# Chapter 3

# Probability

## 3.1 Preliminaries

### 3.1.1 Random Experiments and Sample Space

The main starting idea in Probability theory is the notion of a **random experiment**. By a random experiment we refer to an experiment where we do not know exactly what the outcome is, but know that the outcome will be from a known set of possible outcomes. This known set of possible outcomes is called the **Sample Space** for the experiment, and this set will be denoted by $\Omega$.

**Example 3.1.1** The simplest example is that of tossing a coin. While we do not know what the outcome is, we know that it has to be either a Head or a Tail, which we denote by $H$ and $T$ respectively. Thus in this experiment we have

$$\Omega = \{H, T\} \tag{3.1.1}$$

**Example 3.1.2** Suppose we toss a fair coin thrice and note the sequence of outcomes. Then for this experiment we have

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\} \tag{3.1.2}$$

**Example 3.1.3** For the experiment of throwing a six faced die, with its faces numbered $1, 2, 3, 4, 5, 6$ we have

$$\Omega = \{1, 2, 3, 4, 5, 6\} \tag{3.1.3}$$

**Example 3.1.4** Suppose we throw a coin on the floor and note the coordinates of the centre of the coin when it lands. Let us use a reference $X$ and $Y$ axis and denote the points in the room by $(x, y)$; where

$$a \leq x \leq b \text{ and } c \leq x \leq d$$

Then

$$\Omega \;=\; \{(x, y)\ :\ a \leq x \leq b,\ c \leq y \leq d\} \tag{3.1.4}$$

Thus the first ingredient in Probability Theory is the Sample Space of a random experiment.

### 3.1.2 Events

The next important ingredient in Probability theory is the notion of **Events**. Suppose in the random experiment of rolling a die we may be interested in the possibility of an even number turning up. In this case we are interested in the outcome to be in the subset $\{2, 4, 6\}$ of the sample sapce. In general we may be particularly interested in the outcome being in some special subsets of the sample space. We shall call these as **Elementary Events**. We would like the collection of events we deal with to be set theoretically self contained. What we mean by this is that we would like the collection of events to be such that when we perform the standard set theoretic operations of these events the result is also in this collection of events. We shall now make this idea more specific.
Let $\mathcal{B}$ be the collection of subsets, of the sample sapce $\Omega$, that we are interested in. We would like to have the following properties of $\mathcal{B}$:

1. $\mathcal{B}$ must be a **nonempty collection**. (We have at least some events which are of interest)

2. $\mathcal{B}$ is **closed under complementation**, that is,

$$A \in \mathcal{B} \implies A' \in \mathcal{B} \tag{3.1.5}$$

3. $\mathcal{B}$ is **closed under union**, that is,

$$A, B \in \mathcal{B} \implies A \cup B \in \mathcal{B} \tag{3.1.6}$$

From the above it follows that $\mathcal{B}$ is **closed under finite union**, that is,

$$A_1, A_2, \cdots, A_N \in \mathcal{B} \implies \bigcup_{j=1}^{N} A_j \in \mathcal{B} \tag{3.1.7}$$

4. $\mathcal{B}$ is **closed under intersection**, that is,

$$A, B \in \mathcal{B} \implies A \cap B \in \mathcal{B} \tag{3.1.8}$$

From the above it follows that $\mathcal{B}$ is **closed under finite intersection**, that is,

$$A_1, A_2, \cdots, A_N \in \mathcal{B} \implies \bigcap_{j=1}^{N} A_j \in \mathcal{B} \tag{3.1.9}$$

5. $\mathcal{B}$ is closed under monotonic nondecreasing limits. What we mean by this is the following:
Suppose $\{A_n\}_{n=1,2,\cdots}$ is a nondecreasing sequence of sets in $\mathcal{B}$, that is, $A_n \subseteq A_{n+1}$, for $n = 1, 2, \cdots$. Then we have

$$\lim_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n \tag{3.1.10}$$

We want $\mathcal{B}$ is closed with respect to this limit means that we want

$$\lim_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n \in \mathcal{B} \tag{3.1.11}$$

for monotone non decreasing sequence $\{A_n\}_{n=1,2,\cdots}$ of sets in $\mathcal{B}$.

6. Similarly we want $\mathcal{B}$ is closed under monotonic nonincreasing limits. If $\{A_n\}_{n=1,2,\cdots}$ is a nonincreasing sequence of sets in $\mathcal{B}$, that is, $A_{n+1} \subseteq A_n$, for $n = 1, 2, \cdots$ then we have

$$\lim_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} A_n \tag{3.1.12}$$

We want $\mathcal{B}$ is closed with respect to this limit means that we want

$$\lim_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} A_n \in \mathcal{B} \tag{3.1.13}$$

for monotone non increasing sequence $\{A_n\}_{n=1,2,\cdots}$ of sets in $\mathcal{B}$.

Using DeMorgan's laws we can easily see that if we have Properties 1,2,3 and 5 above then Properties 4 and 6 follows automatically. Hence basically we require 1,2,3 and 5 to be satisfied by $\mathcal{B}$. These ideas lead us to the notion of a $\sigma$-algera.

### 3.1.3 Sigma Algebras

Suppose we have a collection $\mathcal{B}$ of subsets of $\Omega$ which satisfy Properties 1,2,3 and 5 above, that is

$$\mathcal{B} \text{ is a nonempty collection} \tag{3.1.14}$$

$$A \in \mathcal{B} \Longrightarrow A' \in \mathcal{B} \tag{3.1.15}$$

$$A, B \in \mathcal{B} \Longrightarrow A \cup B \in \mathcal{B} \tag{3.1.16}$$

$$\{A_n\}_{n=1,2,\cdots} \in \mathcal{B}, \text{ and } A_n \subseteq A_{n+1} \text{ for } n = 1, 2, \cdots \Longrightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{B} \tag{3.1.17}$$

Consider any sequence $\{B_n\}_{n=1,2,\cdots} \in \mathcal{B}$. Now define

$$A_1 = B_1 \tag{3.1.18}$$

$$A_n = \bigcup_{j=1}^{n} B_j \ , \ n = 2, 3, \cdots \tag{3.1.19}$$

Then clearly we have

$$A_n = \bigcup_{j=1}^{n} B_j \in \mathcal{B} \text{ by (3.1.15)} \tag{3.1.20}$$

$$\bigcup_{j=1}^{n} B_j = \bigcup_{j=1}^{n} A_j \text{ for every } n \tag{3.1.21}$$

$$\bigcup_{j=1}^{\infty} B_j = \bigcup_{j=1}^{\infty} A_j \tag{3.1.22}$$

Clearly $\{A_n\}_{n=1,2,\cdots}$ is a non decreasing sequence in $\mathcal{B}$. Hence by (3.1.17) we get

$$\bigcup_{n=1}^{\infty} A_n \in \mathcal{B} \tag{3.1.23}$$

Hence by (3.1.22) we get

$$\bigcup_{n=1}^{\infty} B_n \in \mathcal{B} \tag{3.1.24}$$

Thus we have

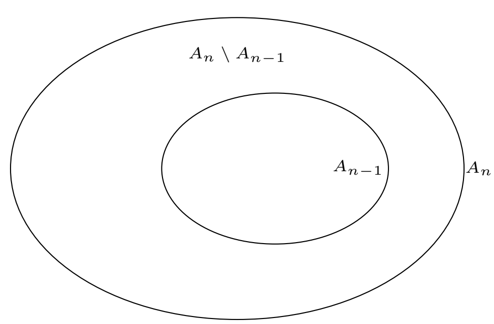
$$B_n \in \mathcal{B} \implies \bigcup_{n=1}^{\infty} B_n \in \mathcal{B} \tag{3.1.25}$$

