

Probability

Notes taken by Runqiu Ye

Lectures by Konstantin Tikhomirov

Carnegie Mellon University

Spring 2026

Contents

1	Measure theory review	3
1.1	Measurable space and mapping	3
1.2	Measure space	4
1.3	π - λ theorem	5
1.4	Extension theorems	5
1.5	Lebesgue Integration	6
1.6	Product measures and Fubini theorem	9
2	Probability theory basics	11
2.1	Distributions and densities	11
2.2	Independence	13
2.3	Convolution	15
2.4	Moments	16

1 Measure theory review

1.1 Measurable space and mapping

Definition (σ -field). A collection of subsets $\Sigma \subset 2^\Omega$ is a σ -field if

- $\emptyset \in \Sigma$.
- If $A \in \Sigma$, then $A^c \in \Sigma$.
- If $\{A_i\}_{i=1}^\infty \subset \Sigma$, then $\bigcup_{i=1}^\infty A_i \in \Sigma$.

The pair (Ω, Σ) is called a measurable space.

Definition (atom). Let Σ be a σ -field. Say $A \in \Sigma$ is an atom if for all $B \in \Sigma$ either $A \subset B$ or $A \cap B = \emptyset$.

Proposition. For all $\omega \in \Omega$, there exists atom $A \in \Sigma$ containing ω if Ω is finite or countable.

Proof. Define $\tilde{A} = \bigcap \{B \in \Sigma : \omega \in B\}$. We can check that $\tilde{A} \in \Sigma$ and \tilde{A} is an atom containing ω . \square

Corollary. If Ω is finite or countable, there exists a partition $\Omega = \bigsqcup_i \Omega_i$, where each Ω_i is an atom of Σ . With this partition, Σ is just the power set with respect to $\{\Omega_i\}_i$.

Definition. If $F \subset 2^\Omega$, then the σ -field generated by F is the smallest σ -field containing all elements of F . Write this σ -field as $\sigma(F)$.

Example. Let $\Omega = \{1, 2, 3, 4, 5\}$ and $F = \{\{2, 3\}, \{3, 4\}\}$. Construct σ -field Σ generated by F . Σ is all possible union of sets from the collection $\{\{2\}, \{3\}, \{4\}, \{1, 5\}\}$.

Definition (measurable mapping). Given two measurable spaces (Ω, Σ) and $(\tilde{\Omega}, \tilde{\Sigma})$. Then $f : \Omega \rightarrow \tilde{\Omega}$ is measurable if $f^{-1}(B) \in \Sigma$ for all $B \in \tilde{\Sigma}$.

Definition (Borel σ -field). Let (T, τ) be a topological space. Then the Borel σ -field $\mathcal{B}(T, \tau)$ is defined as the smallest σ -field containing all open sets.

Definition (product measurable space). Given two measurable spaces (Ω, Σ) and $(\tilde{\Omega}, \tilde{\Sigma})$. We can define the product measurable space as follows: let the ground set be $\Omega \times \tilde{\Omega}$, and let $\Sigma \otimes \tilde{\Sigma}$ be the smallest σ -field containing all rectangles $B \times \tilde{B}$ where $B \in \Sigma$ and $\tilde{B} \in \tilde{\Sigma}$.

More generally, let Λ be an index set and $(\Omega_\lambda, \Sigma_\lambda)_{\lambda \in \Lambda}$. Define the product σ -field $\bigotimes_{\lambda \in \Lambda} \Sigma_\lambda$ be the smallest σ -field containing all elements in the form of $\prod_{\lambda \in \Lambda} B_\lambda$ where $B_\lambda \in \Sigma_\lambda$ and $B_\lambda = \Omega_\lambda$ for all but countably many indices.

Proposition. Let $(\Omega_i, \Sigma_i)_{i=1}^n$ be measurable spaces and $(\prod_{i=1}^n \Omega_i, \bigotimes_{i=1}^n \Sigma_i)$ be the product space. Let (Ω, Σ) be the domain and $f = (f_1, \dots, f_n) : (\Omega, \Sigma) \rightarrow (\prod_{i=1}^n \Omega_i, \bigotimes_{i=1}^n \Sigma_i)$. Suppose f is measurable, then every coordinate projection $f_i : \Omega \rightarrow \Omega_i$ is measurable.

This is also true for arbitrary index set.

Proposition. If $f : (\Omega, \Sigma) \rightarrow (\Omega_f, \Sigma_f)$ and $g : (\Omega, \Sigma) \rightarrow (\Omega_g, \Sigma_g)$, then the concatenation (f, g) is measurable w.r.t. the product space $(\Omega_f \times \Omega_g, \Sigma_f \otimes \Sigma_g)$.

