Probability and Statistics I

Changmin Yu

February 2020

1 Definitions

- **Sample space**: the set of all the outcomes of a random experiment, usually denoted by Ω.
- $\mathcal{F}$ is the σ**-algebra** of Ω, sets of subsets of Ω, sets of events.
- **Probability measure**: a function $\mathbb{P} : \mathcal{F} \to \mathbb{R}$ satisfying:
  - $\mathbb{P}(A) \geq 0$, for all $A \in \mathcal{F}$
  - $\mathbb{P}(\Omega) = 1$
  - If $A_1, A_2, ...$ is a countable sequence of disjoint events, then

$$\mathbb{P}(\cup_i A_i) = \sum_i \mathbb{P}(A_i) \tag{1}$$

- Let B be an event with non-zero probability. The **conditional probability** of any event A given B is defined as

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \tag{2}$$

- Two events are called **independent** if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$
- A **random variable** X is a function $X : \Omega \to \mathbb{R}$.
- A **cumulative distribution function (CDF)** is a function $F_X : \mathbb{R} \to [0, 1]$, which specifies a probability measure as

$$F_X(x) = \mathbb{P}(X \leq x) \tag{3}$$

  A CDF satisfies the following properties:

  - $0 \leq F_X(x) \leq 1$
  - $\lim_{x\to-\infty} F_X(x) = 0$
  - $\lim_{x\to\infty} F_X(x) = 1$

- $x \leq y \Rightarrow F_X(x) \leq F_X(y)$

- When a random variable X takes on a finite set of possible values, i.e., X is a discrete random variable, we could directly specify the probability of each value that the random variable can take. A **probability mass function (PMF)** is a function $p_X : \Omega \to \mathbb{R}$ such that

$$p_X(x) = \mathbb{P}(X = x) \tag{4}$$

  A PMF satisfies the following properties:

  - $0 \leq p_X(x) \leq 1$
  - $\sum_{x \in \mathcal{X}} p_X(x) = 1$
  - $\sum_{x \in A} p_X(x) = \mathbb{P}(X \in A)$

- For continuous random variables, we could define the **probability density function (PDF)** as the derivative of the CDF, i.e.

$$f_X(x) = \frac{dF_X(x)}{dx} \tag{5}$$

  A PDF satisfies the following properties:

  - $f_X(x) \geq 0$
  - $\int_{-\infty}^{\infty} f_X(x)\, dx = 1$
  - $\int_{x \in A} f_X(x)\, dx = \mathbb{P}(X \in A)$

- Let X be a discrete random variable with PMF $p_X(x)$, and $g : \mathbb{R} \to \mathbb{R}$ is an arbitrary function. Then $g(X)$ becomes a random variable, and we could define the **expectation** of $g(X)$ as

$$\mathbb{E}[g(X)] = \sum_{x \in X} g(x) p_X(x) \tag{6}$$

  Similarly, if X is a continuous random variable with PDF $f_X(x)$, then the expected value of $g(X)$ is

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx \tag{7}$$

- The **variance** of a random variable X is defined as

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] \tag{8}$$

  We then immediately have

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \tag{9}$$