Whenever (3.1.25) is true we say that $\mathcal{B}$ is closed under countable union. Thus we have

$$(3.1.14), (3.1.15), (3.1.16) \text{ and } (3.1.17) \implies (3.1.25), \text{ that is },$$
$$\mathcal{B} \text{ is closed under countable union} \tag{3.1.26}$$

Conversely suppose we have (3.1.14),(3.1.15), and (5.3.26). Then clearly (3.1.16) is satisfied. We shall now see that (3.1.10) is also satisfied. We see this as follows:
Let $\{A_n\}_{n=1,2,\cdots}$ be a monotone non decreasing sequece of sets in $\mathcal{B}$. Define

$$B_1 = A_1 \tag{3.1.27}$$
$$B_n = A_n \setminus A_{n-1} \text{ for } n = 2, 3 \cdots \tag{3.1.28}$$



Then we have

1. $B_n \in \mathcal{B}$ for $n = 1, 2, \cdots$

2. $\bigcup_{j=1}^{n} B_j = \bigcup_{j=1}^{n} A_j$ for $n = 1, 2, 3, \cdots$

3. $\bigcup_{j=1}^{\infty} B_j = \bigcup_{j=1}^{\infty} A_j$

Using (3.1.26) we see that $\bigcup_{j=1}^{\infty} B_j \in \mathcal{B}$ and hence $\bigcup_{j=1}^{\infty} A_j \in \mathcal{B}$. This means that $\mathcal{B}$ is closed under monotone non decreasing limits. Thus we get that

$$\begin{aligned}(3.1.14), (3.1.15), (3.1.16), \text{ and } (3.1.18) \text{ hold} \quad &\Longleftrightarrow \\ (3.1.14), (3.1.15) \text{ and } (3.1.26) \text{ hold}& \end{aligned} \tag{3.1.29}$$

This leads us to the following definition:

**Definition 3.1.1** A collection $\mathcal{B}$, of subsets of a set $\Omega$, is said to be a **$\boldsymbol{\sigma}$-algebra** of subsets of $\Omega$ if

1. $\mathcal{B}$ is a non empty collection,

2. $\mathcal{B}$ is closed under complementation, and

3. $\mathcal{B}$ is closed under countable union

From the above definition, and simple set theoretic properties we see that any $\sigma$-algebra $\Sigma$ of subsets of $\Omega$ has the following additional properties:

1. By (3.1.16) we have $\Sigma$ is closed under finite union and closed under monotone non decreasing limits

2. By DeMorgan's laws we have $\Sigma$ is closed under

   (a) finite intersection,

   (b) countable intersection and

   (c) monotone non increasing limits

3. We must have $\Omega \in \Sigma$ and $\phi \in \Sigma$. This follows from the fact that being a non empty collection there must be a set $A \in \Sigma$. Now by closure under complementation we must have $A' \in \Sigma$ and hence by closure under union we have $\Omega = A \cup A' \in \Sigma$. Now by closure under complementation we must have $\phi = \Omega' \in \Sigma$

**Remark 3.1.1** The smallest $\sigma$-algebra is the collection containing only the two sets $\Omega$ and $\phi$, and the largest $\sigma$-algebra is the collection of all subsets of $\Omega$, which is called the **Power Set** of $\Omega$ and denoted by either $\mathcal{P}(\Omega)$ or $2^{\Omega}$.

### 3.1.4 Smallest $\sigma$-algebra Containing a Collection of Subsets

Consider a collection $\mathcal{S}$ of subsets of $\Omega$. This collection $\mathcal{S}$ may or may not be a $\sigma$-algebra. For example, if

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

then let

$$\mathcal{S} = \{A_{12}, A_{34}, A_{56}\} \tag{3.1.30}$$

where

$$A_{12} = \{1, 2\} \tag{3.1.31}$$
$$A_{34} = \{3, 4\} \tag{3.1.32}$$
$$A_{56} = \{5, 6\} \tag{3.1.33}$$

Clearly for many reasons this is not a $\sigma$-algebra. For instance $\Omega \notin \mathcal{S}$, or $\phi \notin \mathcal{S}$. Also it is not closed under union or intersection. Thus given a collection $\mathcal{S}$, of subsets of $\Omega$, it may or not be a $\sigma$-algebra of subsets of $\Omega$. We want to imbed this in a $\sigma$-algebra of subsets of $\Omega$, that is, we want to have a $\sigma$-algebra $\Sigma$ such that $\mathcal{S} \subseteq \Sigma$. Can we do this? Of course we can do this, since we can take $\Sigma$ to be $2^{\Omega}$, the power set of $\Omega$. Then clearly $\mathcal{S} \subseteq 2^{\Omega}$. Consider the above example, and let

$$\Sigma = \{\Omega, \phi, A_{12}, A_{34}, A_{56}, A_{1234}, A_{3456}, A_{1256}\} \tag{3.1.34}$$

where

$$A_{1234} = \{1,2,3,4\} \tag{3.1.35}$$
$$A_{3456} = \{3,4,5,6\} \tag{3.1.36}$$
$$A_{1256} = \{1,2,5,6\} \tag{3.1.37}$$

Then $\Sigma$ is a $\sigma$-algebra, it contains $\mathcal{S}$ and it is smaller than $2^\Omega$ which is also a $\sigma$-algebra that contains $\mathcal{S}$. Thus we may be able to imbed $\mathcal{S}$ in many $\sigma$-algebras. What we want to do is to do this optimally. This means we want the smallest $\sigma$-algebra that contains $\mathcal{S}$. This means that we are looking for a $\sigma$-algebra $\Sigma$ such that

1. $\mathcal{S} \subseteq \Sigma$ and

2. If $\Sigma_1$ is any $\sigma$-algebra that contains $\mathcal{S}$, that is $\mathcal{A} \subseteq \Sigma_1$, then $\Sigma \subseteq \Sigma_1$

Can we find such an optimal $\sigma$-algebra? It can be shown that this is possible and this smallest $\sigma$-algebra containig $\mathcal{S}$ is called the **$\boldsymbol{\sigma}$-algebra generated by $\boldsymbol{\mathcal{S}}$**, and is denoted by $\Sigma(\mathcal{S})$. In the above example the $\Sigma$ given in (3.1.35) is the smallest $\sigma$-algebra generated by the $\mathcal{S}$ in (3.1.31).