Proof. Let $A \times B$ be such that $A \in \Sigma_f$ and $B \in \Sigma_g$. Then the preimage

$$(f, g)^{-1}(A \times B) = f^{-1}(A) \cap g^{-1}(B) \in \Sigma.$$

By definition, the product σ -field is generated by rectangles, so the proof is complete. \square

1.2 Measure space

Definition (measure). Let (Ω, Σ) be a measurable space. Then $\mu : \Sigma \rightarrow [0, \infty]$ is a measure if

- $\mu(\emptyset) = 0$.
- If $A_i \in \Sigma$ is pairwise disjoint then $\mu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$.

Proposition (continuity of measure). If $A_1 \subset A_2 \subset \dots$ is a nested sequence of elements of Σ and μ be any measure on (Ω, Σ) . Then

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{i \rightarrow \infty} \mu(A_i).$$

If $A_1 \supset A_2 \supset \dots$ is a nested sequence of elements of Σ and $\mu(A_n) < \infty$ for some n . Then

$$\mu\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{i \rightarrow \infty} \mu(A_i).$$

Definition. Let (Ω, Σ, μ) be a measure space.

Say μ is σ -finite if there is a representation $\Omega = \bigcup_{i=1}^{\infty} \Omega_i$ where $\Omega_i \in \Sigma$ and $\mu(\Omega_i) < \infty$.

Say μ is a probability measure if $\mu(\Omega) = 1$.

Definition (completion of measure space). Let (Ω, Σ, μ) be a measure space. Let

$$\tilde{\Sigma} = \{A \cup B : A \in \Sigma, B \subset \Omega, \text{there exists } C \in \Sigma \text{ with } \mu(C) = 0 \text{ and } B \subset C\}.$$

We can check $\tilde{\Sigma}$ is a σ -field. If $\tilde{\mu}$ is a measure on $(\Omega, \tilde{\Sigma})$ which agrees with μ on Σ , then $(\Omega, \tilde{\Sigma}, \tilde{\mu})$ is called a completion of (Ω, Σ, μ) .

1.3 π - λ theorem

Definition (π -system). Let Ω be a set and \mathcal{P} be a collection of subsets of Ω . Then \mathcal{P} is a π -system if it is closed with respect to taking finite intersections. That is, $A, B \in \mathcal{P}$ implies $A \cap B \in \mathcal{P}$.

Example. On the real line \mathbb{R} , both $\mathcal{P}_1 = \{(a, b) : a < b\}$ and $\mathcal{P}_2 = \{(-\infty, a] : a \in \mathbb{R}\}$ are π -systems.

Definition (λ -system). Let Ω be a set and \mathcal{L} be a collection of subsets of Ω . Say \mathcal{L} is a λ -system if

- $\emptyset \in \mathcal{L}$.
- $A \in \mathcal{L}$ implies $A^c \in \mathcal{L}$.
- for all countable collection of disjoint elements $A_i \in \mathcal{L}$, we have $\bigcup_{i=1}^{\infty} A_i \in \mathcal{L}$.

Theorem (π - λ theorem). Let Ω be a set, \mathcal{P} be a π -system and \mathcal{L} be a λ -system. Also suppose $\mathcal{P} \subset \mathcal{L}$, then $\sigma(\mathcal{P}) \subset \mathcal{L}$.

Proof. Let $\ell(\mathcal{P})$ be the smallest λ -system on Ω containing \mathcal{P} . The goal is to show that $\ell(\mathcal{P})$ is a σ -field. We need to show that if $A_i \in \ell(\mathcal{P})$ for $1 \leq i < \infty$, then $\bigcup_{i=1}^{\infty} A_i \in \ell(\mathcal{P})$. Note that

$$\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} \left(A_i \setminus \bigcup_{j=1}^{i-1} A_j \right),$$

so it suffices to show that $A, B \in \ell(\mathcal{P})$ implies $A \cap B \in \ell(\mathcal{P})$.

For $A \in \ell(\mathcal{P})$ we define

$$W_A = \{B \subset \Omega : A \cap B \in \ell(\mathcal{P})\}.$$

It can be directly verified that W_A is a λ -system.

