Probability Theory

It is a truth very certain that, when it is not in our power to determine what is true, we ought to follow what is most probable. - Rene Descartes

1. Probability Theory Refresher

Probability is defined as the triplet $(\Omega, \mathcal{E}, P)$, where

- Ω is the *sample space*. It is the set of all possible mutually exclusive and all possible outcomes of a specified experiment. Each possible outcome ω of the set is called a *sample point*.
- $\mathcal{E}$ is a family of events. $\mathcal{E} = \{A, B, C, \ldots\}$, where each event is a set of sample points $\{\omega\}$. An event is an outcome of interest.
- P is a real-valued mapping (function) defined on $\mathcal{E}$. $P(A)$ is said to be the probability of the event A, provided the following axioms are satisfied.

[A_1] For any event A, $P(A) \geq 0$.

[A_2] $P(\Omega) = 1$.

[A_3] If $A \cap B = \varnothing$, then $P(A \cup B) = P(A) + P(B)$.

[A_4] If $A_j \cap A_k = \varnothing, j \neq k$, where $j, k \in \{1, 2, 3, \ldots\}$, then $P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \ldots$

Note that the axiom [A_4] is superfluous if the sample space Ω is finite. Observe the following remarks.

Remarks Let A, and B be any events. Then

1. $P(\varnothing) = 0$, where $\varnothing$ is called the null event.
2. $P(A^c) = (1 - P(A))$.
3. $P(A) \leq P(B)$, if $A \subset B$.
4. $P(A \cup B) = (P(A) + P(B) - P(A \cap B))$. □

Definition *The conditional probability of an event A, given event B has occurred, is given by*:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}, \qquad P(B) \neq 0$$

□

Definition *Events A and B are said to be independent if and only if*

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

□

If the above relationship does not hold, then the events A and B are said to be dependent.

2. Random variables

Definition *A random variable is a function X that maps an event $\omega \in \Omega$ into the real line. That is $X(\omega) \in \mathbb{R}$.* □

The random variable is often denoted as X.

Definition *The distribution function $F_X(.)$ of the random variable X is defined for $x \in \mathbb{R}$ as $F_X(x) = P(X \leq x)$.* □

It is sometimes referred to as the cumulative distribution function. This is a monotonically increasing function. That is, if $x \leq y$ then $F_X(x) \leq F_X(y)$. Also $F_X(x)$ tends to 0 as x tends to $-\infty$, and $F_X(x)$ tends to 1 as x tends to ∞.

Definition *A random variable X is discrete, if its set of possible values is countable. If the random variable X takes on values $x_j, j = 1, 2, 3, \ldots$, then the probabilities $P(X = x_j) = p_X(x_j), j = 1, 2, 3, \ldots$, are called the probability mass function of the random variable X. The probability mass function should satisfy*

$$\begin{aligned} p_X(x_j) &\geq 0, \quad j = 1, 2, 3, \ldots \\ \sum_{j \in \mathbb{P}} p_X(x_j) &= 1 \end{aligned}$$

□

Also $F_X(x) = \sum_{x_j \leq x} p_X(x_j)$.

Definition *A random variable X is said to be continuous, if $X(\Omega)$ is a continuum of numbers. It is assumed that there exists a piecewise continuous function $f_X(.)$ that maps real numbers into real numbers such that*

$$P(a \leq x \leq b) = \int_a^b f_X(x)\, dx$$

The function $f_X(x)$ is called the probability density function. It satisfies the following conditions:

$$f_X(x) \geq 0 \text{ and } \int_{\mathbb{R}} f_X(x)\, dx = 1$$

□

If X is a continuous random variable, then $F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt$. It follows that

$$f_X(x) = \frac{d}{dx} F_X(x)$$

3. Jointly Distributed Random Variables

Jointly distributed random variables are initially defined for two random variables. It can then be extended conveniently to N random variables.

Definitions. *Let X and Y be jointly distributed random variables which take real values.*

1. *Joint distributions.*

   (a) *The joint cumulative distribution function of two random variables X and Y is $F_{X,Y}(\cdot,\cdot)$, where*

   $$F_{X,Y}(x,y) = P(X \le x, Y \le y)$$

   (b) *Let the two random variables X and Y be discrete. The joint probability mass function of the two random variables X and Y is $p_{X,Y}(\cdot,\cdot)$, where*

   $$p_{X,Y}(x,y) = P(X = x, Y = y)$$

   (c) *Let the two random variables X and Y be continuous. The random variables X and Y are jointly continuous if there exists a function $f_{X,Y}(\cdot,\cdot)$ such that*

   $$P\left(X \in \widetilde{A}, Y \in \widetilde{B}\right) = \int_{\widetilde{B}} \int_{\widetilde{A}} f_{X,Y}(x,y)\,dxdy$$

   where $\widetilde{A}$ and $\widetilde{B}$ are any subsets of real numbers. The function $f_{X,Y}(\cdot,\cdot)$ is called the joint probability density function.