**Remark 3.1.2** If $\Omega = \mathbb{R}$, the set of all real numbers and if we consider $\mathcal{S}$ to be the set $\mathcal{I}$ of all intervals then the smallest $\sigma$-algebra generated by this collection of all intervals is called the **Borel $\boldsymbol{\sigma}$-algebra in** $\mathbb{R}$ and is denoted by $\mathcal{B}$. Any set in $\mathcal{B}$ is called a **Borel Set**. $\mathcal{B}$ is also the $\sigma$-algebra generated by the any of the following collection of intervals:

1. $\mathcal{S} =$ collection of all closed intervals

2. $\mathcal{S} =$ collection of all open intervals

3. $\mathcal{S} =$ collection of all left open right closed intervals

4. $\mathcal{S} =$ collection of all right open left closed intervals

5. $\mathcal{S} =$ collection $\mathcal{S}$ of all intervals of the form $(-\infty, x],\ x \in \mathbb{R}$

To conclude, we want the collection of events to be a $\sigma$-algebra of subsets of $\Omega$. Thus we have now two main ingredients for Probability theory, namely the Sample Space and a $\sigma$-algebra of events.

### 3.1.5 Probability Measure

We next look at the next important ingredient in Probability Theory, namely, the notion of a **Probability Measure**. Consider a sample space $\Omega$ and $\sigma$-algenra of events $\mathcal{B}$. What we want to do is to assign some weights to each event $E \in \mathcal{B}$, (which captures the proportion of times the elements of $E$ occur in the total sample space). Thus we want a map

$$\mathcal{P} : \mathcal{B} \longrightarrow \mathbb{R}$$

We shall now look at some natural properties that we shall expect from such a weighting system.

1. Clearly we would like to have the weights to be nonnegative, that is $\mathcal{P}(E) \geq 0$ for every $E \in \mathcal{B}$. Further we shall normalize the total weight to be one, that is $\mathcal{P}(\Omega) = 1$, and we shall, naturaly, assign the weight 0 to the empty set. Thus we need the following properties of $\mathcal{P}$:

$$0 \leq \mathcal{P}(E) \leq 1 \text{ for every } E \in \mathcal{B} \tag{3.1.38}$$

$$\mathcal{P}(\Omega) = 1 \tag{3.1.39}$$

$$\mathcal{P}(\phi) = 0 \tag{3.1.40}$$

2. We also like this weighting system to follow the principle that "the whole is equal to the sum of its parts". What we mean by this is the following:

$$\{E_j\}_{j=1}^{N} \in \mathcal{B} \text{ and } A_i \text{ are mutually pairwise disjoint}$$

$$\Longrightarrow$$

$$\mathcal{P}\left(\bigcup_{j=1}^{N} E_j\right) = \sum_{j=1}^{N} \mathcal{P}(E_j) \tag{3.1.41}$$

From the above properties we can easily see that the following must also hold:

$$\begin{array}{rcl} A \in \mathcal{B} & \Longrightarrow & A' \in \mathcal{B} \\ \Longrightarrow & & \\ \mathcal{P}(\Omega) & = & \mathcal{P}(A \bigcup A') \\ \Longrightarrow & & \\ 1 & = & \mathcal{P}(A) + \mathcal{P}(A') \\ \Longrightarrow & & \\ \mathcal{P}(A') & = & 1 - \mathcal{P}(A) \end{array}$$

We can also see that,

$$\begin{aligned}
A, B \in \mathcal{B} \text{ and } B \subseteq A &\implies A = B \bigcup (A \setminus B) \\
&\implies \\
\mathcal{P}(A) &= \mathcal{P}(B) + \mathcal{P}(A \setminus B) \\
&\implies \\
\mathcal{P}(A \setminus B) &= \mathcal{P}(A) - \mathcal{P}(B)
\end{aligned}$$

Also from the fact that

$$\mathcal{P}(A) = \mathcal{P}(B) + \mathcal{P}(A \setminus B)$$

we see that

$$B \subseteq A \implies \mathcal{P}(B) \leq \mathcal{P}(B) \tag{3.1.42}$$

Thus $\mathcal{P}$ is a monotone function.

3. The next property that we shall demand is a natural continuity of the weighting function. Suppose $\{E_j\}_{j=1}^{\infty}$ is as sequence of sets such that

   (a) $E_j \in \mathcal{B}$ for all $j = 1, 2, \cdots$, and

   (b) The $E_j$ are nondecreasing sets, that is,

   $$E_j \subseteq E_{j+1} \text{ for } j = 1, 2, \cdots$$

   then we have

   $$\lim_{n\to\infty} E_n = \bigcup_{j=1}^{\infty} E_j$$

   We would like the weights $\mathcal{P}(E_n)$ of $E_n$ to converge to the weight $\mathcal{P}(\lim_{n\to\infty} E_n)$. Thus we ask for the following property of $\mathcal{P}$:

   $$\begin{aligned}
   &E_j \text{ is a nondecreasing sequence of sets in } \mathcal{B} \text{ and} \\
   &E_j \subseteq E_{j+1} \text{ for } j = 1, 2, \cdots \implies \\
   &\mathcal{P}(\bigcup_{j=1}^{\infty} E_j) = \lim_{n\to\infty} \mathcal{P}(E_n)
   \end{aligned} \tag{3.1.43}$$

   We call this property "continuity from below" of $\mathcal{P}$

Thus the natural properties of $\mathcal{P}$ that we are looking for are (3.1.38) to (3.1.42). We shall now look at these in a slightly different way. Let us now consider a disjoint sequence of sets $\{A_n\}_{n=1}^{\infty}$ in $\mathcal{B}$. We then define a new sequence of sets, $\{B_n\}_{n=1}^{\infty}$, as follows:

$$B_n = \bigcup_{j=1}^{n} A_j \text{ for } n = 1, 2, \cdots \tag{3.1.44}$$

It is easy to see that

$$\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n \tag{3.1.45}$$

Since $\mathcal{B}$ is a $\sigma$-algebra and $A_n \in \mathcal{B}$, it follows that $B_n \in \mathcal{B}$ for all $n$. $B_n$ is a nondecreasing sequence. Hence by the "'continuity from below " property of $\mathcal{P}$ we have

$$\begin{aligned}
\mathcal{P}(\bigcup_{n=1}^{\infty} B_n) &= \mathcal{P}(\lim_{n\to\infty} B_n) \\
&= \lim_{n\to\infty} \mathcal{P}(B_n) \\
&= \lim_{n\to\infty} \mathcal{P}(\bigcup_{j=1}^{n} A_j) \\
&= \lim_{n\to\infty} \sum_{j=1}^{n} \mathcal{P}(A_j) \text{ by } (3.1.41) \\
&= \sum_{n=1}^{\infty} \mathcal{P}(A_n)
\end{aligned}$$

Using (3.1.43) we get

$$\mathcal{P}(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathcal{P}(A_n)$$

Thus we see that the principle, "the whole is the sum of its parts", holds even when there is an infinite sequence of disjoint parts. We call this property "countable additivity " of $\mathcal{P}$.