Take $A \in \mathcal{P}$, then for any $B \in \mathcal{P}$ we have $A \cap B \in \mathcal{P} \subset \ell(\mathcal{P})$. Hence, $\mathcal{P} \subset W_A$ and thus $\ell(\mathcal{P}) \subset W_A$ for all $A \in \mathcal{P}$, as $\ell(\mathcal{P})$ is the smallest λ -system on Ω containing \mathcal{P} . Now take $A \in \ell(\mathcal{P})$, we have $A \in W_B$ for all $B \in \mathcal{P}$. It follows that $A \cap B \in \ell(\mathcal{P})$ and thus $B \in W_A$. Hence similarly $\ell(\mathcal{P}) \subset W_A$ for all $A \in \ell(\mathcal{P})$.

Now for any pair $B, C \in \ell(\mathcal{P})$, we have $C \in W_B$ and thus $B \cap C \in \ell(\mathcal{P})$. This completes the proof. \square

1.4 Extension theorems

Definition (semi-field). A collection of subsets $S \subset 2^\Omega$ is a semi-field if

- $\emptyset \in S$ and $\Omega \in S$.
- $A, B \in S$ implies $A \cap B \in S$.
- If $A \in S$, then A^c is a finite disjoint union of sets in S .

Theorem (Caratheodory's extension theorem). Let S be a semi-field and let μ be a non-negative function on S satisfying:

- $\mu(\emptyset) = 0$.
- If A_1, \dots, A_n are disjoint and $\bigcup_{i=1}^n A_i \in S$, then $\mu(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n \mu(A_i)$.
- If A_1, A_2, \dots are such that $\bigcup_{i=1}^\infty A_i \in S$, then $\mu(\bigcup_{i=1}^\infty A_i) \leq \sum_{i=1}^\infty \mu(A_i)$.

Then μ admits a unique extension $\bar{\mu}$ which is a measure on \bar{S} , the field (algebra) generated by S . Moreover, if $\bar{\mu}$ is σ -finite then $\bar{\mu}$ admits a unique extension $\tilde{\mu}$ to $\sigma(S)$.

Notation. Let T be any set. Write

$$\mathbb{R}^T = \{(\omega_t)_{t \in T} : \omega_t \in \mathbb{R}\}.$$

Also write \mathcal{R}^T as the σ -field generated by rectangles of the form $\prod_{t \in T} I_t$, where for each $t \in T$, I_t is either a semi-open interval of the form $(a, b]$ with $a < b$ or $I_t = \mathbb{R}$, and $I_t = \mathbb{R}$ for all but finitely many $t \in T$.

Theorem (Kolmogorov's extension theorem). For each finite non-empty subset $J \subset T$, let μ_J be a Borel probability measure in \mathbb{R}^J , and assume that the measures $(\mu_J)_{J \subset T, |J| < \infty}$ are compatible, in the sense that whenever $J_1 \subset J_2 \subset T$ with $0 \leq |J_1| \leq |J_2| < \infty$, $I_j \subset \mathbb{R}$ with $j \in J_1$ are Borel subsets of \mathbb{R} , and

$$\tilde{I}_j = \begin{cases} I_j & (j \in J_1) \\ \mathbb{R} & (j \in J_2 \setminus J_1), \end{cases}$$

one has

$$\mu_{J_2} \left(\prod_{j \in J_2} \tilde{I}_j \right) = \mu_{J_1} \left(\prod_{j \in J_1} I_j \right).$$

Then there exists a unique probability measure μ on $(\mathbb{R}^T, \mathcal{R}^T)$ consistent with $(\mu_J)_{J \subset T, |J| < \infty}$. That is, one has

$$\mu \left(\prod_{t \in T} I_t \right) = \mu_J \left(\prod_{j \in J} I_j \right)$$

whenever $J \subset T$ with $|J| < \infty$ and $I_t = \mathbb{R}$ for all $t \notin J$.

1.5 Lebesgue Integration

Here we provide a proof for dominated convergence theorem that uses the truncation technique, which will be a useful technique later in the course.

Theorem (dominated convergence theorem). Let $\{f_n\}_{n=1}^\infty$ be a sequence of measurable functions on (Ω, Σ, μ) and $g \geq 0$ be another measurable function. Suppose

1. $\int g d\mu < \infty$.
2. $|f_n|(\omega) \leq g(\omega)$ for all $\omega \in \Omega$ and $n \geq 1$.
3. $f_n \rightarrow f$ pointwise.

