Lecture-01: Probability Review

1 Probability Review

Definition 1.1. A **probability space** $(\Omega, \mathcal{F}, P)$ consists of set of all possible outcomes called a **sample space** and denoted by Ω, a collection of subsets $\mathcal{F}$ of sample space called **event space**, and a non-negative set function **probability** $P : \mathcal{F} \to [0,1]$, with the following properties.

1. Event space $\mathcal{F}$ is a σ-algebra, that is it contains an empty set and is closed under complements and countable unions.
2. Set function P satisfies $P(\Omega) = 1$, and is additive for countably disjoint events.

An element of the sample space is called an outcome and an element of event space is called an event.

Example 1.2 (Discrete σ-algebra). For a finite sample space Ω, the event space $\mathcal{F} = \{A : A \subseteq \Omega\}$ consists of all subsets of sample space Ω.

Example 1.3 (Borel σ-algebra). If the sample space $\Omega = \mathbb{R}$, then a **Borel** σ-algebra is generated by half-open intervals by complements and countable unions. That is, $\mathcal{B} = \sigma(\{(-\infty, x] : x \in \mathbb{R}\})$. We make the following observations.

1. From closure under complements, the open interval (x, ∞) belong to $\mathcal{B}$ for each $x \in \mathbb{R}$.
2. From closure under countable unions, the open interval $(-\infty, x) = \cup_{n\in\mathbb{N}}(-\infty, x - \frac{1}{n}]$ belongs to $\mathcal{B}$ for each $x \in \mathbb{R}$.
3. From closure under countable intersections, the singleton $\{x\} = \bigcap_{n\in\mathbb{N}}([x - \frac{1}{n}, \infty) \cap (-\infty, x + \frac{1}{n}])$ belongs to $\mathcal{B}$ for each $x \in \mathbb{R}$.

There is a natural order of inclusion on sets through which we can define monotonicity of probability set function P. To define continuity of this set function, we define limits of sets.

Definition 1.4. For a sequence of sets $(A_n : n \in \mathbb{N})$, we define **limit superior** and **limit inferior** of this set sequence respectively as

$$\limsup_n A_n = \bigcap_{n\in\mathbb{N}} \bigcup_{k\geqslant n} A_k, \qquad \liminf_n A_n = \bigcup_{n\in\mathbb{N}} \bigcap_{k\geqslant n} A_k.$$

It is easy to check that $\liminf A_n \subseteq \limsup A_n$. We say that limit of set sequence exists if $\limsup A_n \subseteq \liminf A_n$, and the limit of the set sequence in this case is $\limsup A_n$.

Theorem 1.5. *Probability set function is monotone and continuous.*

Proof. Consider two events $A \subseteq B$ both elements of $\mathcal{F}$, then from the additivity of probability over disjoint events A and $B \setminus A$, we have

$$P(B) = P(A \cup B \setminus A) = P(A) + P(B \setminus A) \geqslant P(A).$$

Monotonicity follows from non-negativity of probability set function, that is since $P(B \setminus A) > 0$. For continuity from below, we take an increasing sequence of sets $(A_n : n \in \mathbb{N})$, such that $A_n \subseteq A_{n+1}$ for all n. Then, it is clear that $A_n \uparrow A = \cup_n A_n$. We can define disjoint sets $(E_n : n \in \mathbb{N})$, where

$$E_1 = A_1, \qquad E_n = A_n \setminus A_{n-1},\ n \geqslant 2.$$

The disjoint sets E_n's satisfy $\cup_{i=1}^n E_i = A_n$ for all $n \in \mathbb{N}$ and $\cup_n E_n = \cup_n A_n$. From the above property and the additivity of probability set function over disjoint sets, it follows that

$$P(A) = P(\cup_n E_n) = \sum_{n\in\mathbb{N}} P(E_n) = \lim_{n\in\mathbb{N}} \sum_{i=1}^{n} P(E_i) = \lim_{n\in\mathbb{N}} P \cup_{i=1}^n E_i = \lim_{n\in\mathbb{N}} P(A_n).$$

For continuity from below, we take decreasing sequence of sets $(A_n : n \in \mathbb{N})$, such that $A_{n+1} \subseteq A_n$ for all n. We can form increasing sequence of sets $(B_n : n \in \mathbb{N})$ where $B_n = A_n^c$. Then, the continuity from above follows from continuity from above. Continuity of probability for general sequence of converging sets follows from the definition of limsup and liminf of sequence of sets and the continuity of probability function from above and below. □

1.1 Independence

Definition 1.6. For a probability space $(\Omega, \mathcal{F}, P)$, two events $A, B \in \mathcal{F}$ are **independent events** if

$$P(A \cap B) = P(A)P(B). \tag{1}$$

Definition 1.7. A collection of events $\mathcal{E} \subseteq \mathcal{F}$ is called a sub-event space if it is a σ-algebra.

