[AI2613 Lecture 1] Review of Probability Theory

March 2, 2022

1 Probability Space

We start with the notion of probability space. The standard reference for the probability theory is [1].

Definition 1 (Probability Space) *A* probability space *is a tuple* $(\Omega, \mathcal{F}, \mathsf{P}(\cdot))$ *satisfying the following requirements.*

- *The universe* Ω *is a set of "outcomes" (which can be either countable or uncountable).*
- *The set* $\mathcal{F} \subseteq 2^{\Omega}$ *is a* σ*-algebra (the set of all possible "events"). Here we say* $\mathcal{F}$ *is a* σ*-algebra if* $\mathcal{F}$ *satisfies:*
  - $\emptyset, \Omega \in \mathcal{F}$;
  - $\forall A \in \mathcal{F}$*, it holds* $A^c \in \mathcal{F}$;
  - *for any finite or countable sequence of sets* $A_1, \dots, A_n, \dots \in \mathcal{F}$*, it holds that* $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.
- *The probability function* $\mathsf{P}(\cdot) : \mathcal{F} \to [0,1]$ *satisfies*
  - $\mathsf{P}(\emptyset) = 0, \mathsf{P}(\Omega) = 1$;
  - $\mathsf{P}(A^c) = 1 - \mathsf{P}(A)$ *for all* $A \in \mathcal{F}$;
  - *for any finite or countable sequence of* disjoint *sets* $A_1, \dots, A_n, \dots \in \mathcal{F}$*, it holds that* $\mathsf{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathsf{P}(A_i)$.

$A^c \triangleq \Omega \setminus A$.

Let $\mathcal{S} \subseteq 2^{\Omega}$. We use $\sigma(\mathcal{S})$ to denote the minimal σ-algebra containing sets in $\mathcal{S}$. That is, for any $\mathcal{F} \subseteq 2^{\Omega}$, $\mathcal{F} = \sigma(\mathcal{S})$ if and only if (1) $\mathcal{F}$ is a σ-algebra; (2) $\mathcal{S} \subseteq \mathcal{F}$; (3) For any $\mathcal{F}' \subseteq \mathcal{F}$ such that $\mathcal{S} \subseteq \mathcal{F}'$, $\mathcal{F}'$ is not a σ-algebra.

The term "minimal" here is with respect to the set inclusion relation $\subseteq$.

For every $n \in \mathbb{N}$, we use $[n]$ to denote the set $\{1, 2, \dots, n\}$.

Example 1 (Tossing n fair coins) *Let* $\Omega = \{0,1\}^n$, $\mathcal{F} = 2^{\Omega}$ *and for every* $S \in \{0,1\}^n$, $\mathsf{P}(\{S\}) = \frac{1}{2^n}$.

Example 2 (Uniform Reals in $(0,1)$) *The uniform distribution on* $(0,1)$ *is defined as follows:*

- $\Omega = (0,1)$;
- $\mathcal{F}$ *is the* σ*-algebra consisiting of all* Borel sets *on* $(0,1)$*, namely the collection of subsets of* $(0,1)$ *obtained from open intervals by repeatedly taking* countable *unions and complements;*
- $\forall$ *interval* $I = (a,b)$, $P(I) = b - a$ *(This is the* Lebesgue measure*).*

The definition here, although a bit wired at the first glance, is in fact the simplest way to capture our intuition that the probability that a point is in (a,b) should be $b-a$. We cannot take $\mathcal{F} = 2^{\Omega}$ in Example 2 as doing so may include some *non-measurable* sets. In fact, $\mathcal{F}$ is called the *Borel algebra*, which is the smallest σ-algebra containing all open intervals. One can construct a non-Borel set in $(0,1)$ assuming the *axiom of choice*. In fact, the existence of a non-Borel set is independent of Zermelo-Fraenkel set theory without the axiom of choice. We use $\mathscr{R}$ to denote the collection of Borel sets on $\mathbb{R}$. For any $A \subseteq \mathbb{R}$, we use $\mathscr{R}(A)$ to denote $\mathscr{R} \cap 2^A$.

2 Random Variables

Definition 2 (Measurable Space) *Consider a set Ω and a σ-algebra $\mathcal{F}$ on Ω. The tuple $(\Omega, \mathcal{F})$ is called a measurable space.*

Definition 3 (Measurable Function) *Let $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$ be two measurable spaces and $X : \Omega \to \Omega'$ be a function. We say X is a $\mathcal{F}$-*measurable function *if*

$$\forall B' \in \mathcal{F}',\ X^{-1}(B') \in \mathcal{F},$$

$X^{-1}(B') \triangleq \{\omega \in \Omega | X(\omega) \in B'\}$ is the inverse of X.