Then

$$\lim_{n \rightarrow \infty} \int f_n d\mu = \int f d\mu.$$

Proof. **Claim 1.** If h is a function on (Ω, Σ, μ) with $h \geq 0$ and $\int h d\mu < \infty$. Let $\{A_n\}_{n=1}^\infty$ be any sequence of elements of Σ with $\mu(A_n) \rightarrow 0$. Then

$$\int_{A_n} h d\mu \rightarrow 0.$$

Proof of claim. WLOG assume $\mu(A_n) \leq 2^{-n}$ for all n . Define $h_n = h \mathbb{1}_{\bigcup_{i=n}^\infty A_i}$. We then have

1. The sequence $\{h_n\}_{n=1}^\infty$ is monotone.
2. h_n converges to 0 almost everywhere.

Monotone convergence theorem then implies $\lim_{n \rightarrow \infty} \int h_n d\mu = 0$. Meanwhile,

$$0 \leq \int_{A_n} h d\mu \leq \int h_n d\mu,$$

and the proof is complete.

Claim 2. Suppose $h \geq 0$ and $\int h d\mu < \infty$. Let $\{\varepsilon_n\}_{n=1}^\infty$ be a sequence of strictly positive numbers converging to zero. Define

$$B_n = \{\omega \in \Omega : h(\omega) \leq \varepsilon_n\} \in \Sigma.$$

Then

$$\int_{B_n} h d\mu \rightarrow 0.$$

Proof of this claim is left as an exercise.

Now we prove the theorem. Fix $\varepsilon > 0$. By the previous two claims, there exists $M > 0$ and $\delta > 0$ such that

$$\int_{\{g \geq M\}} g d\mu < \varepsilon, \quad \int_{\{g \leq \delta\}} g d\mu < \varepsilon.$$

Let $U = \{\omega : \delta < g(\omega) < M\}$. Since g is integrable, $\mu(U) < \infty$. For $\omega \in U$, let $n_\varepsilon(\omega)$ be the smallest index such that $n \geq n_\varepsilon(\omega)$ implies $|f_n(\omega) - f(\omega)| \leq \varepsilon \mu(U)^{-1}$. It follows that there exists N such that

$$\mu(\{\omega \in U : n_\varepsilon(\omega) > N\}) \leq \frac{\varepsilon}{M}.$$

Then, for $n \geq N$, we have

$$\left| \int_U (f_n - f) d\mu \right| \leq \int_{n_\varepsilon(\omega) \leq N} |f_n - f| d\mu + \int_{n_\varepsilon(\omega) > N} |f_n - f| d\mu \leq 3\varepsilon.$$

Now for $n \geq N$, we have

$$\left| \int (f_n - f) d\mu \right| \leq 3\varepsilon + \int_{U^c} |f - f_n| d\mu \leq 3\varepsilon + 2 \int_{U^c} g d\mu \leq 7\varepsilon.$$

□

Theorem (Markov-Chebyshev inequality). Suppose we have probability measure space $(\Omega, \Sigma, \mathbb{P})$ and $f \geq 0$. Suppose also $\int f d\mathbb{P} < \infty$. Then

$$\mathbb{P}(\{\omega : f(\omega) > t\}) \leq \frac{1}{t} \int f d\mathbb{P}.$$

for all $t > 0$.

Remark. Let $1 \leq p < \infty$. Suppose $f : (\Omega, \Sigma, \mathbb{P}) \rightarrow [0, \infty]$ and $\int f^p d\mathbb{P} < \infty$. Then

$$\mathbb{P}(\{\omega : f(\omega) > t\}) \leq \frac{1}{t^p} \int f^p d\mathbb{P}.$$

for all $t > 0$.

Remark. Suppose $\int e^{\lambda f} d\mathbb{P} < \infty$ for all $\lambda \in \mathbb{R}$ and $f : (\Omega, \Sigma, \mathbb{P}) \rightarrow [0, \infty]$. Then

$$\mathbb{P}(\{\omega : f(\omega) > t\}) \leq \frac{1}{e^{\lambda t}} \int e^{\lambda f} d\mathbb{P}$$

for all $t > 0$ and $\lambda > 0$.

Theorem (Hölder inequality). Let $p, q \in [1, \infty]$ and $p^{-1} + q^{-1} = 1$. Let (Ω, Σ, μ) be a probability space. For any measurable functions f, g , we have

$$\int |fg| d\mathbb{P} \leq \left(\int |f|^p d\mathbb{P} \right)^{1/p} \left(\int |g|^q d\mathbb{P} \right)^{1/q}.$$