Definition 1.8. Two sub-event spaces $\mathcal{G}$ and $\mathcal{H}$ are called independent if any pair of events $(G,H) \in \mathcal{G} \times \mathcal{H}$ are independent. That is,

$$P(G \cap H) = P(G)P(H),\ G \in \mathcal{G}, H \in \mathcal{H}.$$

1.2 Conditional Probability

Definition 1.9. Let $(\Omega, \mathcal{F}, P)$ be a probability space. For events $A, B \in \mathcal{F}$ such that $1 > P(B) > 0$, the conditional probability of event A given event B is defined as

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

Knowing $P(A|B)$, we also know $P(A|B^c)$ if $P(B) < 1$. Indeed, we can compute $P(A|B^c)$ in terms of $P(A|B)$ as

$$P(A|B^c) = \frac{P(A \cap B^c)}{P(B^c)} = \frac{P(A) - P(A \cap B)}{1 - P(B)} = \frac{P(A) - P(B)P(A|B)}{1 - P(B)}. \tag{2}$$

We can check that $P(A \cap \Omega) = P(A)P(\Omega)$ and $P(A \cap \emptyset) = P(A)P(\emptyset)$. Therefore, we can define $P(A|\Omega) = P(A)$ and $P(A|\emptyset) = P(A)$. Hence, if we know the conditional probability on an event $B \in \mathcal{F}$, we know the conditional probability on the event subspace $\{\emptyset, B, B^c, \Omega\}$.

Definition 1.10. We can define conditional probability of event A on a sub-event space $\mathcal{E}$ by the collection $(P(A \mid E) : E \in \mathcal{E})$.

2 Random variables

Definition 2.1. A real valued **random variable** X on a probability space $(\Omega, \mathcal{F}, P)$ is a function $X : \Omega \to \mathbb{R}$ such that for every $x \in \mathbb{R}$, we have

$$X^{-1}(-\infty, x] \triangleq \{\omega \in \Omega : X(\omega) \leqslant x\} \in \mathcal{F}.$$

Recall that the collection $((-\infty, x] : x \in \mathbb{R})$ generates the Borel σ-algebra $\mathcal{B}(\mathbb{R})$. Therefore, it follows that $X^{-1}(\mathcal{B}) \subseteq \mathcal{F}$, since set inverse map X^{-1} preserves complements, unions, and intersections.

Definition 2.2. For a random variable X defined on the probability space $(\Omega, \mathcal{F}, P)$, we define $\sigma(X)$ is the smallest σ-algebra formed by inverse mapping of Borel sets, i.e.

$$\sigma(X) \triangleq \sigma(X^{-1}(-\infty, x] : x \in \mathbb{R}).$$

Note that $\sigma(X)$ is a sub-event space of $\mathcal{F}$ and hence probability is defined for each element of $\sigma(X)$.

Definition 2.3. For a random variable X defined on probability space $(\Omega, \mathcal{F}, P)$, The **distribution function** $F : \mathbb{R} \to [0,1]$ for this random variable X is defined as

$$F(x) = (P \circ X^{-1})(-\infty, x], \text{ for all } x \in \mathbb{R}.$$

Theorem 2.4. *Distribution function F of a random variable X is non-negative, monotone increasing, continuous from the right, and has countable points of discontinuities. Further, if $P \circ X^{-1}(\mathbb{R}) = 1$, then*

$$\lim_{x \to -\infty} F(x) = 0, \qquad \lim_{x \to \infty} F(x) = 1.$$

Proof. Non-negativity and monotonicity of distribution function follows from non-negativity and monotonicity of probability set function, and the fact that for $x_1 < x_2$

$$X^{-1}(-\infty, x_1] \subseteq X^{-1}(-\infty, x_2].$$

Let $x_n \downarrow x$ be a decreasing sequence of real numbers. We take decreasing sets $(A_n \in \mathcal{F} : n \in \mathbb{N})$, where $A_n = X^{-1}(-\infty, x_n] \in \mathcal{F}$. The right continuity of distribution function follows from the continuity from above of probability set functions. $\square$

Example 2.5. One of the simplest family of random variables are indicator functions $\mathbb{1} : \mathcal{F} \times \Omega \to (0,1)$. For each event $A \in \mathcal{F}$, we can define an indicator function as

$$\mathbb{1}_A(\omega) = \begin{cases} 1, & \omega \in A, \\ 0, & \omega \notin A. \end{cases}$$

We make the following observations.

