Lecture Note for MAT8030: Advanced Probability

LI Liying

September 5, 2024

1 Measure theory preliminaries

In this section we cover some basic facts in measure theory and how they integrate into the modern probability theory, which is essential to this field. Most of the material is still with the scope of the celebrated work, *Foundations of the theory of probability*, by Kolmogorov in 1933 ([Kol]), and interesting readers should take a look at.

1.1 Random variables, σ-fields and measures

We start with examples of some random variables (r.v.'s) that the reader should be familiar with from elementary probability. There are two types of r.v.'s encountered in elementary probability: discrete and continuous.

Example 1.1 Examples for discrete r.v.'s.

- **Bernoulli:** $X \sim \mathrm{Ber}(p)$, with $\mathsf{P}(X = 1) = p$, $\mathsf{P}(X = 0) = 1 - p$.
- **binomial:** $X \sim \mathrm{Binom}(n, p)$ with $\mathsf{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$, $k = 0, 1, \ldots, n$.
- **geometry:** $X \sim \mathrm{Geo}(p)$, with $\mathsf{P}(X = k) = (1 - p)^{k-1} p$, $k = 1, 2, \ldots$.
- **Poisson:** $X \sim \mathrm{Poi}(\lambda)$, with $\mathsf{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$, $k = 0, 1, \ldots$.

Example 1.2 Examples for continuous r.v.'s, distribution described by the density function $\mathsf{P}(X \leq a) = \int_{-\infty}^{a} p(x)\,dx$.

- **exponential:** $X \sim \mathrm{Exp}(\lambda)$, with $p(x) = \mathbb{1}_{[0,\infty)}(x) \cdot \lambda e^{-\lambda x}$.
- **uniform:** $X \sim \mathrm{Unif}[a, b]$, with $p(x) = \mathbb{1}_{[a,b]}(x) \cdot \frac{1}{b-a}$.
- **normal/Gaussian:** $X \sim \mathcal{N}(\mu, \sigma^2)$, with $p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$.

Recall that the distribution/law of a r.v. X is determined by its cumulative distribution function (c.d.f.). In particular, sets of the form $\{X \leq a\}$ are *events* of which one can evaluate the probability, denoted by $\mathsf{P}(X \leq a)$.

We can say that $\mathsf{P}(\cdot)$ is a function of events, or a *set function*. A measure $\mathsf{P}(\cdot) : A \mapsto \mathsf{P}(A) \in [0, \infty)$ is a special set-function satisfying the following three properties:

1. **non-negativity**: $\mathsf{P}(A) \geq 0$, $\forall A$.
2. $\mathsf{P}(\varnothing) = 0$.
3. **countable additivity**: for any *disjoint* $A_1, A_2, \ldots$,

$$\mathsf{P}\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} \mathsf{P}(A_n). \tag{1.1}$$

The last property, countable additivity (a.k.a. σ-additivity) is the most important one. It is only with σ-additivity, not finite additivity, that one can get the hands on various limit theorems for integration, and in the context of probability, for mathematical expectation.

Other important properties of measures can be derived from Item 1 to Item 3.

4. **finite additivity** from Items 2 and 3: let $A_{n+1} = A_{n+2} = \cdots = \varnothing$ in (1.1); then

$$\mathsf{P}\Big(\bigcup_{k=1}^{n} A_k\Big) = \sum_{k=1}^{n} \mathsf{P}(A_k).$$

5. **it:monotonicity** from Items 1 and 4: if $A \subset B$, then $A \cap (B \setminus A) = \varnothing$, and hence

$$\mathsf{P}(B) = \mathsf{P}(A) + \mathsf{P}(B \setminus A) \geq \mathsf{P}(A).$$

6. **sub-additivity** from Items 3 and 5: let $\tilde{A}_n = A_n \setminus \big(\bigcup_{k=1}^{n-1} A_k\big) \subset A_n$; then

$$\mathsf{P}\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} \mathsf{P}(\tilde{A}_n) \leq \sum_{n=1}^{\infty} \mathsf{P}(A_n).$$

7. **continuity from above** from Items 2 and 3: if $A_n \downarrow A$ and $\mathsf{P}(A_1) < \infty$, then $\mathsf{P}(A) = \lim_{n\to\infty} \mathsf{P}(A_n)$ $(A = \bigcap_{n=1}^{\infty} A_n)$. In fact, since A_1 is the disjoint union of

$$A_1 = A \cup (A_1 \setminus A_2) \cup (A_2 \setminus A_3) \cup \cdots,$$

we have

$$\mathsf{P}(A_1) = \mathsf{P}(A) + \mathsf{P}(A \setminus A_n) + \sum_{k=n}^{\infty} \mathsf{P}(A_k \setminus A_{k+1}).$$