For any function, we use $\sigma(X)$ to denote the minimal σ-algebra $\mathcal{F}$ such that X is $\mathcal{F}$-measurable.

Definition 4 (Random Variable) *. Let Ω' and $\mathcal{F}'$ in Definition 3 be $\mathbb{R}$ and the Borel algebra $\mathcal{B}$, then X in Definition 3 is a (real-valued) random variable.*

We say a random variable X *discrete* if its range $\mathrm{Ran}(X)$ is countable. In other words, X can only take at most countable many distinct values. Otherwise, we say X is a *continuous* random variable.

The *measurability* of a random variable X captures the intuition that we can safely talk about *the probability of X taking some value.* Intuitively X induces a partition of Ω where two outcomes ω_1 and ω_2 are in the same partition if and only if $X(\omega_1) = X(\omega_2)$. If the partition defined by X is more "coaser" than the partition defined by a σ-algebra $\mathcal{F}$, then X is $\mathcal{F}$ measurable.

Example 3 (Measurable Functions of Tossing a Dice) *. Let $\Omega = [6]$. We have three σ-algebras on Ω: $\mathcal{F}_1 = 2^{[6]}$, $\mathcal{F}_2 = \sigma(\{1,3,5\})$ and $\mathcal{F}_3 = \sigma(\{1,2\})$. Consider three random variables $X_1, X_2, X_3 : \Omega \to \mathbb{R}$ such that $X_1 : \omega \mapsto \omega$, $X_2 : \omega \mapsto \omega \bmod 2$ and $X_3 : \omega \mapsto \mathbf{1}[\omega \le 2]$. Then all these three mappings are $\mathcal{F}_1$-measurable, only X_2 is $\mathcal{F}_2$-measurable and only X_3 is $\mathcal{F}_3$-measurable.*

3 Distribution

Let $(\Omega, \mathcal{F}, \mathsf{P})$ be a probability space and $X : \Omega \to \mathbb{R}$ be a $\mathcal{F}$-measurable random variable. Let $\mathcal{B}$ be the Borel algebra on $\mathbb{R}$. The distribution space $(\mathbb{R}, \mathcal{B}, \mathbf{Pr})$ induced by X is defined as

$$\forall A \in \mathcal{B}, \mathbf{Pr}[A] = \mathbf{Pr}[X \in A] \triangleq \mathbf{P}[X^{-1}(A)].$$

If a function $f : \mathbb{R} \to \mathbb{R}$ satisfies for any $a \le b$:

$$\mathbf{Pr}[a \le X \le b] = \mathbf{Pr}\left[X^{-1}([a,b])\right] = \int_a^b f(x)\,\mathrm{d}x,$$

Then we call $f(x)$ the *probability density function (pdf)* of X.

The function $F(x) \triangleq \mathbf{Pr}[X \le x] = \int_{-\infty}^{x} f(t)\,\mathrm{d}t$ is called the *cumulative distribution function (cdf)* of X.

Example 4 (Exponential Distribution) *If $X \sim \mathrm{Exp}(\lambda)$, or equivalently it follows exponential distribution with rate λ for $\lambda > 0$, then its pdf is*

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & \textit{otherwise.} \end{cases}$$

4 Expectation and Variance

Definition 5 (Expectation) *. Let $(\Omega, \mathcal{F}, \mathsf{P})$ be a probability space and $X : \Omega \to \mathbb{R}$ be a random variable.*

- *For a discrete random variable X, its expectation is*

$$\mathbf{E}[X] \triangleq \sum_{a \in \mathsf{Ran}(X)} a \cdot \mathbf{Pr}[X = a].$$

If Ω is at most countable, we can also write

$$\mathbf{E}[X] = \sum_{\omega \in \Omega} \mathsf{P}(\{\omega\}) \cdot X(\omega).$$

- *For a continuous random variable X with pdf f, its expectation is*

$$\mathbf{E}[X] \triangleq \int_{-\infty}^{\infty} t \cdot f(t)\, \mathrm{d}t.$$

Sometimes it is more convenient to equivalently write the expectation as

$$\mathbf{E}[X] = \int_{\Omega} X(\omega)\mu(d\omega) = \int_{\Omega} X \,\mathrm{d}\mu.$$

using Lebesgue integration.