**Definition 3.1.2** A map

$$\mathcal{P} : \mathcal{B} \longrightarrow \mathbb{R}$$

is said to be countably additive if

$$\{A_n\}_{n=1}^{\infty} \in \mathcal{B} \text{ and } A_n \cap A_m = \phi \text{ if } n \neq m$$
$$\Longrightarrow$$
$$\mathcal{P}(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathcal{P}(A_n) \tag{3.1.46}$$

Thus we have

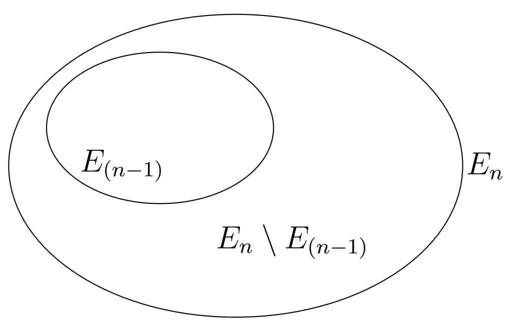
$$\text{Properties (3.1.38) to (3.1.42)} \implies \text{countable additivity of } \mathcal{P} \tag{3.1.47}$$

On the other hand we shall now see that the properties (3.1.38),(3.1.39), (3.1.40) and countable additivity, namely (3.1.45) imply that (3.1.41) and (3.1.42) also hold. For, clearly countable additivity implies finis aditivity and hence (3.1.40) is valid. We have to only show that continuity from below holds. For this, let $\{E_n\}_{n=1}^{\infty}$ be a nondecreasing sequence in $\mathcal{B}$. We define a new sequence $\{A_n\}_{n=1}^{\infty}$, as follows:

$$\begin{aligned} A_1 &= E_1 \\ A_n &= E_n - E_{(n-1)} \text{ for } n \geq 2 \end{aligned}$$

Clearly we have

$$\bigcup_{n=1}^{\infty} E_n = \bigcup_{n=1}^{\infty} A_n$$




The sequence $A_n$ thus defined is in $\mathcal{B}$ and the sets are all pairwise disjoint. Hence countable additivity gives

$$\mathcal{P}(\lim_{n\to\infty} E_n) = \mathcal{P}(\bigcup_{n=1}^{\infty} E_n)$$

$$
\begin{aligned}
&= \mathcal{P}(\bigcup_{n=1}^{\infty} A_n) \\
&= \sum_{n=1}^{\infty} \mathcal{P}(A_n) \\
&= \lim_{n\to\infty} \sum_{j=1}^{n} \mathcal{P}(A_j) \\
&= \lim_{n\to\infty} \left\{ \mathcal{P}(E_1) + \sum_{j=2}^{n} \mathcal{P}(E_j \setminus E_{(j-1)}) \right\} \\
&= \lim_{n\to\infty} \{\mathcal{P}(E_1) + (\mathcal{P}(E_2) - \mathcal{P}(E_1)) + (\mathcal{P}(E_3) - \mathcal{P}(E_2)) + \cdots + \\
&\quad (\mathcal{P}(E_n) - \mathcal{P}(E_{(n-1)}))\} \\
&= \lim_{n\to\infty} \mathcal{P}(E_n)
\end{aligned}
$$

thus proving that $\mathcal{P}$ is continuous from below. Thus we have

$$
\left.\begin{array}{l}
\text{Properties (3.1.38),(3.1.39), (3.1.40) and (3.1.45)} \Longrightarrow \\
\mathcal{P} \text{ is continuous from below}
\end{array}\right\} \qquad (3.1.48)
$$

Thus from (3.1.46) and (3.1.47) we see that

$$
\left.\begin{array}{l}
(3.1.38), (3.1.39), (3.1.40), (3.1..41) \text{ and } (3.1.42) \\
\text{are equivalent to} \\
\text{(3.1.38),(3.1.39),(3.1.40) and (3.1.45)}
\end{array}\right\} \qquad (3.1.49)
$$

Hence we take (3.1.38),(3.1.39),(3.1.40) and (3.1.45) as the defining properties of a Probability measure. We have

**Definition 3.1.3** Let $\Omega$ be a sample space and $\mathcal{B}$ a $\sigma$-algebra of subsets of $\Omega$. A map

$$\mathcal{P} : \mathcal{B} \longrightarrow \mathbb{R}$$

is said to be a probability measure if it satisfies the following properties:

$$0 \leq \mathcal{P}(E) \leq 1 \text{ for all } E \in \mathcal{B} \qquad (3.1.50)$$

$$\mathcal{P}(\Omega) = 1 \qquad (3.1.51)$$

$$\mathcal{P}(\phi) = 0 \qquad (3.1.52)$$

$$
\left.\begin{array}{l}
A_n \in \mathcal{B} \text{ and } A_n \bigcap A_m = \phi \text{ if } n \neq m \\
\Longrightarrow \\
\mathcal{P}(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathcal{P}(A_n)
\end{array}\right\} \qquad (3.1.53)
$$

Every Probability measure satisfies finite additivity and continuity from below. We can also show that a probability measure is continuous from above, that is,

$$\left.\begin{array}{l} A_n \in \mathcal{B} \text{ and } A_{(n+1)} \subseteq A_n \text{ for } n = 1, 2, \cdots \\ \Longrightarrow \\ \mathcal{P}(\bigcap_{n=1}^{\infty} A_n) = \lim_{n\to\infty} \mathcal{P}(A_n) \end{array}\right\} \qquad (3.1.54)$$

# 3.2 Random Variables

## 3.2.1 What is a random variable

Consider the random experiment of rolling a fair die. We have

$$\begin{aligned} \Omega &= \{1, 2, 3, 4, 5, 6\} \\ \mathcal{B} &= \text{The set of all subsets of } \Omega \\ P(j) &= \frac{1}{6} \text{ for } 1 \leq j \leq 6 \end{aligned}$$

Suppose the person rolling the die gets a payment of $j$ rupees if an even number $j$ shows up and has to pay a penalty of $j$ rupees if an odd number $j$ shows up. Then we can express the pay off of this payoff scheme as a function

$$X : \Omega \longrightarrow \mathbb{R}$$

where the values of $X$ are given below:

| $j$ | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $X$ | $-1$ | 2 | $-3$ | 4 | $-5$ | 6 |

We now want to consider such functions on $\Omega$. Let $(\Omega, \mathcal{B}, P)$ be a Probability space. Let

$$X : \Omega \longrightarrow \mathbb{R}$$

be a function defined on the sample space $\Omega$. Then for each $\omega \in \Omega$ we have $X(\omega)$, a real number. We are interested in looking at the values of $\omega$ for which the function $X(\omega)$ lies in a set $B$ in $\mathbb{R}$. We define

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \qquad (3.2.1)$$

Note that $X^{-1}(B)$ is a subset of $\Omega$ for every subset $B$ in $\mathbb{R}$. However this subset $X^{-1}(B)$ need not be an event, that is, $X^{-}(B)$ need not be in $\mathcal{B}$. We would like the function $X$ to be such that $X^{-}(B)$ is an event for "nice" sets $B$ in $\mathbb{R}$. The nice sets we are interested are the Borel subsets of $\mathbb{R}$. When this happens we say that $X$ is a random variable on $(\Omega, \mathcal{B}, P)$. We therefore give the following definition:

**Definition 3.2.1 A function $X : \Omega \longrightarrow \mathbb{R}$ is said to be a random variable on a probability space $(\Omega, \mathcal{B}, P)$ if**

$$\boldsymbol{X^{-1}(B) \in \mathcal{B}} \textbf{ for every Borel set } \boldsymbol{B} \textbf{ in } \mathbb{R}$$

**Remark 3.2.1** In particular we see that if $X$ is a random variable then $X^{-1}(I) \in \mathcal{B}$ for every interval $I$ in $\mathbb{R}$, since every interval is a Borel set.