Theorem (Jensen's inequality). Let (Ω, Σ, μ) be a probability space and f be integrable. Let $\varphi : \overline{\mathbb{R}} \rightarrow \overline{\mathbb{R}}$ be convex and suppose $\varphi(\infty) = \lim_{x \rightarrow \infty} \varphi(x)$ and $\varphi(-\infty) = \lim_{x \rightarrow -\infty} \varphi(x)$. Then

$$\varphi \left(\int f d\mathbb{P} \right) \leq \int \varphi(f) d\mathbb{P}.$$

1.6 Product measures and Fubini theorem

Let $(\Omega_1, \Sigma_1, \mu_1), (\Omega_2, \Sigma_2, \mu_2)$ be σ -finite measure spaces. We already defined the product $\Sigma_1 \otimes \Sigma_2$. To define a product measure, we first consider the algebra of rectangles

$$S = \{A \in \Sigma_1 \otimes \Sigma_2 : A = A_1 \times A_2 \text{ for some } A_1 \in \Sigma_1, A_2 \in \Sigma_2\}.$$

Then we can define $\mu = \mu_1 \times \mu_2$ on S by

$$\mu(A) = \mu_1(A_1)\mu_2(A_2)$$

for $A = A_1 \times A_2$. We can check that the definition is self-consistent. That is, if $A = A_1 \times A_2$ is a countable union of disjoint rectangles $\{A_1^{(j)} \times A_2^{(j)}\}_{j=1}^\infty$, we have

$$\mu(A_1 \times A_2) = \sum_{j=1}^\infty \mu(A_1^{(j)} \times A_2^{(j)}).$$

This can be verified with monotone convergence theorem. Now μ is a premeasure and can be uniquely extended to $\Sigma_1 \otimes \Sigma_2$.

Theorem (Fubini-Tonelli). Let $(\Omega_1, \Sigma_1, \mu_1), (\Omega_2, \Sigma_2, \mu_2)$ be σ -finite measure spaces and let (Ω, Σ, μ) be the product space. Suppose f is measurable on the product space. Suppose either f is non-negative or $\int_\Omega |f| d\mu < \infty$. Then

- $y \mapsto f(x, y)$ is Σ_2 measurable for all $x \in \Omega_1$.
- $x \mapsto \int_{\Omega_2} f(x, y) d\mu_2(y)$ is Σ_1 measurable.
- We have

$$\int_{\Omega_1} \int_{\Omega_2} f(x, y) d\mu_2(y) d\mu_1(x) = \int_\Omega f(x, y) d\mu(x, y).$$

Proof. First suppose $f = \mathbb{1}_A$ for $A \in \Sigma$. Also suppose μ_1, μ_2 are finite. Define section

$$A_x = \{y \in \Omega_2 : (x, y) \in A\}.$$

The goal is to show that $A_x \in \Sigma_2$ for all $x \in \Omega_1$. Define a family of sets

$$\mathcal{F}_x = \{B \in \Sigma : B_x \text{ is } \Sigma_2\text{-measurable}\}.$$

It can be verified that \mathcal{F}_x is a σ -field for all $x \in \Omega_1$. Also, \mathcal{F}_x contains all rectangles and thus $\Sigma \subset \mathcal{F}_x$. Hence, we have shown that $y \mapsto \mathbb{1}_A(x, y) = \mathbb{1}_{A_x}(y)$ is measurable for all $x \in \Omega_1$.

Next we show $x \mapsto \mu_2(A_x)$ is measurable and its integral over Ω_1 is equal to $\mu(A)$. Define

$$\mathcal{U} = \left\{ B \in \Sigma : x \mapsto \mu_2(B_x) \text{ is } \Sigma_1\text{-measurable and } \int_{\Omega_1} \mu_2(B_x) d\mu_1 = \mu(B) \right\}$$

It can be verified that \mathcal{U} is a λ -system. Note that \mathcal{U} also contains all rectangles in Σ . It follows that $\mathcal{U} = \Sigma$ and the proof for indicator functions are complete.

Then use linearity to extend to simple functions, and use monotone convergence theorem to prove the statement for non-negative functions. For the case where f is integrable, consider the positive and negative part about f to complete the proof. \square

2 Probability theory basics

2.1 Distributions and densities

Definition. Let $F : \mathbb{R} \rightarrow [0, 1]$. Suppose F is

- right-continuous.
- non-decreasing.
- $\lim_{t \rightarrow -\infty} F(t) = 0$ and $\lim_{t \rightarrow \infty} F(t) = 1$.

Then F is a cumulative distribution function (CDF).

Remark. If we want to define CDF in \mathbb{R}^2 then the axioms are

- right-continuous: $F(\tilde{s}, \tilde{t}) \rightarrow F(s, t)$ as $t \downarrow \tilde{t}$ and $s \downarrow \tilde{s}$.
- coordinate-wise non-decreasing.
- $\lim_{s, t \rightarrow \infty} F(s, t) = 1$, $\lim_{s \rightarrow -\infty} F(s, t) = 0$ for any t , and $\lim_{t \rightarrow -\infty} F(s, t) = 0$ for any s .
- For a rectangle with bottom left vertex (a_1, a_2) and top right vertex (b_1, b_2) ,

$$F(b_1, b_2) - F(b_1, a_2) - F(a_1, b_2) + F(a_1, a_2) \geq 0.$$

Now we can connect the notion of CDF with randomness.