2. *Marginal distributions.*

   (a) *As y tends to $\infty, F_{X,Y}(x,y)$ tends to $F_X(x)$. Similarly as x tends to ∞, $F_{X,Y}(x,y)$ tends to $F_Y(y)$. $F_X(\cdot)$ and $F_Y(\cdot)$ are called marginal cumulative distribution functions of X and Y respectively.*

   (b) *Let X and Y be both discrete random variables with joint probability mass function $p_{X,Y}(\cdot,\cdot)$. Then*

   $$p_X(x) = \sum_y p_{X,Y}(x,y), \quad \text{and} \quad p_Y(y) = \sum_x p_{X,Y}(x,y)$$

   where $p_X(\cdot)$ and $p_Y(\cdot)$ are called the marginal mass functions of X and Y respectively.

(c) *Let X and Y be both continuous random variables with joint probability density function $f_{X,Y}(\cdot,\cdot)$. Then*

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy, \quad \textit{and} \quad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx$$

where $f_X(\cdot)$ and $f_Y(\cdot)$ are called marginal probability density functions of X and Y respectively.

3. *Conditional distributions.*

(a) *Let X and Y be both discrete random variables with joint probability mass function $p_{X,Y}(\cdot,\cdot)$. Conditional probability mass function of X, given $Y = y$ and $p_Y(y) > 0$, is defined by*

$$p_{X|Y}(x \mid y) = P(X = x \mid Y = y), \qquad \forall\, x$$

(b) *Let X and Y be both continuous random variables with joint probability density function $f_{X,Y}(\cdot,\cdot)$. Conditional probability density function of X, given $Y = y$ and $f_Y(y) > 0$ is defined by*

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}, \qquad \forall\, x$$

□

4. Expectation

Definition *The expectation or mean of a random variable X denoted by $E(X)$ is:*

$$E(X) = \int_{-\infty}^{\infty} x\,dF_X(x)$$

Specifically:

1. *If X is a discrete random variable.*

$$E(X) = \sum_{x:p_X(x)>0} x\,p_X(x)$$

2. *If X is a continuous random variable.*

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)\,dx$$

provided the integral exists. □

5. Variance, Standard Deviation and Covariance

Definition *The variance $Var(X)$ of a random variable X, is*:

$$\begin{aligned} Var(X) &= E\left((X - E(X))^2\right) \\ &= E\left(X^2\right) - (E(X))^2 \end{aligned}$$

□

Definition *The standard deviation σ_X of a of a random variable X is*:

$$\sigma_X = \sqrt{Var(X)}$$

Definition *The covariance $Cov(X, Y)$ of a random variables X and Y is*:

$$\begin{aligned} Cov(X,Y) &= E\left((X - E(X))(Y - E(Y))\right) \\ &= E(XY) - E(X)E(X) \end{aligned}$$

□

Definition *Let σ_X and σ_Y be the standard deviation of the random variables X and Y respectively. Let $\sigma_X \neq 0$ and $\sigma_Y \neq 0$. The correlation coefficient $Cor(X, Y)$ of these random variables is*:

$$Cor(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

□

6. Moment Generating Function

Definition. *The moment generating function of random variable X is given by $\mathcal{M}_X(t) = E\left(e^{tX}\right)$.*

(a) *If X is discrete random variable, then $\mathcal{M}_X(t) = \sum_x e^{tx} p_X(x)$.*

(b) *If X is continuous random variable, then $\mathcal{M}_X(t) = \int_{-\infty}^{\infty} e^{tx} f_X(x)\, dx$.*

□

7. Independent Random Variables

Definition *Random variables X and Y are stochastically independent (or simply independent) random variables if*

$$F_{X,Y}(x,y) = F_X(x) F_Y(y), \quad \forall\ x, y \in \mathbb{R}$$

□

Remarks Let X and Y be independent random variables. Then

1. $E(XY) = E(X)E(Y)$. Note that the reverse is not true. That is, $E(XY) = E(X)E(Y)$ does not imply independence of X and Y.
2. $Var(X+Y) = Var(X) + Var(Y)$
3. $Cov(X,Y) = 0$. □

8. Examples of Some Distributions

Some examples of discrete and continuous time distributions are given.

8.1. Discrete Time Distributions.

Bernoulli Distribution

A random variable X is said to have a Bernoulli distribution, if the probability mass function of X is given by

$$p_X(x) = \begin{cases} (1-p), & \text{if } x = 0 \\ p, & \text{if } x = 1 \end{cases}$$

where $0 \leq p \leq 1$. Also $E(X) = p$, $Var(X) = p(1-p)$.

Binomial Distribution

A random variable X is said to have a binomial distribution, if the probability mass function of X is given by

$$p_X(x) = \begin{cases} \begin{pmatrix} n \\ x \end{pmatrix} p^x q^{n-x}, & \text{if } x = 0, 1, 2, \ldots, n \\ 0, & \text{otherwise} \end{cases}$$

where $0 \leq p \leq 1, q = (1-p)$, and $n \in \mathbb{P}$. Also $E(X) = np$, $Var(X) = npq$. Note that if $n = 1$, then we obtain a Bernoulli distribution.