### 3.2.2 Probability Measure Induced on $\mathbb{R}$ by a Random Variable

Consider a random variable $X$ on a probability space $(\Omega, \mathcal{B}, P)$. Now we look at $\mathbb{R}$ as a sample space of a random experiment with the Borel sets as the events, that is now we look at $\Omega_1 = \mathbb{R}$ and $\mathcal{B}_1 = \mathcal{B}$. For any Borel set $B \in \mathcal{B}$ we have $X^{-1}(B) \in \mathcal{B}$ and hence we can define $P(X^{-1}(B)$. We denote this by $\mathcal{P}_X(B)$. Thus we have a function

$$\mathcal{P}_X : \mathcal{B} \longrightarrow \mathbb{R} \tag{3.2.2}$$

We observe the following properties of the function $\mathcal{P}_X$:

1. We have

$$\mathcal{P}_X(B) \quad = \quad P(X^{-1}(B))$$

   Since the right hand side is a probability we get

$$0 \leq \mathcal{P}_X(B) \leq 1 \text{ for every } B \in \mathcal{B} \tag{3.2.3}$$

2. Since $X^{-1}(\mathbb{R}) = \Omega$ and $X^{-1}(\phi) = \phi$ we get

$$\begin{aligned} \mathcal{P}_X(\mathbb{R}) = P(\Omega) &= 1 && (3.2.4) \\ \mathcal{P}_X(\phi) = P(\phi) &= 0 && (3.2.5) \end{aligned}$$

3. If $B_1, B_2, \cdots, B_N$ is a finite collection of mutually disjoint Borel sets in $\mathbb{R}$ then we have

$$X^{-1}\left(\bigcup_{j=1}^{N} B_j\right) \;=\; \bigcup_{j=1}^{N} X^{-1}(B_j)$$

Since $E_j = X^{-1}(B_j) \in \mathcal{B}$ and they are disjoint we get by the finite additivity property of the probability measure $P$,

$$\begin{aligned} P\left(\bigcup_{j=1}^{N} X^{-1}(B_j)\right) &= \sum_{j=1}^{N} P\left(X^{-1}(B_j)\right) \\ &\Longrightarrow \\ \mathcal{P}_X\left(\bigcup_{j=1}^{N} B_j\right) &= \sum_{j=1}^{N} \mathcal{P}_X(B_j) \end{aligned}$$

Hence we get that
$\boldsymbol{P_X}$ **is finitely additive**

4. Similarly we get if $B_n$, $n = 1, 2, 3, \cdots$ is a sequence of mutually disjoint Borel sets in $\mathbb{R}$ then

$$X^{-1}\left(\bigcup_{n=1}^{\infty} B_n\right) \;=\; \bigcup_{n=1}^{\infty} X^{-1}(B_n)$$

Since $E_n = X^{-1}(B_n) \in \mathcal{B}$ for $n = 1, 2, 3, \cdots$, and they are disjoint we get by the countable additivity property of the probability measure $P$,

$$\begin{aligned} P\left(\bigcup_{n=1}^{\infty} X^{-1}(B_n)\right) &= \sum_{n=1}^{\infty} P\left(X^{-1}(B_n)\right) \\ &\Longrightarrow \\ \mathcal{P}_X\left(\bigcup_{n=1}^{\infty} B_n\right) &= \sum_{j=1}^{\infty} \mathcal{P}_X(B_n) \end{aligned}$$

Hence we get that
$\boldsymbol{\mathcal{P}_X}$ **is countably additive additive**

The above properties show that $\mathcal{P}_X$ is a probability measure on the Borel sets of $\mathbb{R}$. We call this the probability measure $\mathcal{P}_X$ on the Borel $\sigma$-algebra $\mathcal{B}$ in $\mathbb{R}$, as the measure induced by $P$ and $X$. We shall now see that this

induced probability measure $\mathcal{P}_X$ can be studied in an easier manner.
Let $X$ be a random variable and $x$ be any real number. Then consider the interval,

$$I_x = (-\infty, x] \tag{3.2.6}$$

Then since $I_x$ is a Borel set,

$$X^{-1}(I_x) \in \mathcal{B} \tag{3.2.7}$$

This means

$$\{\omega : X(\omega) \in I_x\} \in \mathcal{B} \tag{3.2.8}$$

which gives

$$\{\omega : X(\omega) \leq x\} \in \mathcal{B} \quad \forall x \in \mathbb{R} \tag{3.2.9}$$

We observe the following : Suppose now $X : \Omega \longrightarrow \mathbb{R}$ is such that (3.2.9) is true. Then let

$$\mathcal{B}_X = \left\{A \subseteq \mathbb{R} : X^{-1}(A) \in \mathcal{B}\right\} \tag{3.2.10}$$

$\mathcal{B}_X$ is the collection of all the subsets of $\mathbb{R}$ whose preimage by $X$ in $\Omega$ is in $\mathcal{B}$. We can easily see, (using the definition of a $\sigma$-algebra), that $\mathcal{B}_X$ is a $\sigma$-algebra of subsets of $\mathbb{R}$ and since this contains intervals of the form $I_x = (-\infty, x)$, for all $x \in \mathbb{R}$, (by our assumption that (3.2.9) holds). Hence this collection must contain all Borel sets, since the Borel sets form the smallest $\sigma$-algebra containing intervals of this form. Hence we have

$$\begin{aligned} B \in \mathcal{B} &\implies B \in \mathcal{B}_X \\ &\implies X^{-1}(B) \in \mathcal{B} \\ &\implies X \text{is a random variable} \end{aligned}$$

Thus in order to verify that $X$ is random variable it is enough to verify that (3.2.9) hods and hence we take this as the definition of a random variable. We have

**Definition 3.2.2 A function $X : \Omega \longrightarrow \mathbb{R}$ is said to be a random variable on a probability space $(\Omega, \mathcal{B}, P)$ if**

$$\boldsymbol{X^{-1}(-\infty, x] \in \mathcal{B} \text{ for every Borel set } B \text{ in } \mathbb{R}}$$

**that is**

$$\boldsymbol{\{\omega : -\infty < X(\omega) \leq x\} \in \mathcal{B} \text{ for all } x \in \mathbb{R}}$$

**Example 3.2.1** Let us consider the random experiment of tossing a coin thrice. We have

$$\begin{aligned} \Omega &= \{\text{HHH,HHT,HTH,THH,HTT,THT,TTH,TTT}\} \\ \mathcal{B} &= 2^{\Omega} \end{aligned}$$

Let

$$X : \Omega \longrightarrow \mathbb{R}$$

be defined as

$$X(\omega) = \text{the number of Heads in } \omega$$

Thus

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

We have

$$\begin{aligned} \{\omega : X(\omega) \le x\} &= \phi \quad \text{if } x < 0 \\ &= \{\text{TTT}\} \quad \text{if } 0 \le x < 1 \\ &= \{\text{HTT,THT,TTH}\} \quad \text{if } 1 \le x < 2 \\ &= \{\text{HHT,HTH,THH}\} \quad \text{if } 2 \le x < 3 \\ &= \{\text{HHH}\} \quad \text{if } 3 \le x < \infty \end{aligned}$$

In all these cases

$$\{\omega : X(\omega) \le x\} \in \mathcal{B}$$

Hence $X$ is a random variable on this probability space.

