Introduction to Probability
Chapter 3: Random Variable

Dr. Nitin Gupta
Department of Mathematics
Indian Institute of Technology Kharagpur,
Kharagpur - 721 302, INDIA.

Outline

1. Random variable
2. Probability mass function (PMF)
3. Probability density function (PDF)
4. Cummulative distribution function (CDF)

References

1. Probability and statistics in engineering by Hines et al (2003) Wiley.
2. Mathematical Statistics by Richard J. Rossi (2018) Wiley.
3. Probability and Statistics with reliability, queuing and computer science applications by K. S. Trivedi (1982) Prentice Hall of India Pvt. Ltd.

Random Variable

Definition (Random Variable)

Let (Ω, f, P) be probability model. A real valued function X defined on sample space Ω is a random variable if for all $x \in \mathbb{R}$, $\{\omega : X(\omega) \leq x\} \in f$, i.e. $X^{-1}((-\infty, x]) \in f$. i.e. $X^{-1}((-\infty, x])$ is an event.

Example

E: Toss a coin, then $\Omega = \{H, T\}$. Suppose X counts the number of heads. Consider $f = \{\phi, \{H\}, \{T\}, \Omega\}$. Then

$$\mathrm{X}^{-1}((-\infty, x]) = \begin{cases} \phi, & x < 0 \\ \{T\}, & 0 \leq x < 1 \\ \{H, T\}, & x \geq 1 \end{cases}$$

is in f, i.e., $X^{-1}((-\infty, x])$ an event. Hence X is a random variable.

Random Variable

Example

E: Toss a coin two times, then $\Omega = \{HH, HT, TH, TT\}$. Suppose X counts the number of heads. Consider f as power set of Ω. Then

$$\mathrm{X}^{-1}((-\infty, x]) = \begin{cases} \phi, & x < 0 \\ \{TT\}, & 0 \leq x < 1 \\ \{TT, HT, TH\}, & 1 \leq x < 2 \\ \Omega, & x \geq 2 \end{cases}$$

is an event. Hence X is a random variable.

Discrete Random Variable

- X is a discrete type random variable if we can associate a number $p_X(x) = P(X = x)$ with each outcome x in the range space R_X such that $p_X(x) \geq 0$ and $\sum_{x \in R_X} p_X(x) = 1$.
- The pair $(x, p_X(x)),\ x \in R_X$ is called probability distribution and $p_X(x)$ is called the probability mass function (PMF).

Example

Let we toss the coin three times. Then sample space is $S = \{HHH, HHT, HTH, THH, TTH, THT, HTT, TTT\}$; where H, T denote head and tail, respectively. Let X denote number of heads in tossing coin three times. Then $R_X = \{0, 1, 2, 3\}$. The probability distribution is

x	0	1	2	3
$P(X = x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Continuous Random Variable

- Let X be continuous type random variable if we can associate a function called probability density function (PDF) $f(x)$ such that

  1. $f(x) \geq 0$, for all x
  2. $\int_{-\infty}^{\infty} f(x)dx = 1$, and
  3. $P(a \leq X \leq b) = \int_{a}^{b} f(x)dx$, for $-\infty < a < b < \infty$.

Example

Let X be the time to failure (in days) of an electronic device has the following PDF

$$f(x) = \begin{cases} 0, & x < 0 \\ \lambda e^{-\lambda x}, & x \geq 0 \end{cases}$$

Here for $0 < a < b$, $P(a < X < b) = \int_a^b \lambda e^{-\lambda x} dx = e^{-\lambda a} - e^{-\lambda b}$. Then the electronic device has life more than 10 days but less than 30 days is

$$P(10 < X < 30) = e^{-10\lambda} - e^{-30\lambda}.$$

Cumulative Distribution Function (CDF)

- For random variable X, the CDF is defined as $F_X(x) = P(X \leq x)$.
- For discrete case: for random variable X, the CDF is defined as $F_X(x) = P(X \leq x) = \sum_{x_i \leq x} p_X(x_i)$, where $p(x_i)$ is the PMF. Note that $P(X = a) = F(a) - F(a-)$.
- For continuous case: for random variable X, the CDF is defined as $F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)du$, where $f(x)$ is PDF. Note that $\frac{d}{dx}F(x) = f(x)$ and $P(X = a) = 0$.

