Probability and Statistics

MA-202

Summary

Reference

- Rohatgi, V. K., & Saleh, A. M. E. (2015). An Introduction to Probability and Statistics. John Wiley & Sons.

Random Variable

Recall that for an experiment with sample space Ω, a random variable X (usually denoted by capital letters) is a real-valued function $X : \Omega \to \mathbb{R}$ such that pre-image of every interval in $\mathbb{R}$ is an event of Ω.

For example,

- If in the study of the ecology of a lake, X, the r.v. may be depth measurements at randomly chosen locations. Then X is a random variable. The range for X is the minimum depth possible to the maximum depth possible

- Tossing a coin three times,

$$\Omega = \{HHH, HHT, HTH, HTT, TTT, TTH, THT, THH\}$$

  Define: $X : \Omega \to \mathbb{R}$ as number of heads. Then, X is a random variable with support (possible values or range) $\{0, 1, 2, 3\}$

Given an experiment and the corresponding set of possible outcomes (the sample space), a random variable associates a particular number with each outcomes. Mathematically a random variable denoted by X is a real valued function from Ω to $\mathbb{R}$, i.e, $X : \Omega \to \mathbb{R}$.[1]

Distribution Function

A real-valued function F defined on $(-\infty, \infty)$ that is

a) **Non-Decreasing:** If $x < y$, then $F(x) \leq F(y)$.

b) **Right Continuous:** $F(x^+) := \lim_{h \to 0} F(x+h) = F(x)$ for all $x \in \mathbb{R}$.

c) $\lim_{x \to -\infty} F(x) = 0$.

d) $\lim_{x \to \infty} F(x) = 1$.

is called a distribution function.

- F is non-decreasing and can have only jump discontinuities.
- $F(x)$ can be
  - a step function
  - an absolutely continuous function
  - a mix of above

Cumulative Distribution Function (CDF)

Let X be a random variable defined over $(\Omega, \mathcal{F}, P)$. Define a point function $F_X(\cdot)$ on $(-\infty, \infty)$ as follows:

$$F_X(x) = P(\omega \in \Omega : X(\omega) \leq x) = P(X \leq x), \quad \forall\, x \in \mathbb{R}.$$

The function $F_X(\cdot)$ is called the distribution function of RV X. One can prove that $F_X(\cdot)$ is indeed a distribution function.

Based on $F_X(\cdot)$ (i.e., pure jumps, absolutely continuous or mix of these), we can classify X as discrete, continuous or mixed type random variable.[2]

Discrete Random Variable

Definition 1. *An RV X defined on $(\Omega, \mathcal{F}, P)$ is said to be of the discrete type, or simply discrete, if there exists a countable set $E \subseteq \mathcal{S}$ such that $P\{X \in E\} = 1$. Let X take on the value x_i, with*

[1] For a more formal and precise definition, refer the book mentioned above.

[2] One can alternatively define random variables directly, without referring to $F_X(x)$.

probability p_i $(i = 1, 2, \ldots)$. *We have*

$$P\{\omega : X(\omega) = x_i\} = p_i,$$

Then, $\sum_{i=1}^{\infty} p_i = 1$ *and* $p_i \geq 0$ *for all* i.

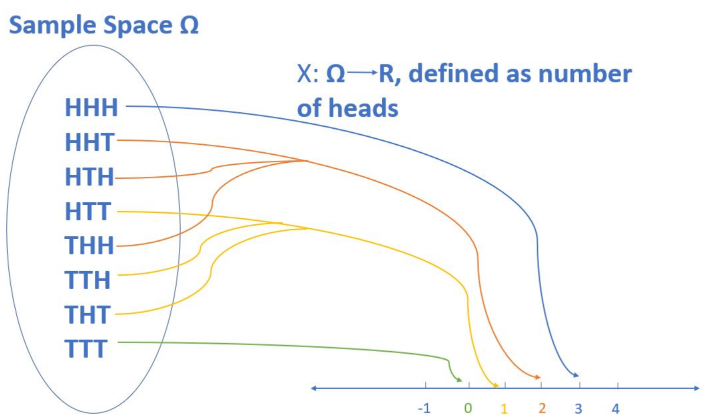
Remark 1. *The points of* E *that have positive mass are called jump points or points of increase of the DF of* X, *and their probabilities are size of jumps of the DF.*

Definition 2. *The collection of numbers* $\{p_i\}$ *satisfying* $P\{X = x_i\} = p_i \geq 0$, *for all* i *and* $\sum_{i=1}^{\infty} p_i = 1$ *is called the probability mass function (PMF) of RV* X.