Suppose X -random real-valued variable on $(\Omega, \Sigma, \mathbb{P})$ that is almost everywhere finite. Define

$$F_X(t) = \mathbb{P}\{X(\omega) \leq t\}$$

for $-\infty < t < \infty$. It can be verified that F_X is a CDF.

Conversely, for any CDF F , there exists a probability space $(\Omega, \Sigma, \mathbb{P})$ and a real valued random variable on $(\Omega, \Sigma, \mathbb{P})$ with CDF F .

Definition. If X is random variable on $(\Omega, \Sigma, \mathbb{P})$ real valued and a.e. finite. Then we can define the induced Borel probability measure μ_X on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ by

$$\mu_X(B) := \mathbb{P}\{X \in B\}$$

for all $B \in \mathcal{B}_{\mathbb{R}}$.

Now suppose μ is any Borel probability measure on \mathbb{R} . Consider probability space $(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \mu)$ and formal identity mapping id on \mathbb{R} . Then $\mu_{\text{id}} \equiv \mu$.

Theorem. There is a one-to-one correspondence between the family of CDFs and the family of Borel probability measure on \mathbb{R} .

Proof. For any Borel probability measure μ , $F_\mu(t) = \mu((-\infty, t])$ is a valid CDF.

Conversely, for any CDF F , there exists unique probability measure μ_F on \mathbb{R} such that $\mu_F((-\infty, t]) = F(t)$ for all $-\infty < t < \infty$. This is a corollary of Caratheodory extension theorem. For detailed proof see notes or textbook. \square

Remark. Suppose $X = (X_1, X_2)$ is a random vector in \mathbb{R}^2 . We can define

$$F_X(s, t) = \mathbb{P}\{X_1 \leq s, X_2 \leq t\}.$$

Corresponding results are also true.

Definition (Probability mass function). Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space and $X : \Omega \rightarrow \mathbb{R}$ be random variable. Suppose there exists $S \subset \mathbb{R}$ countable such that $\mathbb{P}\{X \in S\} = 1$. We can define the probability mass function (PMF) f_X via

$$f_X(t) = \mathbb{P}\{X = t\}$$

for $t \in \mathbb{R}$. Due to the restriction, this gives complete description of the distribution, and we can construct CDF F_X via

$$F_X(t) = \sum_{s \leq t} f_X(s).$$

This sum makes sense since the $f_X(s) = 0$ for all but countably many s . Conversely, we can also reconstruct f_X from a CDF F_X .

Definition (Probability density function). Suppose F is a CDF which is absolutely continuous. That is, there exists Borel measurable non-negative function ρ on \mathbb{R} such that

$$F(t) = \int_{-\infty}^t \rho(s) ds$$

for all $-\infty < t < \infty$. This implies F is almost everywhere differentiable and the derivative is ρ . In this case, say ρ is the density function.

If random variable X is such that F_X is absolutely continuous, then the corresponding ρ_X is the probability density function for X .

Remark. Recall that a Borel σ -finite measure μ on the real line is absolutely continuous w.r.t the Lebesgue measure m on \mathbb{R} if $\mu(A) = 0$ whenever $A \in \mathcal{B}_{\mathbb{R}}$ is Lebesgue null. In this case, Randon-Nikodym theorem implies existence of non-negative Borel measurable function f such that $\mu(A) = \int_A f dm$.

Theorem. Suppose X RV on $(\Omega, \Sigma, \mathbb{P})$ is real-valued and a.e. finite. The following are equivalent:

1. F_X is absolutely continuous.
2. μ_X is absolutely containing w.r.t. Lebesgue measure.

Moreover, ρ_X is also the derivative of μ_X w.r.t. Lebesgue measure. That is, for any $A \in \mathcal{B}_{\mathbb{R}}$,

$$\mu_X(A) = \int_A \rho_X(t) dt.$$

2.2 Independence

Definition. Say two events $A, B \in \Sigma$ are independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

It is easy to verify that A, B are independent implies A^c, B are independent.

Remark. Suppose $\mathbb{P}(B) > 0$, then the conditional probability of A given B is defined as

$$\mathbb{P}(A | B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Then, independence of A and B is equivalent to $\mathbb{P}(A | B) = \mathbb{P}(A)$.

