

Instructors: Purushottam Kar, Neeraj Misra
Authors: Gurpreet Singh
Email: guggu@iitk.ac.in

RANDOM VARIABLES

1. Introduction

In probability and statistics, a random variable, random quantity, aleatory variable, or stochastic variable is a variable whose possible values are numerical outcomes of a random phenomenon.

Definition 3.1 (Random Variable). A random variable defined over a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a function $X : \Omega \rightarrow \mathbb{R}$ that satisfies the following conditions

1. For every $r \in \mathbb{R}$, the set $\{X \leq r\} = A_r = \{w \in \Omega \mid X(w) \leq r\}$ satisfies $A_r \in \mathcal{F}$ i.e. the set $\{X \leq r\}$ is a measurable event for every $r \in \mathbb{R}$
2. The probabilities of events $X = \infty$ and $X = -\infty$ are zero

As a function, a random variable is required to be measurable, which rules out certain pathological cases where the quantity which the random variable returns is infinitely sensitive to small changes in the outcome.

Random variables throw away (non-useful) information.

Following is an example of a random variable, using the coin toss example.

Example 3.1 (Coin-Toss). Consider an event of a simple (fair) coin toss. We have only two outcomes, Heads (H) or Tails (T). Therefore, $\Omega = \{H, T\}$

Consider a mapping $X : \Omega \rightarrow \{-1, 1\}$ such that

$$X = \begin{cases} +1 & \text{if outcome is Heads} \\ -1 & \text{if outcome is Tails} \end{cases}$$

Since this is measurable, we can say that this is a valid random variable.

Random Variables can be both discrete as well as continuous. We will discuss continuous random variables later.

1.1 Other Examples of Common Random Variables

These are some of the commonly used (discrete) random variables

Example 3.2 (Indicator Random Variables). The Indicator Random Variables for an event E over an outcome w is defined as

$$\mathbb{I}[E] = \begin{cases} 1 & \text{if } w \in E \\ 0 & \text{if } w \notin E \end{cases}$$

Example 3.3 (Bernoulli Random Variables). The Bernoulli Random Variable defines a Bernoulli Variable assigning binary values to events e.g. toss of a coin

Example 3.4 (Rademacher Random Variables). Defines a Rademacher Variable assigning values $\{-1, 1\}$ to outcomes of binary (fair) events.

2. Cumulative Distribution Function

Before we understand Cumulative Distribution Function, it is better to understand a few other terms first.

Definition 3.2 (Support of a Random Variable). The support of a random variable (say X) is the set of points $\{x\}$ such that the probability $\mathbb{P}[x - \epsilon \leq X \leq x + \epsilon]$ is non-zero *i.e.* $S_X = \{x \mid \mathbb{P}[X = x] \neq 0\}$

Note. The support cannot be an empty set

Remark. For a random variable, if the support is a countable set, then the random variable is said to be discrete. Whereas, if $\forall x \in \mathbb{R}, \lim_{\epsilon \rightarrow 0} \mathbb{P}[X \in [x - \epsilon, x + \epsilon]] = 0$, then the random variable is said to be continuous.

Definition 3.3 (Probability Mass Function). For a discrete random variable (such as in Examples 3.1 to 3.4), the probability mass function (pmf) gives the probability that the random variable is exactly equal to some value, based on the mapping from the sample space.

Example 3.5 (Die). For a fair die, there are six possible outcomes. Therefore, we can define $\Omega = \{1, 2, 3, 4, 5, 6\}$. We can also define a random variable $X : \Omega \rightarrow \{1, 2, 3, 4, 5, 6\}$ such that

$$X(w) = \text{value}(w)$$

Since this is measurable, we can say that X is a valid random variable. Since we know the probabilities of all outcomes ($\frac{1}{6}$), we can write the pmf as

$$f_X(x) = \begin{cases} \frac{1}{6} & \text{if } x \in \{1, 2, 3, 4, 5, 6\} \\ 0 & \text{else} \end{cases}$$

Definition 3.4 (Probability Density Function). For a continuous random variable (examples later), the probability density function (pdf) defines the probability density of the random variable at a value. The integral of the pdf across an interval gives the probability that the value of the variable lies within the same interval.

Cumulative Distribution Function (CDF) defines a probability that a defined random variable X takes on a value less than or equal to a real number $x \in \mathbb{R}$ *i.e.* $F : \mathbb{R} \rightarrow [0, 1]$

$$F_X(x) = \mathbb{P}[X \leq x]$$

Some properties of a CDF are given below

1. $\lim_{x \rightarrow \infty} F_X(x) = 1$ and $\lim_{x \rightarrow -\infty} F_X(x) = 0$
2. F is not-decreasing
3. $\mathbb{P}[X > x] = 1 - F_X(x)$
4. F is right continuous *i.e.* $F_X(x^+) = F_X(x)$
5. $\mathbb{P}[x_1 < X \leq x_2] = F_X(x_2) - F_X(x_1)$
6. $\mathbb{P}[X = x] = F_X(x) - F_X(x^-)$

Note. For some continuous random variables (say X), we can find a function f_X such that

$$F_X(x) = \int_{-\infty}^x f_X(t) dt$$

However, it is not necessary that a pdf always exists. For cases in which it exists, the random variable is said to be absolutely continuous

Note. The CDF F_X might be absolutely continuous, even if the $<++>$

Example 3.6 (Coin-Toss). Same as in Example 1.1. we consider the random variable X , where

$$X = \begin{cases} +1 & \text{if outcome is Heads} \\ -1 & \text{if outcome is Tails} \end{cases}$$

Since we are considering a fair coin, we can write the probability of getting Heads as $\frac{1}{2}$, and the same for getting tails. Hence, we can write the Probability Mass Function as follows

$$f_X(x) = \begin{cases} 0.5 & \text{if } x \in \{-1, 1\} \\ 0 & \text{else} \end{cases}$$

Hence, we can write the cumulative distribution function

$$F_X(x) = \begin{cases} 0 & x < -1 \\ 0.5 & x \geq -1 \text{ and } x < 1 \\ 1 & x \geq 1 \end{cases}$$