Lecture-02: Probability Review

1 Probability Review

Definition 1.1 (σ-algebra). A collection $\mathcal{A}$ of subsets of a set Ω is called a σ-algebra if
(i) it contains the empty set,
(ii) it is closed under complements, and
(iii) it is closed under countable unions.

Definition 1.2. A **probability space** $(\Omega, \mathcal{F}, P)$ consists of
(i) set of all possible outcomes called a **sample space** denoted by Ω,
(ii) a σ-algebra over sample space called **event space** denoted by $\mathcal{F}$, and
(iii) a set function **probability** denoted by $P : \mathcal{F} \to [0,1]$ such that (a) P is non-negative, (b) $P(\Omega) = 1$, and (c) P 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*.

Definition 1.3. A collection of events $\mathcal{E} \subseteq \mathcal{F}$ is called a sub-event space if it is a σ-algebra over Ω.

Definition 1.4. For a family of events $\mathcal{A} \subseteq \mathcal{F}$, the sub-event space generated by the family $\mathcal{A}$ is the smallest sub-event space containing the family $\mathcal{A}$ and denoted by $\sigma(\mathcal{A})$.

Remark 1. The sub-event space $\sigma(\mathcal{A})$ contains all the elements of $\mathcal{A}$ and the complements and countable unions of generated sets.

Example 1.5 (Discrete σ-algebra). For a finite sample space Ω, the discrete event space $\mathcal{P}(\Omega) \triangleq \{A : A \subseteq \Omega\}$ consists of all subsets of sample space Ω, and is sometimes also denoted by 2^{Ω}. In particular, $\mathcal{P}(\Omega)$ consists of singleton set $\{\omega\}$ for each outcome $\omega \in \Omega$.

Example 1.6 (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, $\mathcal{B}(\mathbb{R}) \triangleq \sigma(\{(-\infty, x] : x \in \mathbb{R}\})$. We make the following observations.

1. From closure under complements, the open interval (x, ∞) belong to $\mathcal{B}(\mathbb{R})$ 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 $\mathcal{B}(\mathbb{R})$ for each $x \in \mathbb{R}$.
3. From closure under complements, half-closed intervals $[x, \infty)$ belongs to $\mathcal{B}(\mathbb{R})$ for each $x \in \mathbb{R}$.
4. From closure under finite intersections, the closed set $[x, y] = [x, \infty) \cap (-\infty, y]$ belongs to $\mathcal{B}(\mathbb{R})$ for each $x, y \in \mathbb{R}$.
5. From closure under countable intersections, the singleton $\{x\} = \bigcap_{n\in\mathbb{N}}([x - \frac{1}{n}, x + \frac{1}{n}])$ belongs to $\mathcal{B}(\mathbb{R})$ for each $x \in \mathbb{R}$.

1.1 Limits of sets and continuity of probability

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

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

Remark 2. It is easy to check that $\liminf A_n \subseteq \limsup A_n$. To see this, we observe that $\cap_{k \geqslant n} A_k \subseteq A_k$ for all $k \geqslant n$. This implies that $\cap_{k \geqslant n} A_k \subseteq \cup_{k \geqslant \ell} A_k$ for all $n, \ell \in \mathbb{N}$ and hence $\cap_{k \geqslant n} A_k \subseteq \cap_{\ell \in \mathbb{N}} \cup_{k \geqslant \ell} A_k$ for all $n \in \mathbb{N}$ and the result follows.

Definition 1.8. We say that limit of set sequence $(A_n : n \in \mathbb{N})$ exists if $\limsup A_n \subseteq \liminf A_n$, and the limit of the set sequence in this case is denoted as $\lim A_n \triangleq \limsup A_n = \liminf A_n$.

Theorem 1.9. *Probability set function is monotone and continuous.*

Proof. Consider a probability set function $P : \mathcal{F} \to [0,1]$.

Monotonicity: 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 $P(B \setminus A) > 0$.

Continuity: For continuity from below, we take a nondecreasing sequence of sets $(A_n : n \in \mathbb{N})$, such that $A_n \subseteq A_{n+1}$ for all n. We observe that $A_\infty \triangleq \lim_n A_n = \cup_{n \in \mathbb{N}} A_n$ and $A_n \uparrow A_\infty$. We can define disjoint sets $(E_n : n \in \mathbb{N})$, where $E_1 = A_1$ and $E_n = A_n \setminus A_{n-1}$ for all $n \geqslant 2$. 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

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

For continuity from below, we take a nonincreasing sequence of sets $(A_n : n \in \mathbb{N})$, such that $A_{n+1} \subseteq A_n$ for all n. We can form nondecreasing sequence of sets $(B_n : n \in \mathbb{N})$ where $B_n = A_n^c$. Then, the continuity from below follows from the 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.