All the terms are positive, and the left hand side is finite, so the tail of the infinite sum must converges to 0, and hence

$$\mathsf{P}(A) = \lim_{n\to\infty} \mathsf{P}(A_1) - \mathsf{P}(A \setminus A_n) - \sum_{k=n}^{\infty} \mathsf{P}(A_k \setminus A_{k+1}) = \lim_{n\to\infty} \mathsf{P}(A_1) - \mathsf{P}(A_1 \setminus A_n) = \lim_{n\to\infty} \mathsf{P}(A_n).$$

8. **continuity from below** from Items 2, 3, 5 and 7: if $A_n \uparrow A$, then $\mathsf{P}(A) = \lim_{n\to\infty} \mathsf{P}(A_n)$.

Noting that $\mathsf{P}(A_n)$ is increasing, by sub-additivity,

$$\mathsf{P}(A) \leq \sum_{n=2}^{\infty} \mathsf{P}(A_n \setminus A_{n-1}) = \lim_{n\to\infty} \mathsf{P}(A_n).$$

If $\mathsf{P}(A) = \infty$, there is nothing else to prove. Otherwise, $\mathsf{P}(A) < \infty$, and $A - A_n \downarrow \varnothing$. Then by continuity from above,

$$0 = \mathsf{P}(\varnothing) = \lim_{n\to\infty} \mathsf{P}(A \setminus A_n) = \lim_{n\to\infty} \mathsf{P}(A) - \mathsf{P}(A_n).$$

We also need to impose some conditions on the domain of the set-function $\mathsf{P}(\cdot)$. The domain should behave well under countable union/intersection. This leads to the definition of σ-algebra.

Definition 1.1 *Let Ω be any non-empty set. A collection of subsets of Ω, $\mathcal{F}$, is a σ-algebra (or σ-field), if*

1. $\Omega \in \mathcal{F}$,
2. $A \in \mathcal{F}$ *implies* $A^c \in \mathcal{F}$,
3. *(closure under countable union)* $A_n \in \mathcal{F}$ *implies* $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

Example 1.3
1. The smallest σ-algebra: $\mathcal{F} = \{\varnothing, \Omega\}$.
2. The largest σ-algebra: $\mathcal{F} = \{$ all subsets of $\Omega\}$.

A set Ω equipped with a σ-algebra $\mathcal{F}$ is called a *measurable space*, written in a pair $(\Omega, \mathcal{F})$.

Proposition 1.1 *Let $\mathcal{F}$ be a σ-field. Then*

- $\varnothing \in \mathcal{F}$,
- $A \subset B$, $A, B \in \mathcal{F}$ *imply* $B \setminus A \in \mathcal{F}$,
- *(closure under countable intersection)* $A_n \in \mathcal{F}$ *implies* $\bigcap_{n=1}^{\infty} A_n \in \mathcal{F}$.

Definition 1.2 *A probability space, or probability triple, $(\Omega, \mathcal{F}, \mathsf{P})$ is such that $(\Omega, \mathcal{F})$ is a measurable space and $\mathsf{P} : \mathcal{F} \to [0, 1]$ is a measure with $\mathsf{P}(\Omega) = 1$.*

Definition 1.3 *A random variable (r.v.) $X = X(\omega) : \Omega \to \mathbb{R}$ is a map from a probability space $(\Omega, \mathcal{F}, \mathsf{P})$ to $\mathbb{R}$, such that*

$$\{\omega : X(\omega) \le a\} \in \mathcal{F}, \quad \forall a \in \mathbb{R},$$

or written more compactly, $X^{-1}\big((-\infty, a]\big) \in \mathcal{F}$ for all $a \in \mathbb{R}$.

Let us recall some basic facts about the pre-image map φ^{-1} for *any* map $\varphi : U \to V$. It is defined by

$$\varphi^{-1}(W) := \{u \in U : \varphi(u) \in W\}.$$

Proposition 1.2 *The map φ^{-1} commutes with most set operations, in particular:*

- $\varphi^{-1}(W_1 \cap W_2) = \varphi^{-1}(W_1) \cap \varphi^{-1}(W_2)$,
- $\varphi^{-1}(W_1 \cup W_2) = \varphi^{-1}(W_1) \cup \varphi^{-1}(W_2)$,
- $\varphi^{-1}(W^c) = \big(\varphi^{-1}(W)\big)^c$.

