THE CHINESE UNIVERSITY OF HONG KONG
Department of Mathematics
MATH4240 - Stochastic Processes - 2024/25 Term 2

Chapter 0 Review on Probability

I. Probability Space. A probability space is a triple $(\Omega, \mathcal{F}, P)$.

- Ω is a set called the sample space. An element $\omega \in \Omega$ is called an outcome.
- $\mathcal{F}$ is a nonempty set of subsets of Ω, called the event space (whose elements called events), such that
  (a) $\Omega \in \mathcal{F}$.
  (b) If $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$.
  (c) If $A_i \in \mathcal{F}$, $i = 1, 2, \cdots$, then $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$.

  A collection of subsets with these three properties is called a σ-algebra or σ-field.
- $P : \mathcal{F} \to [0, 1]$ is called the probability measure over the event space $\mathcal{F}$, satisfying
  (a) $P(\Omega) = 1$.
  (b) $0 \le P(A) \le 1$, $\forall\, A \in \mathcal{F}$.
  (c) $P(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$, $\forall\, \{A_i\}_{i=1}^{n}$ (n can be finite or infinite) which is disjoint.

Conditional probability: Let A, B be two events. The probability that B happens given that A occurs is denoted by

$$P(B|A) := \frac{P(A \cap B)}{P(A)} \quad \text{for } P(A) \neq 0. \tag{1}$$

To compute $P(A \cap B)$, one may use formulas

$$P(A \cap B) = P(B|A)P(A) \text{ or } P(A \cap B) = P(A|B)P(B). \tag{2}$$

We say A and B are independent if $P(B|A) = P(B)$, i.e.

$$P(A \cap B) = P(A)P(B). \tag{3}$$

Let A be fixed, $P_A(\cdot) := P(\,\cdot\,|A)$ is called the conditional probability measure.

For any event B, to compute $P(B)$, we may first find *all* possible events that cause B, for instance, Ω is the union of disjoint events $A_1, \cdots, A_n$ and under this disjoint decomposition we also know how to compute $P(B|A_i)$ and $P(A_i)$ for each i. Then

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i). \tag{4}$$

Moreover, we can also compute the probability of each cause event A_i subject to the caused event B in the way that

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)}. \tag{5}$$

This is the so-called Bayes' formula.

II. Random Variables and Distributions. A random variable (r.v.) X on $(\Omega, \mathcal{F}, P)$ is a function from Ω to $\mathbb{R}$, that is to assign each outcome with a real value. X is called a discrete r.v. if the range of X is a discrete set. X is called a continuous r.v. if the range of X is an interval of $\mathbb{R}$, for instance.

Discrete r.v.: Assume that the range of X is given by $S = \{k\}_{k=0}^{N}$ (N can be finite or infinite). S is called the state space.

$$p_k = P(X = k), \quad k = 0, 1, \cdots, N, \tag{6}$$

is called the probability density function (p.d.f.) of X. Here $X = k$ means the event

$$\{X = k\} = \{\omega \in \Omega : X(\omega) = k\} \in \mathcal{F}. \tag{7}$$

Note

$$0 \leq p_k \leq 1, \quad \sum_{k \in S} p_k = 1. \tag{8}$$

The following examples are important:

(a) Binomial r.v.: It means a r.v. X having the p.d.f.:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad 0 \leq k \leq n. \tag{9}$$

For instance, we perform n independent trials. At each trial, the success probability is p and the failure probability is $1 - p$. Let X be the number of successes in n trials. Then, X is a binomial r.v. given as above.

(b) Geometric r.v.: Let X denote the number of trials for the first success, then

$$P(X = k) = p(1-p)^{k-1}, \quad k = 1, 2, \ldots$$

is the probability that the first occurrence of success requires k independent trials.

(c) Poisson r.v.: It means a r.v. X having the p.d.f.:

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \cdots, \tag{10}$$

where $\lambda > 0$ is called the rate parameter. There are many models obeying the Poisson distribution. A general model is given as follows. An event can occur 0, 1, 2, $\cdots$ times in an interval. The average number of events in an interval is designated $\lambda > 0$. Let X

be the NO of events observed in an interval. Then, X is a Poisson r.v. given as above. For instance, X may denote the NO of arrivals in a unit time with $\lambda > 0$ meaning the rate of arrivals. Note that given $\lambda > 0$, by letting $n \to \infty$ with $np = \lambda$, the binomial distribution converges to the Poisson distribution, i.e.

$$\lim_{n\to\infty, np=\lambda>0} \binom{n}{k} p^k (1-p)^{n-k} = e^{-\lambda}\frac{\lambda^k}{k!}, \tag{11}$$

for each $k = 0, 1, 2, \cdots$.

