Section Notes for PSTAT 213B

Haosheng Zhou

Jan, 2025

Contents

Week 1

Elements of Measure Theory

The measure-based probability theory is established on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. It consists of the sample space Ω, the sigma field $\mathscr{F}$ (definition?) on Ω, and the probability measure $\mathbb{P} : \mathscr{F} \to [0, 1]$. The probability measure satisfies the axioms of probability (definition?), which contains the essential "countable additivity". Within this framework, a random variable $X : \Omega \to \mathbb{R}$ is just a mapping $\omega \mapsto X(\omega)$ that is $\mathscr{F}$-measurable, i.e., $\{\omega : X(\omega) \in B\} \in \mathscr{F}, \forall B \subset \mathbb{R}$ Borel.

The Borel sigma field, denoted $\mathscr{B}_{\mathbb{R}}$, is the collection of all Borel measurable subsets of $\mathbb{R}$. Mathematically speaking, it is defined as $\mathscr{B}_{\mathbb{R}} := \sigma(\mathscr{Q})$, where $\mathscr{Q} := \{(a, b] : a, b \in \mathbb{R}\}$ is a π-system. That is the reason the CDF is defined as $F_X(x) = \mathbb{P}(X \leq x)$. To connect the CDF with the distribution/law of the r.v., we define the law of X, denoted μ_X, as the probability measure on $(\mathbb{R}, \mathscr{B}_{\mathbb{R}})$ induced by X: $\mu_X(B) = \mathbb{P}(X \in B)$ for any Borel set $B \in \mathscr{B}_{\mathbb{R}}$. The following exercise proves that the CDF characterizes the law of a random variable, which is a standard application of the $\pi - \lambda$ theorem.

Exercise 1. *Given μ, ν as two probability measures on $(\Omega, \mathscr{F})$, and $\mathscr{Q}$ is a π-system such that $\sigma(\mathscr{Q}) = \mathscr{F}$. If $\mu_1(A) = \mu_2(A)$ for $\forall A \in \mathscr{Q}$, prove that $\mu_1(A) = \mu_2(A)$ for $\forall A \in \mathscr{F}$.*

Consequently, prove that if μ_X, μ_Y are the laws of r.v. X and Y, then $F_X \equiv F_Y$ iff $\mu_X(B) = \mu_Y(B)$ for $\forall B \in \mathscr{B}$.

Hint. Consider $\mathscr{H} := \{A \in \mathscr{F} : \mu_1(A) = \mu_2(A)\}$ and prove that it is a λ-system. □

The next core topics are integration and the convergence theorems. By definition, the expectation $\mathbb{E}X := \int_\Omega X(\omega)\, d\mathbb{P}(\omega)$ is nothing else but the Lebesgue integral of the r.v. (measurable function) X. By a simple change of variable $x = X(\omega)$, we recover the important identity $\mathbb{E}X = \int_{\mathbb{R}} x\, d\mu_X(x)$ (prove it on your own). The key insight from the Lebesgue integration is that the behavior of the function on any zero-Lebesgue-measure set can be neglected, which naturally motivates the definition of "almost sure" and "almost everywhere" in measure theory. Any pointwise properties that are required to prove some relationships of Lebesgue integrals can be reduced to almost sure/almost everywhere properties without any costs.

One of the key problems w.r.t. the integration is when one has a sequence of random variables X_n and $X_n \stackrel{a.s.}{\to} X(\omega)$ $(n \to \infty)$ (pointwise can be reduced to almost sure convergence since neglected by the Lebesgue integral). We hope to understand under what conditions $\mathbb{E}X_n \to \mathbb{E}X$ $(n \to \infty)$ holds, i.e. the interchange of the limit and the integration is allowed. Three most important convergence theorems provide the sufficient conditions: (please check)

- Monotone convergence theorem (MCT): $X_n \geq 0$ $a.s.$ for $\forall n,\ \omega \in \Omega$, and X_n is $a.s.$ increasing in n.
- Dominated convergence theorem (DCT): $\sup_n |X_n| \leq Y$ $a.s.$, and $\mathbb{E}Y < \infty$.
- Bounded convergence theorem (BCT): $\sup_n |X_n| \leq M$ $a.s.$, where $M \in \mathbb{R}$.

If one has a weaker condition (e.g., only non-negativity but no monotonicity), it is possible to derive a weaker conclusion, stated by the Fatou's lemma (please check).