Let X be a r.v. on $(\Omega, \mathcal{F}, \mathsf{P})$, and let $\mathcal{B} = \{A \text{ s.t. } X^{-1}(A) \in \mathcal{F}\}$. Definition 1.3 and Proposition 1.2 imply that $\mathcal{B}$ contains all the intervals in $\mathbb{R}$. Moreover, since $\mathcal{F}$ is a σ-algebra,

$$X^{-1}(I_n) \in \mathcal{F} \quad \Longrightarrow \quad X^{-1}\Big(\bigcup_{n=1}^{\infty} I_n\Big) = \bigcup_{n=1}^{\infty} X^{-1}(I_n) \in \mathcal{F}.$$

This implies that $\mathcal{B}$ is also a σ-algebra. As we will see in the next section, $\mathcal{B}$ is the *Borel σ-algebra*, which is the most important class of σ-algebras in probability theory.

1.2 Construction of σ-algebra and (probability) measures

Simply put, the Borel σ-algebra is the *smallest* σ-algebra containing by open sets. To understand what is "smallest", we start with the following observation.

Lemma 1.3 *1. If $\mathcal{F}_1$ and $\mathcal{F}_2$ are two σ-algebras on Ω, then $\mathcal{F}_1 \cap \mathcal{F}_2$ is also a σ-algebra.*

2. If $\mathcal{F}_\gamma, \gamma \in \Gamma$ are σ-algebras on Ω, where Γ is an arbitrary index set (countable or uncountable), then $\bigcap_{\gamma \in \Gamma} \mathcal{F}_\gamma$ is also a σ-algebra.

Proposition 1.4 *Let $\mathcal{A}$ be a collection of subsets in Ω. Then there exists a smallest σ-algebra containing $\mathcal{A}$, called* the σ-algebra generated by $\mathcal{A}$ *and written $\sigma(\mathcal{A})$, in the sense that if $\mathcal{G} \supset \mathcal{A}$ is a σ-algebra, then $\sigma(\mathcal{A}) \subset \mathcal{G}$.*

Proof: Take $\sigma(\mathcal{A}) = \bigcap_{\mathcal{F}\ \sigma\text{-algebra}: \mathcal{F} \supset \mathcal{A}} \mathcal{F}$. $\square$

Definition 1.4 (Borel σ-algebra) *Let M be a metric space (or any topological space). The* Borel σ-algebra $\mathcal{B}(M)$ *is the σ-algebra generated by all the open sets in M.*

Example 1.4

- $\mathcal{B}(\mathbb{R}) = \sigma\Big((-\infty, a], a \in \mathbb{R}\Big)$.
- $\mathcal{B}(\mathbb{R}^d) = \sigma\Big((-\infty, a_1] \times \cdots \times (-\infty, a_d], a_i \in \mathbb{R}\Big)$.

Proposition 1.5 *A map $X(\omega)$ on $(\Omega, \mathcal{F}, \mathsf{P})$ is a r.v. if and only if $X^{-1}(A) \in \mathcal{F}$ for any $A \in \mathcal{B}(\mathbb{R})$.*

This is the usual definition for r.v.'s.

Now let us take about the distribution of a r.v. X. One can check that $\mu = \mathsf{P} \circ X^{-1}$ defined by

$$\mu(A) = \mathsf{P}(\{\omega : X(\omega) \in A\}), \quad A \in \mathcal{B}(\mathbb{R}),$$

is a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. We call μ the *distribution/law* of X. Clearly, $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$ is a probability space. For most of the practical application, say computing expectation, variance, etc, it is enough to understand the distribution of a r.v., not the original probability measure P on some abstract space that can be potentially be very complicate. Another obvious advantage is that the distributions of all r.v.'s are probability measures live on the *same* measurable space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Note that the c.d.f. of a r.v. can be read from its distribution:

$$F_X(a) = \mathsf{P}(X \le a) = \mu\big((-\infty, a]\big), \quad a \in \mathbb{R}.$$

The central topic for this section is to understand how the c.d.f. determines μ.

Theorem 1.6 *Every increasing, right-continuous function $F : \mathbb{R} \to [0,1]$ with $F(-\infty) = 0$ and $F(\infty) = 1$ uniquely determines a probability measure μ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.*

Along the way we will learn how to construct σ-algebras and (probability) measures.

References

[Kol] A.N. Kolmogorov. *Foundations of the Theory of Probability (English Translation).*