Definition. Let A_1, \dots, A_n be events. Say they are mutually independent if for any $\emptyset \neq I \subset [n]$, we have

$$\mathbb{P}\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} \mathbb{P}(A_i).$$

This is equivalent to saying that for every $2 \leq i \leq n$, the event A_i is independent from any event generated by A_1, \dots, A_{i-1} , or A_i is independent from $\sigma(A_1, \dots, A_{i-1})$.

Remark. The events A_1, \dots, A_n are called k -wise independent if any k -subset of the events are mutually independent. For $k < n$, this notion is strictly weaker than mutual independence of all n events. As an example, consider \mathbb{P} to be the uniform distribution on $\{1, \dots, 4\}$. Let $A_1 = \{1, 2\}$, $A_2 = \{1, 3\}$, and $A_3 = \{2, 3\}$. Then they are pairwise independent but not mutually independent.

Definition. A collection of events $\{A_\lambda\}_{\lambda \in \Lambda}$ on $(\Omega, \Sigma, \mathbb{P})$ are mutually independent if any finite subset of events are mutually independent.

Definition. Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space. Two σ -subfields are independent if for any $A \in \Sigma_1$ and $B \in \Sigma_2$, A, B are independent.

Definition. Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space and X, Y be two real-valued random variables. Say X and Y are independent if

$$\mathbb{P}\{X \in A, Y \in B\} = \mathbb{P}\{X \in A\} \mathbb{P}\{Y \in B\}$$

for any $A, B \in \mathcal{B}_{\mathbb{R}}$.

Equivalently, let Σ_X, Σ_Y be the σ -field generated by X and Y . Then independence of X and Y is equivalent to independence of Σ_X and Σ_Y .

Now we explore how this connect with product structure.

Proposition. Let $(\Omega_1, \Sigma_1, \mathbb{P}_1)$ and $(\Omega_2, \Sigma_2, \mathbb{P}_2)$ be two probability spaces and let $(\Omega, \Sigma, \mathbb{P})$ be the product space. Let X and Y be two random variables on $(\Omega, \Sigma, \mathbb{P})$. Suppose there exists some measurable functions such that $X(\omega_1, \omega_2) = g(\omega_1)$, and $Y(\omega_1, \omega_2) = h(\omega_2)$. Then X and Y are independent.

Proof. Let $A, B \in \mathcal{B}_{\mathbb{R}}$. Then

$$\begin{aligned} \mathbb{P}\{X \in A, Y \in B\} &= \mathbb{P}\{(\omega_1, \omega_2) : \omega_1 \in g^{-1}(A), \omega_2 \in h^{-1}(B)\} \\ &= \mathbb{P}\left(\{\omega_1 \in g^{-1}(A)\} \times \{\omega_2 \in h^{-1}(B)\}\right) \\ &= \mathbb{P}_1(\omega_1 \in g^{-1}(A)) \mathbb{P}_2(\omega_2 \in h^{-1}(B)). \end{aligned}$$

However,

$$\begin{aligned} \mathbb{P}_1(\omega_1 \in g^{-1}(A)) &= \mathbb{P}\{(\omega_1, \omega_2) : \omega_1 \in g^{-1}(A), \omega_2 \in \Omega_2\} \\ &= \mathbb{P}\{X \in A\}, \end{aligned}$$

and similarly for Y . □

Remark. Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space and suppose X, Y be two random variables that are independent and a.e. finite. They then generate two Borel probability measure μ_X and μ_Y on \mathbb{R} . Define a product probability space as of $(\mathbb{R}^2, \mathcal{B}_{\mathbb{R}^2}, \mu_X \times \mu_Y)$. Define $\tilde{X}(x, y) = x$ and $\tilde{Y}(x, y) = y$ as random variables on the product space. By definition, \tilde{X} is equidistributed with X . That is, $\mu_{\tilde{X}} = \mu_X$ and $F_{\tilde{X}} = F_X$. Similarly $\mu_{\tilde{Y}} = \mu_Y$. Also, \tilde{X}, \tilde{Y} are independent. Now (X, Y) and (\tilde{X}, \tilde{Y}) have the same distribution.

Remark. If X and Y are independent, then their joint distribution $F_{(X,Y)}$ is uniquely determined by the individual distributions of F_X, F_Y . Indeed,

$$F_{(X,Y)}(s, t) = F_X(s)F_Y(t).$$

Remark. Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space and suppose X, Y be two random variables that are independent. Suppose they have densities ρ_X, ρ_Y , then the distribution density of vector (X, Y) is $\rho_{(X,Y)}(s, t) = \rho_X(s)\rho_Y(t)$.