The next important concept is the Radon-Nikodym derivative, which often comes up when studying conditional expectation. The problem of interest is that, when having two probability measures $\mathbb{P},\mathbb{Q}$ on the same measurable space $(\Omega,\mathscr{F})$, if the following representation exists:

$$\mathbb{P}(A)=\int_A f(\omega)\,d\mathbb{Q}(\omega),\ \forall A\in\mathscr{F}, \tag{1}$$

for some measurable function $f:\Omega\to\mathbb{R}$. As a necessary condition, for any $A\in\mathscr{F}$ such that $\mathbb{Q}(A)=0$, it holds that $\mathbb{P}(A)=0$. We call this property the absolute continuity of $\mathbb{P}$ w.r.t. $\mathbb{Q}$, denoted $\mathbb{P}<<\mathbb{Q}$. Amazingly, this is also the sufficient condition! The Radon-Nikodym theorem states that, if $\mathbb{P}<<\mathbb{Q}$ for two probability measures (actually σ-finite, so it applies for λ, the Lebesgue measure on $\mathbb{R}$), then such measurable function f must exist and is almost surely unique such that equation (1) holds. We call such a function f the Radon-Nikodym derivative and denote it by $\frac{d\mathbb{P}}{d\mathbb{Q}}$. The following exercise states some fundamental properties of the RN-derivative.

Exercise 2. *Argue that $\frac{d\mathbb{P}}{d\mathbb{Q}}$ is actually a r.v. on the probability space $(\Omega,\mathscr{F},\mathbb{Q})$. Check that $\mathbb{E}_{\mathbb{Q}}\frac{d\mathbb{P}}{d\mathbb{Q}}=1$, where $\mathbb{E}_{\mathbb{Q}}$ denotes the expectation under measure $\mathbb{Q}$.*

If $\frac{d\mathbb{P}}{d\mathbb{Q}}>0,\ \mathbb{Q}-a.s.$, prove that $\frac{d\mathbb{Q}}{d\mathbb{P}}$ exists and that $\frac{d\mathbb{Q}}{d\mathbb{P}}=\frac{1}{\frac{d\mathbb{P}}{d\mathbb{Q}}}$ under both $\mathbb{P}-a.s.$ sense and $\mathbb{Q}-a.s.$ sense.

Hint. To identify $\frac{d\mathbb{Q}}{d\mathbb{P}}$, use the almost sure uniqueness of the RN-derivative. □

Why does the notation of the RN derivative look like a differential between two measures? The following exercise provides the intuition.

Exercise 3. *For disjoint sets $A\in\mathscr{F}$ and $\Delta A\in\mathscr{F}$ (understood as the perturbation in A), calculate $\frac{\mathbb{P}(A\cup\Delta A)-\mathbb{P}(A)}{\mathbb{Q}(A\cup\Delta A)-\mathbb{Q}(A)}$. Prove that this difference quotient is equal to $f(\omega_0)$ if $\Delta A=\{\omega_0\}$ and $\mathbb{Q}(\{\omega_0\})>0$.*

Argue intuitively that $\frac{\mathbb{P}(A\cup\Delta A)-\mathbb{P}(A)}{\mathbb{Q}(A\cup\Delta A)-\mathbb{Q}(A)}\to f(\Delta A)$ as $\mathbb{Q}(\Delta A)\to 0$, which is interpreted as the relative rate of change in the measure when A receives an infinitesimal perturbation.

Hint. Definition. □

The Lebesgue decomposition theorem implies that for the law μ_X of r.v. X, there exists a unique decomposition of the law $\mu_X=\mu+\nu$ such that $\mu<<\lambda$ (absolute continuous w.r.t. Lebesgue measure) and $\nu\perp\lambda$ (singular w.r.t. Lebesgue measure). We say $\nu\perp\lambda$ if there exists disjoint$A,B\in\mathscr{B}$ such that $A\cup B=\mathbb{R}$, but $\nu(A)=\lambda(B)=0$. This provides the classification of random variables into discrete, continuous, and singular r.v. In some sense, that is why probability densities and probability mass functions are both called "likelihood" in statistics.

Exercise 4. *Consider the probability space $(\Omega=[0,1],\mathscr{F}=\mathscr{B}_{[0,1]},\mathbb{P}=\lambda)$, where λ is the Lebesgue measure.*

If $X(\omega)=\begin{cases}0 & \text{if } \omega<\frac{1}{2}\\ 1 & \text{if } \omega\geq\frac{1}{2}\end{cases}$, find out μ,ν in the Lebesgue decomposition of μ_X (discrete).

If $X(\omega)=\omega$, find out μ,ν in the Lebesgue decomposition of μ_X (continuous).

Show that there exists a Borel measurable set C which is uncountable, but has zero Lebesgue measure. Consider r.v. X that is supported on C. Prove that $\mu\equiv 0$ in the Lebesgue decomposition of μ_X. This is neither a discrete r.v. nor a continuous r.v. (singular).