Continuous r.v.: Assume that there is a nonnegative function $f(\cdot)$ such that

$$P(a \le X \le b) = \int_a^b f(t)\,dt, \quad -\infty < a < b < \infty. \tag{12}$$

Then, X is a continuous r.v. and f is called the p.d.f. of X. Here $a \le X \le b$ means the event $\{a \le X \le b\} = \{\omega \in \Omega : a \le X(\omega) \le b\}$. Note

$$f(x) \ge 0, \quad \int_{-\infty}^{\infty} f(x)\,dx = 1. \tag{13}$$

The following are important examples:

(a) Uniform p.d.f.:

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{if } a \le x \le b, \\ 0 & \text{otherwise.} \end{cases} \tag{14}$$

(b) Exponential p.d.f.:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \ge 0, \\ 0 & \text{otherwise.} \end{cases} \tag{15}$$

(c) Normal p.d.f.:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} := N(\mu, \sigma^2). \tag{16}$$

See below for the meaning of μ and $\sigma > 0$. $N(0,1)$ is called the standard normal distribution.

III. Expectation and Variance. The expectation (or mean) of X is defined by

$$\mu = E(X) := \sum_{k\in S} k p_k \quad \text{or} \quad \int_{-\infty}^{\infty} x f(x)\,dx. \tag{17}$$

The 2nd moment of X is defined by

$$E(X^2) := \sum_{k\in S} k^2 p_k \quad \text{or} \quad \int_{-\infty}^{\infty} x^2 f(x)\,dx. \tag{18}$$

The variance of X is defined by

$$\sigma^2 = \text{Var}\,(X) := E(X-\mu)^2 = E(X^2) - \mu^2. \tag{19}$$

<u>Conditional Expectation</u>: In the discrete case, suppose that (X, Y) has a joint p.d.f.:

$$p(x_i, y_j) = P(X = x_i, Y = y_j). \tag{20}$$

Then,

$$E(Y|X = x_i) = \sum_j y_j P(Y = y_j | X = x_i) = \sum_j y_j \frac{p(x_i, y_j)}{p(x_i)}, \tag{21}$$

where $p(x_i) := \sum_j p(x_i, y_j)$ is the p.d.f. of X. Therefore, fixing Y, we may regard $E(Y|X)$ as a r.v. with the p.d.f. given above. In the continuous case, suppose that (X, Y) has a joint p.d.f. $f(x, y)$ such that

$$P(X \le x, Y \le y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v)\, dv du. \tag{22}$$

Then,

$$E(Y|X = x) = \int_{-\infty}^{\infty} y \frac{f(x, y)}{f(x)}\, dy, \tag{23}$$

where $f(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$ is the p.d.f. of X. Similar to the discrete case, fixing Y, $E(Y|X)$ can be regarded as a continuous r.v. with the p.d.f. given above.

IV. Sequence of r.v.'s By repeating a random experiment at time $n = 0, 1, \cdots$ independently, we obtain a sequence of *independent and identically distributed* (i.i.d.) r.v. $\{X_n\}_{n=0}^{\infty}$. To describe $\{X_n\}_{n=0}^{\infty}$, we have the following two basic theorems in probability:

- Law of Large Numbers: Assume $\mu = E(X_n)$ for each n. The weak law of large numbers says that for any $\epsilon > 0$,

  $$\lim_{n\to\infty} P\left(\left|\frac{X_0 + \cdots + X_{n-1}}{n} - \mu\right| \ge \epsilon\right) = 0. \tag{24}$$

  The strong law of large numbers says that

  $$P\left(\lim_{n\to\infty} \frac{X_0 + \cdots + X_{n-1}}{n} = \mu\right) = 1. \tag{25}$$

- Central Limit Theorem: Assume $\mu = E(X_n)$ and $\sigma^2 = \text{Var}(X_n)$ for each n. The central limit theorem says that the p.d.f. of

  $$\frac{X_0 + \cdots + X_{n-1} - n\mu}{\sigma\sqrt{n}} \tag{26}$$

  tends to the standard normal p.d.f. $N(0, 1)$ as $n \to \infty$.

However, in many cases $\{X_n\}_{n=0}^{\infty}$ may not be independent, and indeed there exists a sort of dependence relation. In general, $\{X_n\}_{n=0}^{\infty}$ is called a (discrete) stochastic process and $\{X_t\}_{t\ge 0}$ is called a continuous stochastic process. The goal of this elementary course is to consider the "Markov" process (to be defined) in the discrete and continuous time.

——End, updated on Jan 9th——