Example 5 (Expectation of Exponential Distribution) *Let $X \sim \mathsf{Exp}(\lambda)$ for $\lambda > 0$, then*

$$\mathbf{E}[X] = \int_0^{\infty} t \cdot \lambda e^{-\lambda t}\, \mathrm{d}t = \frac{1}{\lambda}.$$

Definition 6 (Variance) *The variance of a random variable X is*

$$\mathbf{Var}[X] \triangleq \mathbf{E}\left[(X - \mathbf{E}[X])^2\right] = \mathbf{E}\left[X^2\right] - \mathbf{E}[X]^2.$$

Proposition 7 *Let $X_1, \ldots, X_n$ be random variables where n is a finite constant. Then*

$$\mathbf{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbf{E}[X_i].$$

5 Conditional Probability

Definition 8 (Conditional Probability) *Let $(\Omega, \mathcal{F}, \mathsf{P})$ be a probability space. Let $A, B \in \mathcal{F}$ be two events with $\mathsf{P}(B) > 0$. The conditional probability of A given B is*

$$\mathsf{P}(A \mid B) \triangleq \frac{\mathsf{P}(A \cap B)}{\mathsf{P}(B)}.$$

This is well-defined since we know from the definition of σ-algebra that $A \cap B \in \mathcal{F}$.

In the following, we define the notion of *conditional expectation* for those *discrete* random variables.

Definition 9 (Conditional Expectation) *Let $(\Omega, \mathcal{F}, \mathsf{P})$ be a probability space. Let $A \in \mathcal{F}$ be an event with $\mathsf{P}(A) > 0$. Let $X : \Omega \to \mathbb{R}$ be a* discrete *random variable. The conditional expectation of X conditioned on A is*

$$\mathbf{E}[X \mid A] \triangleq \sum_{a \in \mathsf{Ran}(X)} a \cdot \mathbf{Pr}[X = a \mid A].$$

Let $Y : \Omega \to \mathbb{R}$ be another discrete random variable. The conditional expectation of X conditioned on Y, written as $\mathbf{E}[X \mid Y]$, is a random variable $f_Y : \Omega \to \mathbb{R}$ such that

$$\forall \omega \in \Omega : \ f_Y(\omega) = \mathbf{E}\left[X \,\middle|\, Y^{-1}(Y(\omega))\right] = \mathbf{E}[X \mid Y = Y(\omega)]. \tag{1}$$

Proposition 10

- $\mathbf{E}[X \mid Y]$ *is $\sigma(Y)$-measurable.*
- $\mathbf{E}[\mathbf{E}[X \mid Y]] = \mathbf{E}[f_Y] = \mathbf{E}[X]$.

Proof.

- Since the value of $\mathbf{E}[X \mid Y]$ is determined by $Y(\omega)$, it is clearly $\sigma(Y)$-measurable.
- We compute $\mathbf{E}[f_Y]$ by definition.

$$\begin{aligned}
\mathbf{E}[f_Y] &= \sum_{y \in \mathsf{Ran}(Y)} \mathbf{Pr}[Y = y] \cdot \mathbf{E}[X \mid Y = y] \\
&= \sum_{y \in \mathsf{Ran}(Y)} \mathbf{Pr}[Y = y] \cdot \sum_{x \in \mathsf{Ran}(X)} \mathbf{Pr}[X = x \mid Y = y] \cdot x \\
&= \sum_{x \in \mathsf{Ran}(X)} x \cdot \sum_{y \in \mathsf{Ran}(Y)} \mathbf{Pr}[Y = y] \cdot \mathbf{Pr}[X = x \mid Y = y] \\
&= \sum_{x \in \mathsf{Ran}(X)} x \cdot \sum_{y \in \mathsf{Ran}(Y)} \mathbf{Pr}[X = x \wedge Y = y] \\
&= \sum_{x \in \mathsf{Ran}(X)} x \cdot \mathbf{Pr}[X = x] \\
&= \mathbf{E}[X].
\end{aligned}$$

□

6 *Conditional Expectation for General Random Variables*

The definition of conditional expectation for continuous random variables is more subtle. For example, if $X, Y \sim N(0,1)$ are two independent random variables following standard normal distribution, then intuitively $\mathbf{E}[X \mid Y = 0]$ should be identical to $\mathbf{E}[X]$, which is zero. However, we cannot directly adopt the definition before since $\mathbf{Pr}[Y = 0] = 0$.

Definition 11 *Let* $(\Omega, \mathcal{F}, \mathsf{P})$ *be the probability space. Let* X *be a random variable with* $\mathbf{E}[|X|] < \infty$. *The conditional expectation* $\mathbf{E}[X \mid Y]$ *is a* $\sigma(Y)$*-measurable random variable* f_Y *satisfying*

$$\forall A \in \sigma(Y), \int_A f_Y \, \mathrm{d}P = \int_A X \, \mathrm{d}P.$$

The existence and uniqueness of f_Y follow from Radon-Nikodym theorem.

References

[1] Rick Durrett. *Probability: theory and examples*, volume 49. Cambridge university press, 2019. 1