Hint. Consider Cantor set C. $\square$

After clearing up those concepts for a single r.v., we start to talk about random vectors, most importantly, the concept of independence between two r.v. X and Y. The independence of X and Y is defined as the independence between sigma fields $\sigma(X)$ and $\sigma(Y)$. Two sigma fields $\mathscr{A}, \mathscr{B}$ are defined to be independent if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B)\,,\ \forall A \in \mathscr{A}, B \in \mathscr{B}. \tag{2}$$

The following theorem is again an application of the $\pi - \lambda$ theorem.

Exercise 5. *$\mathscr{Q}, \mathscr{R}$ are π-systems and subsets of $\mathscr{F}$. If $\mathscr{Q}, \mathscr{R}$ are independent, then $\sigma(\mathscr{Q}), \sigma(\mathscr{R})$ are independent. Consequently, prove that X, Y are independent iff $\mathbb{P}(X \leq x, Y \leq y) = \mathbb{P}(X \leq x)\,\mathbb{P}(Y \leq y)$, $\forall x, y \in \mathbb{R}$.*

Hint. Fix $A \in \mathscr{Q}$, consider $\mathscr{H}_A := \{B \in \sigma(\mathscr{R}) : \mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B)\}$ and prove $\mathscr{Q}, \sigma(\mathscr{R})$ are independent. Then fix $B \in \mathscr{R}$ and perform the similar argument once more. $\square$

To build a larger probability space where two (finitely many) random variables on $(\Omega_1, \mathscr{F}_1, \mathbb{P})$ and $(\Omega_2, \mathscr{F}_2, \mathbb{Q})$ can live, one sets the sample space as $\Omega_1 \times \Omega_2$ and wishes to build a sigma field $\mathscr{F}_1 \otimes \mathscr{F}_2$ on $\Omega_1 \times \Omega_2$. The procedure is similar to that in one dimension, except setting $\mathscr{F}_1 \otimes \mathscr{F}_2 := \sigma(\mathscr{Q})$, where the π-system $\mathscr{Q} := \{A \times B : A \in \mathscr{F}_1, B \in \mathscr{F}_2\}$ is the collection of measurable rectangles. $\mathscr{F}_1 \otimes \mathscr{F}_2$ is called the product sigma field, which is typically larger than $\mathscr{F}_1 \times \mathscr{F}_2$. Concerning the product probability measure $\mathbb{P} \otimes \mathbb{Q} : \mathscr{F}_1 \otimes \mathscr{F}_2 \to [0, 1]$, it is also only defined on $\mathscr{Q}$ through $(\mathbb{P} \otimes \mathbb{Q})(A \times B) := \mathbb{P}(A) \cdot \mathbb{Q}(B)$ and later extended to the whole $\mathscr{F}_1 \otimes \mathscr{F}_2$. That is the construction of the product probability space $(\Omega_1 \times \Omega_2, \mathscr{F}_1 \otimes \mathscr{F}_2, \mathbb{P} \otimes \mathbb{Q})$. Actually, Kolmogorov's extension lemma guarantees the conclusion for uncountably many r.v., provided the consistency condition.

As the last topic, we shall talk about Fubini's theorem when it comes to double integration/summation. If $\mathbb{P}, \mathbb{Q}$ are both sigma finite measures, the key condition is that either the integrand f is a.s. non-negative or integrable, i.e., $\int_{\Omega_1 \times \Omega_2} |f(\omega_1, \omega_2)|\, d(\mathbb{P} \otimes \mathbb{Q})(\omega_1, \omega_2) < \infty$. Under this, one can interchange the order of integration, i.e.

$$\int_{\Omega_1 \times \Omega_2} f(\omega_1, \omega_2)\, d(\mathbb{P} \otimes \mathbb{Q})(\omega_1, \omega_2) = \int_{\Omega_2} \int_{\Omega_1} f(\omega_1, \omega_2)\, d\mathbb{P}(\omega_1)\, d\mathbb{Q}(\omega_2) = \int_{\Omega_1} \int_{\Omega_2} f(\omega_1, \omega_2)\, d\mathbb{Q}(\omega_2)\, d\mathbb{P}(\omega_1). \tag{3}$$

The following exercise is one of the most important implications of Fubini's theorem to keep in mind.

Exercise 6. *Prove $\mathbb{E}|X|^p = \int_0^\infty p y^{p-1} \mathbb{P}(|X| > y)\, dy$ for $\forall p > 0$.*

Hint. Apply Fubini's theorem using $|x|^p = \int_0^{|x|} p y^{p-1}\, dy$. $\square$

Week 2

Almost Sure and Convergence in Probability

The following exercise tests the understanding on almost sure and convergence in probability.