Example

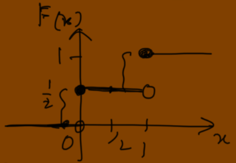
Example (1)

Let the random variable X denote the number of defects in a device having the CDF as

x	0	1	Total
$p(x)$	1/2	1/2	1

$$F(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{2}, & 0 \le x < 1 \\ 1, & x \ge 1 \end{cases}$$

$p(x) = P(X = x)$



$P(X = \frac{1}{2}) = F(\frac{1}{2}) - F(\frac{1}{2}-) = 0$

Then $P(X = 1) = F(1) - F(1-) = 1 - \frac{1}{2} = \frac{1}{2}$ and $P(X = 0) = F(0) - F(0-) = \frac{1}{2} - 0 = \frac{1}{2}$. Also $P\left(0 \le X \le \frac{3}{2}\right) = P\left(X \le \frac{3}{2}\right) - P(X < 0) = F\left(\frac{3}{2}\right) - F(0-) = 1 - 0 = 1.$

$P(0 < X \le \frac{3}{2}) + P(X=0) = F(\frac{3}{2}) - \cancel{F(0)} + \cancel{F(0)} - F(0-)$

Example (2)

A pen drive has either 2GB, 4GB, 8GB, 16GB or 64GB of memory. Let Y denote the amount of memory in a purchased pen drive with probability mass function as $p(2) = 0.05$, $p(4) = 0.10$, $p(8) = 0.35$, $p(16) = k$, $p(64) = 0.10$, where k is a constant and $p(x) = P(Y = x)$. We want to determine the value of k and the CDF of Y. Since $p(2) + p(4) + p(8) + p(4) + p(16) + p(64) = 1$, therefore $k = 0.4$. The CDF of rv Y is

$$F(x) = \begin{cases} 0, & x < 2 \\ 0.05, & 2 \leq x < 4 \\ 0.15, & 4 \leq x < 8 \\ 0.50, & 8 \leq x < 16 \\ 0.90, & 16 \leq x < 64 \\ 1, & x \geq 64 \end{cases}$$

Example (3)

Let X denote the number of trials to get a success has the following PMF

$$p(x) = q^{x-1}p, \;\; x = 1, 2, \ldots; \;\; p + q = 1.$$

Then for any positive integer x,

$$F(x) = \sum_{y \leq x} p(y) = \sum_{y=1}^{x} q^{y-1}p = p\sum_{y=0}^{x-1} q^y = p \times \frac{1-q^x}{1-q} = 1 - q^x.$$

Hence the CDF of X is

$$F(x) \;\; = \;\; \begin{cases} 0, & x < 1 \\ 1 - q^{[x]}, & x \geq 1 \end{cases}$$

where $[x]$ is the largest integer $\leq x$ (e.g., $[4.7] = 4$).

Example (4)

Let magnitude X denote the dynamic load on a bridge (in Newtons). The PDF of X is

$$f(x) = \begin{cases} x - \frac{7}{4}, & 0 \leq x \leq 4 \\ 0 & \text{otherwise} \end{cases}$$

For any number $0 \leq x \leq 4$,

$$F(x) = \int_{-\infty}^{x} f(y)dy = \int_{0}^{x} \left(y - \frac{7}{4} \right) dy = \frac{x^2}{2} - \frac{7}{4}x.$$

Then the CDF of X is

$$F(x) = \begin{cases} 0, & x < 0 \\ \frac{x^2}{2} - \frac{7}{4}x, & 0 \leq x < 4 \\ 1, & x \geq 4 \end{cases}$$

Properties of CDF

- $0 \leq F_X(x) \leq 1, \forall x$.
- $F_X(x)$ is non-decreasing in x.
- $F_X(x)$ is right continuous.
- $\lim_{x \to \infty} F_X(x) = 1$ and $\lim_{x \to -\infty} F_X(x) = 0$.

Summary

This chapter introduced the concept of random variable. Then the discrete and continuous random variables are defined. The probability mass function (PMF), the probability density function (PDF) and the cummulative distribution function (CDF) are introduced after that.