Let $\mathcal{R}_X$ denote the set of values taken by the random variable X. (This is just the Range of the function $X$ and is a subset of $\mathbb{R}$). We generally come across two types of random variables.

**Discrete RVs**:
These are random variables for which $\mathcal{R}_X$, that is the values taken by $X$, is either a finite set of real numbers or an infinite sequence of real numbers.

**Nondiscrete RVs**:
These are typically those for which $\mathcal{R}_X$ is either a finite interval or a semiinfinite interval or the full real line.

## 3.3 PMF, CDF and PDF

We have seen above that a function

$$X : \Omega \longrightarrow \mathbb{R}$$

is a random variable if

$$\{\omega \in \Omega : -\infty < X(\omega) \leq x\} \in \mathcal{E} \text{ for all } x \in \mathbb{R}$$

For brevity we shall write the set $\{\omega \in \Omega : -\infty < X(\omega) \leq x\}$ as $\{X(\omega) \leq x\}$ Hence

$$P\{X(\omega) \leq x\}$$

is well-defined. This is called the CUMULATIVE DISTRIBUTION FUNCTION (of the random variable $X$) and is denoted by $F_X(x)$. We have

**Definition 3.3.1 Let $X : \Omega \longrightarrow \mathbb{R}$ be a random variable on a probability space $(\Omega, \mathcal{E}, P)$. The cumulative distribution function (cdf) of $X$ is defined as,**

$$\boldsymbol{F_X(x) = P(\{X(\omega) \leq x\})}$$

For a discrete random variable, for any value $x_j \in \mathcal{R}_X$, we denote by $p_j$ the Probability of $X$ attaining this value, that is,

$$p_j \quad = \quad P(\{X = x_j\}) \tag{3.3.11}$$

The function $p_X$ defined as

$$p_X : \mathcal{R}_X \longrightarrow \mathbb{R} \tag{3.3.12}$$

defined as

$$p_X(x_j) = p_j = P(\{X = x_j\}) \tag{3.3.13}$$

is called the **Probability Mass Function** or in short **PMF** of the random variable $X$.
Let us look at some simple examples of random variables and their cumulative distribution function.

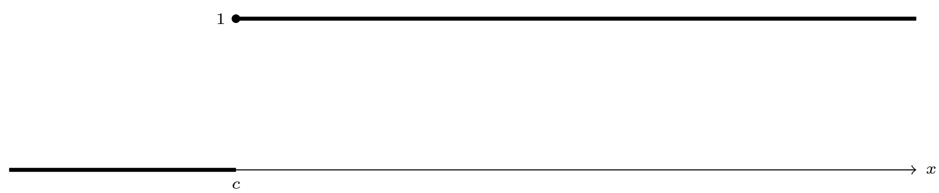
**Example 3.3.1** Consider a random variable $X(\omega)$ which takes only one value say $c$ on the probability space $(\Omega, \mathcal{E}, P)$.
Then we have

$$F_X(x) = P\{\omega : X(\omega) \le x\}$$

is given by

$$\begin{aligned} F_X(x) &= 0 \quad \text{if } \ x < c \\ &= 1 \quad \text{if } \ c \le x < \infty \end{aligned}$$

Thus $F_X(x)$ is a step function, with a jump of one unit at the point $c$, as shown below:




Suppose now $X(\omega)$ takes two values $c_1$ and $c_2$ such that

$$\begin{aligned} P\{\omega : X(\omega) = c_1\} &= p \\ P\{\omega : X(\omega) = c_2\} &= q \end{aligned}$$

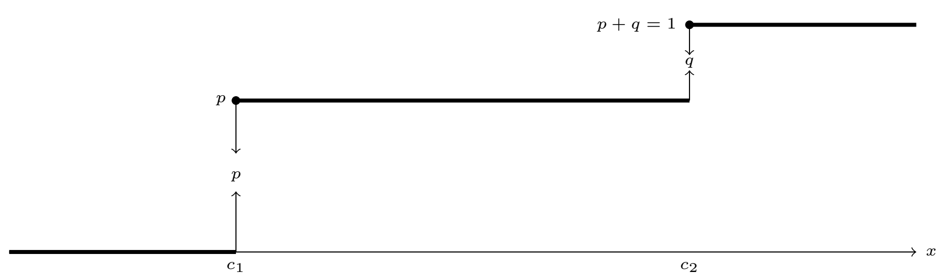
where

$$p + q = 1 \ \ , \ \ 0 \le p, q \le 1.$$

w.l.g. we assume $c_1 < c_2$. Then we have

$$\begin{aligned} F_X(x) &= 0 \qquad \text{if } \ -\infty < x < c_1 \\ &= p \qquad \text{if } \ c_1 \le x < c_2 \\ &= p + q = 1 \ \text{ if } \ c_2 \le x < \infty \end{aligned}$$

$F_X(x)$ is again a step function with two steps, a jump of $p$ at the point $c_1$ and a further jump of $q$ at the point $c_2$, as shown below:



In general if $X(\omega)$ takes $k$ values $c_1 < c_2 < \cdots < c_k$ such that

$$P\{\omega : X(\omega) \leq c_j\} = p_j$$

where

$$0 \leq p_j \leq 1, \ \sum_{j=1}^{k} p_j = 1$$

then

$$F_X(x) = \begin{cases} 0 & \text{if} \quad -\infty < x < c_1 \\ p_1 & \text{if} \quad c_1 \leq x < c_2 \\ p_1 + p_2 & \text{if} \quad c_2 \leq x < c_3 \\ \cdots\cdots\cdots & \\ p_1 + p_2 + \cdots + p_j & \text{if} \quad c_j \leq x < c_{j+1} \\ \cdots\cdots\cdots & \\ 1 & \text{if} \quad c_k \leq x < \infty \end{cases}$$

The graph of $F_X(x)$ will be a piecewise constant graph having jumps $p_j$ at $c_j$, for $j = 1, 2, \cdots, k$.
This is typical of discrete random variables.

**Properties of CDF**:
We shall now look at some properties of the cumulative distribution function $F_X(x)$ of a random variable $X$.