□

1.2 Independence

Definition 1.10. For a probability space $(\Omega, \mathcal{F}, P)$, two events $A, B \in \mathcal{F}$ are **independent events** if

$$P(A \cap B) = P(A)P(B).$$

Definition 1.11. 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, for all $G \in \mathcal{G}$ and $H \in \mathcal{H}$, we have

$$P(G \cap H) = P(G)P(H).$$

1.3 Conditional Probability

Definition 1.12. 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

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

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

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

Remark 3. Recall that the collection $\{(-\infty, x] : x \in \mathbb{R}\}$ generates the Borel σ-algebra $\mathcal{B}(\mathbb{R})$. Therefore, it follows that $X^{-1}(\mathcal{B}(\mathbb{R})) \subseteq \mathcal{F}$, 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 the inverse mapping of Borel sets, i.e.

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

Definition 2.3. A random variable X is **independent** of an event subspace $\mathcal{E}$, if $\sigma(X)$ and $\mathcal{E}$ are independent event subspaces.

Definition 2.4. Two random variables X,Y defined on the same probability space are independent if $\sigma(X)$ and $\sigma(Y)$ are independent event subspaces.

Definition 2.5. For a random variable X defined on probability space $(\Omega,\mathcal{F},P)$, the corresponding **distribution function** $F : \mathbb{R} \to [0,1]$ is defined as

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

Remark 4. Recall that probabilities are defined only for events. We note that $\sigma(X)$ is a sub-event space of $\mathcal{F}$ and hence probability is defined for each element of $\sigma(X)$. In particular for a random variable X, the probabilities are defined for generating events $X^{-1}(-\infty,x] \in \mathcal{F}$, and denoted by $F(x) = P \circ X^{-1}(-\infty,x]$ for each $x \in \mathbb{R}$. It follows that for each event $A \in \sigma(X)$, we can find the probability $P(A)$ in terms of F.

Remark 5. 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 and Y are independent if and only if for all $x,y \in \mathbb{R}$, we have

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

Theorem 2.6. *Distribution function F of a random variable $X : \Omega \to \mathbb{R}$ is nonnegative, monotone nondecreasing, 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. Nonnegativity and monotonicity of distribution function follows from nonnegativity 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 $x_n \downarrow x_\infty$ be a decreasing sequence of real numbers. We take decreasing sets $A \in \mathcal{F}^{\mathbb{N}}$, where $A_n \triangleq X^{-1}(-\infty,x_n] \in \mathcal{F}$ for $n \in \mathbb{N}$. Then $A_n \downarrow A_\infty$, and the right continuity of distribution function follows from the continuity from above of probability set functions. Countable discontinuities follow from the fact that $\lim_{x\to\infty} F_X(x) \leqslant 1$. By taking sequences $a_n \downarrow -\infty$ and $b_n \uparrow \infty$, we define sequence of monotone sets $A_n \triangleq X^{-1}(-\infty,a_n] \downarrow \emptyset$ and $B_n \triangleq X^{-1}(-\infty,b_n] \uparrow \Omega$. The result follow from the continuity of probability function. □

Example 2.7. 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. $g^{-1}(-\infty, x] \in \mathcal{B}(\mathbb{R})$ 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 6. The expectation is only defined for random variables. For an event A, the probability $P(A)$ equals expectation of the indicator random variable $\mathbb{1}_A$.

Remark 7. 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 $E \in \mathcal{E}$,

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

3 Random Vectors

Definition 3.1. 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**.

Remark 8. Since each X_i is a random variable, the joint event $\cap_{i \in [n]} X_i^{-1}(-\infty, x_i] \in \mathcal{F}$. That is, $\sigma(X) \triangleq \sigma(\cap_{i=1}^n \left\{ X_i^{-1}(-\infty, x_i] : x \in \mathbb{R}^n \right\})$.

Definition 3.2. We define projection operators $\pi_i : \mathbb{R}^n \to \mathbb{R}$ as $\pi_i(x) \triangleq x_i$ for any vector $x \in \mathbb{R}^n$.

Remark 9. We observe that $\pi_i^{-1}(-\infty, x_i] = \{y \in \mathbb{R}^n : y_i \leqslant x_i\}$.

Proposition 3.3. *A random mapping $X : \Omega \to \mathbb{R}^n$ is a random vector if $X^{-1}(\pi_i^{-1}(-\infty, x_i]) \in \mathcal{F}$ for all $i \in [n]$ and $x_i \in \mathbb{R}$.*

Proof. Recall that a random mapping $X : \Omega \to \mathbb{R}^n$ is a random vector if $\pi_i(X)$ is a random variable for each $i \in [n]$. Since $X_i^{-1}(-\infty, x_i] = X^{-1} \circ \pi_i^{-1}(-\infty, x_i]$, the result follows. ☐

Remark 10. We observe that $\sigma(X) = \sigma(X^{-1} \circ \pi_i^{-1}(-\infty, x_i] : x_i \in \mathbb{R}, i \in [n])$.

Definition 3.4. The joint distribution of random vector X is defined for all $x \in \mathbb{R}^n$ as

$$F_X(x_1, \dots, x_n) \triangleq P\left(\cap_{i \in [n]} X_i^{-1}(-\infty, x_i] \right).$$

Definition 3.5. A random vector $X : \Omega \to \mathbb{R}^n$ is independent if the joint distribution is product of marginals. That is,

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

Definition 3.6. Two random vectors X, Y defined on the same probability space are independent if $\sigma(X)$ and $\sigma(Y)$ are independent event subspaces.