The CDF $F_X(\cdot)$ of X is given by:

$$F_X(x) = P(X \leq x) = \sum_{\{i; x_i \leq x\}} p_i.$$

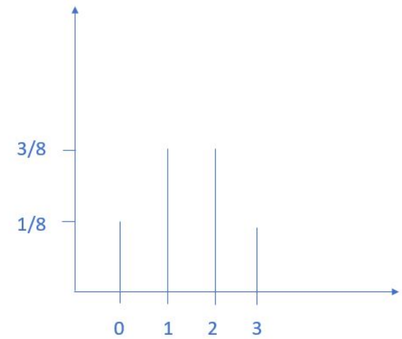
Example 1. *Consider again the experiment of tossing a coin three times. Define:* $X : \Omega \to \mathbb{R}$ *as number of heads. That is,*



$$\begin{aligned}
P(X = 1) &= P(HTT, THT, TTH) = 3/8 \\
P(X = 2) &= P(HHT, HTH, THH) = 3/8 \\
P(X = 0) &= P(TTT) = 1/8 \\
P(X = 3) &= P(HHH) = 1/8.
\end{aligned}$$

Example 2. *Suppose a game is to be played by throwing a fair die. The rules are as follows:*

x	0	1	2	3
P(X=x)	1/8	3/8	3/8	1/8



- *you win 1 Rs. if face 2 turns up.*
- *you win 2 Rs. if face 4 turns up.*
- *you lose 1.5 Rs. if face 6 turns up.*
- *you neither win nor lose if any other face turns up.*

Define X as a random variable giving the amount won on any throw of die. What is the support of X? Find the corresponding PMF?

Solution: *Clearly, X can take values $\{-1.5, 0, 1, 2\}$ which implies that X is a discrete random variable.*

$$\begin{aligned}
P(X = 1) &= P(\textit{face 2 turns up}) = 1/6 \\
P(X = 2) &= P(\textit{face 4 turns up}) = 1/6 \\
P(X = -1.5) &= P(\textit{face 6 turns up}) = 1/6 \\
P(X = 0) &= P(\textit{either of faces 1,3,5 turn up}) = 3/6.
\end{aligned}$$

Continuous Random Variable

Let X be an RV defined on $(\Omega, \mathcal{F}, P)$ with DF F. Then X is said to be of the continuous type (or simply, continuous) if F is absolutely continuous, that is, if there exists a nonnegative function $f(x)$ such that for every real number x we have

$$F(x) = \int_{-\infty}^{x} f(t)dt.$$

The function f is called the probability density function (PDF) of the RV X.

Note that $f \geq 0$ and satisfies $\lim_{x\to+\infty} F(x) = F(+\infty) = \int_{-\infty}^{\infty} f(t)dt = 1$.

Let a and b be any two real numbers with $a < b$. Then

$$P\{a < X < b\} = \int_{a}^{b} f(t)dt.$$

Note that $F'(x) = f(x)$.

Remark 2. *In the discrete case, $P(X = a)$ is the probability that X takes the value a. In the continuous case, $f(a)$ is not the probability that X takes the value a. Indeed, if X is of the continuous type, it assumes every value with probability 0.*

Corollary 1. *Let X be a random variable defined over $(\Omega, \mathcal{F}, P)$, $F_X(\cdot)$ its distribution function and $a, b \in \mathbb{R}$ with $a < b$. We define the following notation,*

$$F_X(x-) := \lim_{h\to 0} F_X(x-h) = P(X < x).$$

Further, we have

a) $P(X = a) = P(X \leq a) - P(X < a) = F_X(a) - F_X(a^-)$, *i.e., size of the jump in F_X at a. Note that for a continuous random variable, $P(X = a) = 0$, $\forall\ a$ (since F_X is continuous for a continuous random variable).*

b) $P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F_X(b) - F_X(a)$.

c) $P(a < X < b) = P(X < b) - P(X \leq a) = F_X(b^-) - F_X(a) = F_X(b) - F_X(a) - P(X = b)$.

d) $P(a \leq X < b) = P(X < b) - P(X < a) = F_X(b^-) - F_X(a^-) = F_X(b) - F_X(a) - P(X = b) + P(X = a)$.

e) $P(a \leq X \leq b) = P(X \leq b) - P(X < a) = F_X(b) - F_X(a^-) = F_X(b) + - F_X(a) + P(X = a)$.