1. $\mathbb{1}_A$ is a random variable for each $A \in \mathcal{F}$. This follows from the fact that

$$\mathbb{1}_A^{-1}(-\infty, x] = \begin{cases} \emptyset, & x < 0, \\ A^c, & x \in [0,1), \\ \Omega, x \geqslant 1. \end{cases}$$

2. The distribution function F for the random variable $\mathbb{1}_A$ is given by

$$F(x) = \begin{cases} 0, & x < 0, \\ P(A^c), & x \in [0,1), \\ 1, & x \geqslant 1. \end{cases}$$

2.1 Expectation

Let $g : \mathbb{R} \to \mathbb{R}$ be a Borel measurable function, i.e. $g^{-1}(-\infty, x] \in \mathcal{B}$ for all $x \in \mathbb{R}$. Then, the **expectation** of $g(X)$ for a random variable X with distribution function F is defined as

$$\mathbb{E}g(X) = \int_{x \in \mathbb{R}} g(x) dF(x).$$

Remark 1. Recall that probabilities are defined only for events. For a random variable X, the probabilities are defined for generating events $X^{-1}(-\infty, x] \in \mathcal{F}$ by $F(x) = P \circ X^{-1}(-\infty, x]$.

Remark 2. The expectation is only defined for random variables. For an event A, the probability $P(A)$ equals expectation of the indicator random variable $\mathbb{1}_A$.

2.2 Random Vectors

Definition 2.6. If $X_1, \dots, X_n$ are random variables defined on the same probability space $(\Omega, \mathcal{F}, P)$, then the vector $X \triangleq (X_1, \dots, X_n)$ is a random mapping $\Omega \to \mathbb{R}^n$ and is called a random vector. Since each X_i is a random variable, the joint event $\cap_{i \in [n]} X_i^{-1}(-\infty, x_i] \in \mathcal{F}$, and the joint distribution of random vector X is defined as

$$F_X(x_1, \dots, x_n) = P\left(\cap_{i \in [n]} X_i^{-1}(-\infty, x_i]\right), \quad \text{for all } x \in \mathbb{R}^n.$$

Definition 2.7. A random variable X is **independent** of the event subspace $\mathcal{E}$, if $\sigma(X)$ and $\mathcal{E}$ are independent event subspaces.

Remark 3. Since $\sigma(X)$ is generated by the collection $(X^{-1}(-\infty, x] : x \in \mathbb{R})$, it follows that X is independent of $\mathcal{E}$ if and any if for all $x \in \mathbb{R}$ and event $E \in \mathcal{E}$,

$$\mathbb{E}[\mathbb{1}_{\{X \leqslant x\}} \mathbb{1}_E] = P(\{X \leqslant x\} \cap E) = P(\{X \leqslant x\})P(E) = \mathbb{E}\mathbb{1}_{\{X \leqslant x\}} \mathbb{E}\mathbb{1}_A.$$

Definition 2.8. Two random variables X, Y are independent if $\sigma(X)$ and $\sigma(Y)$ are independent event subspaces.

Remark 4. Since $\sigma(X)$ and $\sigma(Y)$ are generated by collections $(X^{-1}(-\infty, x] : x \in \mathbb{R})$ and $(Y^{-1}(-\infty, y] : y \in \mathbb{R})$, it follows that the random variables X, Y are independent if and only if for all $x, y \in \mathbb{R}$, we have

$$F_{X,Y}(x, y) = F_X(x) F_Y(y).$$

Definition 2.9. A random vector $X : \Omega \to \mathbb{R}^n$ is independent if the joint distribution is product of marginals. That is,

$$F_X(x) = \prod_{i=1}^{n} F_{X_i}(x_i), \quad \text{for all } x \in \mathbb{R}^n.$$