1. $0 \leq F_X(x) \leq 1$ ; $\forall\, x \in \mathbb{R}$
   This is because $F_X(x)$ is the probability of the event $\{\omega : X(\omega) \leq x\}$

2. $x_1 < x_2 \implies F_X(x_1) \leq F_X(x_2)$, that is, $F_X(x)$ is a nondecreasing function.
   This follows from the fact

$$\{\omega : X(\omega) \leq x_1\} \ \subseteq \ \{\omega : X(\omega) \leq x_2\}$$

3. $\lim\limits_{x \longrightarrow \infty} F_X(x) = 1$.
   This follows from the fact for any increasing sequence of real numbers (with $x_n \to \infty$), we have the sequence of sets $E_n = (X \leq x_n)$ is nondecreasing and hence

$$\lim_{n \to \infty} E_n \ = \ \bigcup_{n=1}^{\infty} E_n = \{\omega \in \Omega : X(\omega) < \infty\}$$

Hence by the continuity property of the probability we get

$$\begin{aligned}
P(\lim_{n\to\infty} E_n) &= \lim_{n\to\infty} \mathcal{P}(E_n) \\
\Longrightarrow & \\
P(\Omega) &= \lim_{n\to\infty} F_X(x_n) \\
\Longrightarrow & \\
1 &= \lim_{x\to\infty} F_X(x)
\end{aligned}$$

4. Similarly we can show that $\lim_{x\longrightarrow-\infty} F_X(x) = 0$

5. $F_X(x)$ is right continuous at every $x \in \mathbb{R}$ ; i.e.,

$$\lim_{h\longrightarrow 0+} F_X(x+h) = F_X(x).$$

This follows from the continuity from above property of the probability measure. We have

$$E_n = \left\{X(\omega) \le x + \frac{1}{n}\right\}$$

is a sequence of events decreasing to the event

$$E = \{X(\omega) \le x\}$$

Hence by the property of continuity from above of the probability function we get

$$\begin{aligned}
\mathcal{P}(E) &= \lim_{n\to\infty} \mathcal{P}(E_n) \\
\Longrightarrow & \\
\mathcal{P}(\{X(\omega) \le x\}) &= \mathcal{P}(\left\{X(\omega) \le x + \frac{1}{n}\right\}) \\
\Longrightarrow & \\
F_X(x) &= \lim_{n\longrightarrow\infty} F_X(x + \frac{1}{n}) \\
\Longrightarrow & \\
F_X(x) &= F_X(x+)
\end{aligned}$$

Thus $F_X(x)$ is right continuous at every point $x \in \mathbb{R}$.

If a function $F(x)$ has to be the cumulative distribution function of a random variable, it must satisfy the above five properties.
Next we shall look at the probabilities of the sets $\{X(\omega) \in I\}$ where $I$ is any interval. We observe the following:

1. Since the interval $(x, \infty)$ is the complement of the interval $(-\infty, x]$ we get

$$\mathcal{P}(\{X(\omega) \in (x, \infty\}) = \mathcal{P}(\{X(\omega) \in (-\infty, x]\})$$

Thus

$$\mathcal{P}(\{X(\omega) \in (x, \infty)\}) = 1 - F_X(x) \qquad (3.3.14)$$

2. Next we observe that, if $a < b$ then,

$$\begin{aligned} (a, b] &= (-\infty, b] \setminus (-\infty, a] \\ &\Longrightarrow \\ \mathcal{P}(\{X(\omega) \in (a, b]\}) &= \mathcal{P}(\{X(\omega) \in (-\infty, b]\}) - \mathcal{P}(\{X(\omega) \in (-\infty, a]\}) \\ &\Longrightarrow \\ \mathcal{P}(\{X(\omega) \in (a, b]\}) &= F_X(b) - F_X(a) \end{aligned}$$

Thus we have

$$\mathcal{P}(\{X(\omega) \in (a, b]\}) = F_X(b) - F_X(a) \qquad (3.3.15)$$

3. The sequence intervals

$$I_n = (-\infty, x - \frac{1}{n}]$$

increase to the interval

$$I = (-\infty, x)$$

Hence by the continuity from below property of probability we have

$$\begin{aligned} \mathcal{P}(\{X(\omega) \in (-\infty, x)\}) &= \lim_{n\to\infty} \mathcal{P}(\left\{X(\omega) \in (-\infty, x - \frac{1}{n}]\right\}) \\ &\Longrightarrow \\ \mathcal{P}(\{X(\omega) \in (-\infty, x)\}) &= \lim_{n\to\infty} F_X(x - \frac{1}{n}) \\ &= F_X(x-), \text{ (the left hand limit of } F_X(x) \text{ at the point } x) \end{aligned}$$

Thus we have

$$\mathcal{P}(\{X(\omega) \in (-\infty, x)\}) \;=\; F_X(x-) \tag{3.3.16}$$

4. Analogously, if $a \leq b$ we can show that

$$\begin{aligned}\mathcal{P}(\{X(\omega) \in (a, b)\}) &= F_X(b-) - F_X(a) &(3.3.17)\\ \mathcal{P}(\{X(\omega) \in [a, b)\}) &= F_X(b-) - F_X(a-) &(3.3.18)\\ \mathcal{P}(\{X(\omega) \in [a, b]\}) &= F_X(b) - F_X(a-) &(3.3.19)\end{aligned}$$

5. We can write the singleton set $\{x\}$ as

$$\{X = x\} \;=\; \{X \in (a, x]\} \setminus \{X \in (a, x)\} \text{ for any } a < x$$

Hence we get

$$\begin{aligned}\mathcal{P}(\{X = x\}) &= \mathcal{P}(\{X \in (a, x]\}) - \mathcal{P}(\{X \in (a, x)\})\\ &= [F_X(x) - F_X(a)] - [F_X(x-) - F_X(a)]\\ &= F_X(x) - F_X(x-)\end{aligned}$$

Thus we have

$$\mathcal{P}(\{X = x\}) \;=\; F_X(x) - F_X(x-) \tag{3.3.20}$$

A random variable is said to be **continuous** if $F_X(x)$ is continuous at all $x$, that is, if $F_X(x)$ is also left continuous, (since we know that it is already right continuous). For continuous random variables we have $F_X(x-) = F_X(x+) = F_X(x)$ for all $x \in \mathbb{R}$. Hence we have

$$\mathcal{P}(\{X = x\}) \;=\; 0 \text{ for any continuous random variable } X \tag{3.3.21}$$

From this it follows that for a continuous random variable, for any finite interval $I$ whose left and right end points are $a$ and $b$ respectively,

$$\mathcal{P}(\{X \in I\}) \;=\; F_X(b) - F_X(a) \tag{3.3.22}$$

irrespective of whether the end points are in $I$ or not.
**Probability Density Function (PDF)**:
Consider a continuous random variable $X$ with CDF given by $F_X(x)$. Since

$F_X(x)$ is a continuous nondecreasing function, its derivatives exist except possibly at a sequence of points in $\mathbb{R}$, (the sequence can be arranged in an increaseing order). Let $f_X(x)$ be the function derived as follows:

$$f_X(x) = \begin{cases} \frac{d}{dx}F_X(x) \text{ whenever the derivative exists at } x \text{ and,} \\ \text{any arbitrary nonnegative real value at other points} \end{cases} \tag{3.3.23}$$

This function $f_X(x)$ is called the Probability Density Function of the random variable $X$. Since the function $F_X(x)$ is nondecreasing, its derivative is nonnegative, whenever it exists. Hence we have

$$f_X(x) \geq 0 \text{ for all } x \in \mathbb{R} \tag{3.3.24}$$

Moreover, we have

$$F_X(x) = \int_{-\infty}^{x} f_X(s)ds \tag{3.3.25}$$

Since $F_X(-\infty, \infty) = 1$ we have

$$\int_{-\infty}^{\infty} f_X(x)dx = 1 \tag{3.3.26}$$

For any finite interval $I$ whose left and right end points are $a$ and $b$ respectively, we have

$$\mathcal{P}(\{X \in I\}) = \int_{a}^{b} f_X(x)dx \tag{3.3.27}$$

irrespective of whether the end points are in $I$ or not.