Exercise 7. *Consider the probability space* $([0,1], \mathscr{B}_{[0,1]}, \lambda)$*, on which there exists a sequence of random variables* $X_n(\omega) := \mathbb{I}_{(0,\frac{1}{n})}(\omega)$*. Judge if this sequence of r.v. converges a.s./in probability, and identify the limit.*

Repeat the exercise for a sequence of independent r.v. Y_n *on the same probability space such that* $Y_n \stackrel{d}{=} X_n, \ \forall n$*.*

Hint. $\{X_n\}$ converges almost surely but not $\{Y_n\}$. Independence matters. □

Exercise 8 (Convergence for i.i.d. r.v.)**.** *Refer to Lemma 14 (1), (3) in 2024 notes.*

Week 3

Exercise 9. *Let $\{X_n\}$ be i.i.d. random variables following $\mathcal{E}(1)$, and $M_n := \max_{1\leq i\leq n}\{X_i\}$, show that*

$$\frac{M_n}{\log n} \xrightarrow{a.s.} 1 \ (n \to \infty). \tag{4}$$

Note that we have already proved $\limsup_{n\to\infty} \frac{X_n}{\log n} = 1$ a.s. in the homework.

Hint. Prove the following conclusions under the almost sure sense: $\limsup_{n\to\infty} \frac{M_n}{\log n} \geq 1$, $\limsup_{n\to\infty} \frac{M_n}{\log n} \leq 1$, $\liminf_{n\to\infty} \frac{M_n}{\log n} \geq 1$, among which the first one is obvious.

To prove $\limsup_{n\to\infty} \frac{M_n}{\log n} \leq 1$, split the maximum w.r.t. n terms into the first N terms (finitely many) and the tail $n - N$ terms (infinitely many). Use $\limsup_{n\to\infty} \frac{X_n}{\log n} = 1$ to bound the tail part.

To prove $\liminf_{n\to\infty} \frac{M_n}{\log n} \geq 1$, apply Borel-Cantelli and prove $\forall \varepsilon \in (0,1)$, $\sum_{n=1}^{\infty} \mathbb{P}\left(\frac{M_n}{\log n} < 1-\varepsilon\right) < \infty$. □

Exercise 10. *Show that*

$$d(X,Y) := \mathbb{E}\frac{|X-Y|}{1+|X-Y|} \tag{5}$$

is a metric on the space of random variables (the equality is under almost sure sense). Show that $X_n \xrightarrow{p} X$ $(n \to \infty)$ iff $d(X_n, X) \to 0$, which shows that convergence in probability can be embedded into a metric space.

Hint. The triangle inequality of d follows from the fact that $x \mapsto \frac{x}{1+x}$ is increasing. The equivalence in convergence follows from BCT and Markov inequality. □

Exercise 11. *Show that for $1 \leq p < q \leq \infty$, L^q convergence implies L^p convergence.*

Hint. Holder's inequality. □

Exercise 12. *Check that for $X_n \equiv X$, $\{X_n\}$ is U.I. iff $X \in L^1$. This provides the motivation of the def of U.I.*

If $X \in L^1$, prove that $\forall \varepsilon > 0$, $\exists \delta > 0$, such that for any $A \in \mathscr{F}, \mathbb{P}(A) < \delta$, $\mathbb{E}|X|\mathbb{I}_{|X|\in A} < \varepsilon$ holds.

Hint. By Cauchy principle, $\forall \varepsilon > 0$, $\exists M > 0$ such that $\int_{|X(\omega)|\geq M} |X(\omega)|\, d\mathbb{P}(\omega) < \frac{\varepsilon}{2}$. □

Exercise 13. *If $\{X_n\}$ and $\{Y_n\}$ are U.I., show that $\{X_n + Y_n\}$ is U.I.*

Hint. Definition. □

Exercise 14. *Show that $\{X_n\}$ is U.I., if one of the following conditions holds:*

(1): (moment) exists $\varepsilon > 0$, such that $\sup_n \mathbb{E}|X_n|^{1+\varepsilon} < \infty$.

(2): (dominated) exists $Y \in L^1$, such that $\sup_n |X_n| \leq Y$ a.s..

Explain why condition (1) cannot be weakened to $\sup_n \mathbb{E}|X_n| < \infty$.

Hint. Part (1): consider $\mathbb{E}\frac{|X_n|}{|X_n|^{1+\varepsilon}}|X_n|^{1+\varepsilon}\mathbb{I}_{|X_n|\geq\lambda}$. Use the fact that $\frac{x}{x^{1+\varepsilon}} \to 0$ $(x \to +\infty)$ to connect with the moment condition. Part (2): by definition.

Counterexample: $X_n(\omega) = n\mathbb{I}_{(0,\frac{1}{n})}(\omega)$. Contradiction follows from Vitali's convergence theorem. □