Remark. If X and Y are independent random variable, and f, g are measurable functions. Then $f(X)$ and $g(Y)$ are independent as well.

Remark. Given probability space $(\Omega, \Sigma, \mathbb{P})$ and random variable X . It may not exists another random variable Y that is independent from X on the same probability space. See the following example.

Example. As an example, consider $([0, 1], \mathcal{B}_{[0,1]}, m)$ and $X(\omega) = \omega$, so X is uniform on $[0, 1]$. The goal is to construct variable Y such that $Y \sim \text{Bernoulli}(\frac{1}{2})$.

Proof. Let $A \in \mathcal{B}_{[0,1]}$ and $t \in [0, 1]$. Define the density of set A at point t to be

$$\lim_{\varepsilon \rightarrow 0} \frac{m(A \cap [t - \varepsilon, t + \varepsilon])}{2\varepsilon}.$$

The Lebesgue density theorem says this is well defined and takes values in $\{0, 1\}$ for m -a.e. $t \in [0, 1]$.

Suppose such Y exists, we can view Y as an indicator of a set $A \subset [0, 1]$ of probability $\frac{1}{2}$. That is, $Y = \mathbb{1}_A$ and $\mathbb{P}(A) = \frac{1}{2}$. Choose any point $t \in (0, 1)$ such that density of A at t is well-defined. WLOG assume the density is 1. Pick $\varepsilon > 0$ such that $m(A \cap [t - \varepsilon, t + \varepsilon]) \geq \frac{3}{2}\varepsilon$. Now,

$$\begin{aligned} m(A \cap [t - \varepsilon, t + \varepsilon]) &= \mathbb{P}\{Y = 1, X \in [t - \varepsilon, t + \varepsilon]\} \\ &= \mathbb{P}\{Y = 1\} \mathbb{P}\{X \in [t - \varepsilon, t + \varepsilon]\} \\ &= \varepsilon, \end{aligned}$$

a contradiction.

Alternatively, we can derive a contradiction using $m(A) = \mathbb{P}\{Y = 1, X \in A\}$. □

Remark. The goal is to have statements “independent” from the underlying probability space.

2.3 Convolution

Definition (convolution). Let μ, ν be Borel probability measures on \mathbb{R} . The convolution of μ and ν is a probability measure on \mathbb{R} such that

$$(\mu * \nu)(S) = \int_{\mathbb{R}^2} \mathbb{1}_S(x + y) d(\mu \times \nu)$$

for all $S \in \mathcal{B}_{\mathbb{R}}$.

Remark. Suppose both μ and ν have densities ρ_μ and ρ_ν . That is, $\mu \ll m$ and $\nu \ll m$. It follows that

$$\begin{aligned} (\mu * \nu)(S) &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \mathbb{1}_S(x + y) \rho_\mu(x) dx \right) \rho_\nu(y) dy \\ &= \int_{\mathbb{R}} \left(\int_S \rho_\mu(w - y) dw \right) \rho_\nu(y) dy \\ &= \int_S \left(\int_{\mathbb{R}} \rho_\mu(w - y) \rho_\nu(y) dy \right) dw \end{aligned}$$

Note that this implies $\mu * \nu \ll m$ and

$$\rho_{\mu * \nu}(w) = \int_{\mathbb{R}} \rho_\mu(w - y) \rho_\nu(y) dy,$$

which is the convolution of function ρ_μ and ρ_ν .

Remark. If X and Y are two random variables and μ_X and μ_Y are the corresponding induced Borel measure \mathbb{R} , then $\mu_{X+Y} = \mu_X * \mu_Y$.

Remark. Suppose X and Y are independent and their CDF is F_X and F_Y , then

$$F_{X+Y}(t) = \int_{\mathbb{R}} F_X(t-w) d\mu_Y(w) = \int_{\mathbb{R}} F_X(t-w) dF_Y(w).$$

2.4 Moments

In this section we explore the computation and basic properties of moments.

Definition. Let X be a random variable on $(\Omega, \Sigma, \mathbb{P})$. The p -th absolute moment of X is

$$\mathbb{E} |X|^p = \int_{\Omega} |X|^p d\mathbb{P}.$$

This is always well-defined but can be infinite.

If p is positive integer, the p -th moment of X is

$$\mathbb{E} X^p = \int_{\Omega} X^p d\mathbb{P}$$

whenever it is defined.

In particular, $\mathbb{E}X$ is the mean or expectation, the variance is $\text{var}(X) = \mathbb{E}(X - \mathbb{E}X)^2$ whenever the expectation is defined, and the standard deviation is $\sqrt{\text{var}(X)}$.