Geometric Distribution

A random variable X is said to have a geometric distribution, if the probability mass function of X is given by

$$p_X(x) = \begin{cases} pq^x, & \text{if } x \in \mathbb{N} \\ 0, & \text{otherwise} \end{cases}$$

where $0 < p \leq 1, q = (1-p)$. Also $E(X) = q/p$, $Var(X) = q/p^2$.

If X is a geometrically distributed random variable, with parameter p, then

$$P(X \geq m+n \mid X \geq m) = P(X \geq n) \quad \text{for } m, n \in \mathbb{N}$$

Poisson Distribution

A random variable X is said to have a Poisson distribution, if the probability mass function of X is given by

$$p_X(x) = \begin{cases} e^{-\lambda}\frac{\lambda^x}{x!}, & \text{if } x = 0, 1, 2, \ldots \\ 0, & \text{otherwise} \end{cases}$$

where $0 < \lambda$. Also $E(X) = \lambda$, $Var(X) = \lambda$.

8.2. Continuous Time Distributions.

Uniform Distribution

A random variable X is said to have a uniform distribution, if the probability density function of X is given by

$$f_X(x) = \begin{cases} \frac{1}{(b-a)}, & \text{if } x \in [a, b] \\ 0, & \text{otherwise} \end{cases}$$

where $a, b \in \mathbb{R}$, and $a < b$. Also $E(X) = (a+b)/2$, $Var(X) = (b-a)^2/12$. Also

$$F_X(x) = \begin{cases} 0, & \text{if } x < a \\ \frac{(x-a)}{(b-a)}, & \text{if } x \in [a, b] \\ 1, & \text{if } x > b \end{cases}$$

Exponential Distribution

A random variable X is said to have a exponential distribution, if the probability density function of X is given by

$$f_X(x) = \begin{cases} 0, & \text{if } x \in (-\infty, 0) \\ \lambda e^{-\lambda x}, & \text{if } x \in [0, \infty) \end{cases}$$

where $\lambda > 0$. Also $E(X) = 1/\lambda$, $Var(X) = 1/\lambda^2$. And

$$F_X(x) = \begin{cases} 0, & \text{if } x \in (-\infty, 0) \\ \left(1 - e^{-\lambda x}\right), & \text{if } x \in [0, \infty) \end{cases}$$

If X is an exponentially distributed random variable, with parameter λ, then

$$P(X \geq t + u \mid X \geq u) = P(X \geq t) \qquad \text{for } t, u \in [0, \infty)$$

This equation is said to represent the memoryless property of the exponential distribution. Exponential distribution is the only continuous time distribution, that possesses the memoryless property.

Normal Distribution

A random variable X is said to have a normal (or Gaussian) distribution, if the probability density function of X is given by

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \qquad x \in \mathbb{R}$$

where $\mu \in \mathbb{R}$ and $\sigma > 0$. Also $E(X) = \mu$, $Var(X) = \sigma^2$.

If a normal random variable has $\mu = 0$, and $\sigma = 1$, then it is called a standard normal random variable. The probability density function $\phi(x)$, and its cumulative distribution function $\Phi(x)$ are given by

$$\begin{aligned} \phi(x) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \qquad x \in \mathbb{R} \\ \Phi(x) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{y^2}{2}} dy, \qquad x \in \mathbb{R} \end{aligned}$$

The function $\Phi(x)$ is generally evaluated numerically.

Cauchy Distribution

A random variable X is said to have a Cauchy distribution, if the probability density function of X is given by

$$f_X(x) = \frac{1}{\pi\beta\left[1 + \left(\frac{x-\alpha}{\beta}\right)^2\right]}, \qquad x \in \mathbb{R}$$

where $\alpha \in \mathbb{R}$, and $\beta \in (0, \infty)$. The mean and variance of this random variable do not exist. Its cumulative distribution function $F_X(x)$ is

$$F_X(x) = \frac{1}{2} + \frac{1}{\pi} \tan^{-1}\left(\frac{x-\alpha}{\beta}\right), \qquad x \in \mathbb{R}$$

Pareto Distribution

A random variable X is said to have a Pareto distribution, if the probability density function of X is given by

$$f_X(x) = \begin{cases} 0, & \text{if } x \le x_0 \\ \frac{\alpha}{x_0}\left(\frac{x_0}{x}\right)^{\alpha+1}, & \text{if } x > x_0 \end{cases}$$

where $x_0, \alpha \in (0, \infty)$. Note that x_0 is called the location parameter, and α is called the shape parameter. Also

$$\begin{aligned} E(X) &= \frac{\alpha x_0}{(\alpha - 1)}, \text{ for } 1 < \alpha \\ Var(X) &= \frac{\alpha x_0^2}{(\alpha - 2)} - \left(\frac{\alpha x_0}{\alpha - 1}\right)^2, \text{ for } 2 < \alpha \\ F_X(x) &= \begin{cases} 0, & \text{if } x \le x_0 \\ 1 - \left(\frac{x_0}{x}\right)^{\alpha}, & \text{if } x > x_0 \end{cases} \end{aligned}$$

Pareto distribution decays much more slowly than the exponential distribution. This distribution has recently found application in modeling internet traffic.