q < \infty$. Then

$$\mathbb{E}\,|X|^p \leqslant (\mathbb{E}\,|X|^q)^{p/q}.$$

In particular,

- the p-norm is **smaller**; by taking the root we get $\|X\|_p \leqslant \|X\|_q$,
- the p-norm is **finite**, so if X is q-integrable it is p-integrable,
- **convergence** in L^q implies convergence in L^p.

To **prove** it, note that

$$\mathbb{E}\,|X|^p = \mathbb{E}\left[(|X|^q)^{p/q}\right] \leqslant (\mathbb{E}\,|X|^q)^{p/q}$$

where the inequality is Jensen for concave $x \mapsto x^{p/q}$.

3.6.3 Integrating with respect to Lebesgue–Stieltjes measures

For F, G corresponding to μ, ν measures on $(\mathbb{R}, \mathcal{B})$,

(a) $\int_{\langle a,b]} [F(y) - F(a)]\, G(\mathrm{d}y) = (\mu \otimes \nu)\,\{(x,y)\colon a < x \leqslant y < b\}$,

(b) $\displaystyle\int_{\langle a,b]} F(y)G(\mathrm{d}y) + \int_{\langle a,b]} G(y)F(\mathrm{d}y) = F(b)G(b) - F(a)G(a) + \sum_{x\in\langle a,b]} \mu(\{x\})\nu(\{x\}).$

(c) In particular, if $F = G$ is continuous,

$$\int_{\langle a,b]} 2F(y)F(\mathrm{d}y) = F(b)^2 - F(a)^2.$$

Firstly, denote $D_{a,b} = \{(x,y)\colon a < x \leqslant y \leqslant b\}$. Then,

$$\begin{aligned}(\mu\otimes\nu)(D_{a,b}) &= \int \mu(D^y_{a,b})\nu(\mathrm{d}y)\\ &= \int_{\langle a,b]} \mu(\langle a,y])\nu(\mathrm{d}y)\\ &= \int_{\langle a,b]} [F(b) - F(a)]\, G(\mathrm{d}y).\end{aligned}$$

Secondly,

$$\begin{aligned}\int_{\langle a,b]} F(y)G(\mathrm{d}y) &= \int_{\langle a,b]} \left[F(y) - F(a) + F(a)\right] G(\mathrm{d}y) \\ &= \int_{\langle a,b]} \left[F(y) - F(a)\right] G(\mathrm{d}y) + F(a)\left[G(b) - G(a)\right] \\ &= (\mu \otimes \nu)(D_{a,b}) + F(a)\left[G(b) - G(a)\right].\end{aligned}$$

Analogously,

$$\int_{\langle a,b]} G(y)F(\mathrm{d}y) = (\mu \otimes \nu)(D'_{a,b}) + G(a)\left[F(b) - F(a)\right]$$

where $D'_{a,b} = \{(x,y) \colon a < y \leqslant x \leqslant b\}$. The final result follows from

$$\begin{aligned}(\mu \otimes \nu)(D_{a,b}) + (\mu \otimes \nu)(D'_{a,b}) &= (\mu \otimes \nu)(D_{a,b} \cup D'_{a,b}) + (\mu \otimes \nu)(D_{a,b} \cap D'_{a,b}) \\ &= \left[F(b) - F(a)\right]\left[G(b) - G(a)\right] + \sum_{x \in \langle a,b]} \mu(\{x\})\nu(\{x\}).\end{aligned}$$

References

[1] R. Durrett. **Probability: Theory and Examples**. 5th ed. Cambridge University Press, Apr. 2019. ISBN: 9781108591034 9781108473682. DOI: 10.1017/9781108591034. URL: https://www.cambridge.org/core/product/identifier/9781108591034/type/book (visited on 10/16/2025).

[2] G. B. Folland. **Real analysis: modern techniques and their applications**. eng. 2. ed. A Wiley-Interscience publication. New York Weinheim: Wiley, 1999. ISBN: 9780471317166.

[3] O. Kallenberg. **Foundations of Modern Probability**. en. Vol. 99. Probability Theory and Stochastic Modelling. Cham: Springer International Publishing, 2021. ISBN: 9783030618704 9783030618711. DOI: 10.1007/978-3-030-61871-1. URL: https://link.springer.com/10.1007/978-3-030-61871-1 (visited on 10/16/2025).