**Remark 3.3.1** For a discrete random variable, the pdf will involve the delta function. If the discrete random variable $X$ takes the values $x_1 < x_2 < x_3 < \cdots x_j < \cdots$ with probabilities $p_1, p_2, \cdots p_j, \cdots$ then we have to define

$$f_X(x) = \sum_j p_j \delta(x - x_j) \tag{3.3.28}$$

**Functions of a Random Variable**

Let $X$ be a random variable on a probability space $(\Omega, \mathcal{B}, P)$. (Without loss of generality let us assume that $X$ is a continuous random variable). Let $\mathcal{R}_X$ be the range of $X$. Suppose

$$g : \mathcal{R}_X \longrightarrow \mathbb{R}$$

is an increasing function defined on $\mathcal{R}_X$. Then for every $y \in$ Range of $g$ there exists a unique $x \in \mathcal{R}_X$ such that $g(x) = y$ and hence $x = g^{-1}(y)$. We define a new random variable as follows:

$$Y = g(X)$$

Then we have, for any $y \in \mathbb{R}$

$$\begin{array}{rcl}
Y(\omega) \leq y & \Longleftrightarrow & g(X(\omega)) \leq y \\
& \Longleftrightarrow & X(\omega) \leq g^{-1}(y) \\
\Longrightarrow & & \\
P(Y \leq y) & = & P(X \leq g^{-1}(y)) \\
& = & F_X(g^{-1}(y)) \\
\Longrightarrow & & \\
F_Y(y) = P(g(X) \leq y) & = & F_X(g^{-1}(y))
\end{array}$$

Similarly for reasonably smooth functions $g$ we define $Y = g(X)$ and

$$F_Y(y) \;=\; P(g(X) \leq y) \tag{3.3.29}$$

**Example 3.3.2** Let $X$ be a random variable with CDF $F_X(x)$. Let $Y = |X|$. Then $Y$ is a nonnegative random variable and hence $F_Y(y) = 0$ for $y < 0$. For $y \geq 0$ we have

$$\begin{array}{rcl}
P(Y \leq y) & = & P(|X| \leq y) \\
& = & P(-y \leq X \leq y) \\
& = & F_X(y) - F_X(-y-)
\end{array}$$

If $X$ is a continuous random variable then $F_Y(-y-) = F_Y(-y)$ and we get

$$F_Y(y) \;=\; F_X(y) - F_X(-y)$$

**Example 3.3.3** Let $X$ be a continuous random variable with CDF $F_X(x)$. Let $Y = aX + b$ where $a \neq 0$. We have

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(aX + b \leq y) \\ &= P(aX \leq y - b) \end{aligned}$$

Case 1: $a > 0$
In this case we have

$$\begin{aligned} P(Y \leq y) &= P(aX \leq y - b) \\ &= P(X \leq \frac{y-b}{a}) \\ &= F_X\left(\frac{y-b}{a}\right) \end{aligned}$$

Case 2: $a < 0$
We have

$$\begin{aligned} P(Y \leq y) &= P(aX \leq y - b) \\ &= P(X \geq \frac{y-b}{a}) \\ &= 1 - P(X < \frac{y-b}{a}) \\ &= 1 - F_X\left(\frac{y-b}{a}\right) \text{ (since } X \text{ is continuous)} \end{aligned}$$

Thus we have

$$F_Y(y) = \begin{cases} F_X\left(\frac{y-b}{a}\right) & \text{if } a > 0 \\ 1 - F_X\left(\frac{y-b}{a}\right) & \text{if } a < 0 \end{cases}$$

Differentiating we get

$$\begin{aligned} f_Y(y) &= \begin{cases} \frac{1}{a} f_X\left(\frac{y-b}{a}\right) & \text{if } a > 0 \\ -\frac{1}{a} f_X\left(\frac{y-b}{a}\right) & \text{if } a < 0 \end{cases} \\ &= \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right) \text{ for all } y \in \mathbb{R} \text{ and all } a \in \mathbb{R}, (a \neq 0) \end{aligned}$$

**Example 3.3.4** Consider a continuous real valued random variable $X$ with CDF $F_X(x)$. Let $Y = X^2$. Then $Y$ is a nonnegative random variable. Hence $F_Y(y) = 0$ for $y < 0$. For $y \geq 0$ we have

$$\begin{aligned} P(Y \leq y) &= P(X^2 \leq y) \\ &= P(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}) \end{aligned}$$

Hence we get

$$F_Y(y) = \begin{cases} 0 & \text{for} \quad y < 0 \\ F_X(\sqrt{y}) - F_X(-\sqrt{y}) & \text{for} \quad y > 0 \end{cases}$$

Differentiating we get the pdf as

$$f_Y(y) = \begin{cases} 0 & \text{for} \quad y < 0 \\ \frac{f_X(\sqrt{y})}{2\sqrt{y}} + \frac{f_X(-\sqrt{y})}{2\sqrt{y}} & \text{for} \quad y > 0 \end{cases}$$

# 3.4 Discrete Random Variables

We shall now consider some standard models of discrete random variables. We shall first consider discrete random variables which take only a finite number of values. We shall denote the set of values taken by the random variable $X$ by $\mathcal{R}_X$

## 3.4.1 Bernoulli Random Variables

A random variable which takes only one value, say $C$, is the constant random variable and for this random variable we have the PMF defined as

$$P(X = C) = 1 \tag{3.4.1}$$

Therefore the simplest nontrivial random variable is that which taks two values. Without loss of generality we shall take the two values to be 0 and 1. Hence we have

$$\mathcal{R}_X = \{0, 1\} \tag{3.4.2}$$

Hence we have the PMF

$$\begin{aligned} p_1 &= P(X=1) \\ p_0 &= P(X=0) \end{aligned}$$

If we denote $p_1$ by $p$ then we must have $p_0 = 1 - p$. Thus we take a real number $p$ such that $0 < p < 1$ and have the PMF as

$$p_1 = P(X=1) = p \tag{3.4.3}$$

$$p_0 = P(X=0) = 1 - p \tag{3.4.4}$$

Such random variables are called Bernoulli Random Variables

**Bernoulli Random Variable**

$\mathcal{R}_X = \{0, 1\}$

**PMF:**

$p_1 = P(X=1) = p$

$p_0 = P(X=0) = 1 - p$

$0 < p < 1$

### 3.4.2 Uniform Random Variables

We shall next consider random variables taking $N$ values in general. Let us say these $N$ values are $x_1, x_2, \cdots, x_N$. Then we have

$$\mathcal{R}_X = \{x_1, x_2, \cdots, x_N\} \tag{3.4.5}$$

Suppose the random variable takes these values with equal probability $p$ then we have

$$p_k = P(X = x_k) = p \text{ for } k = 1, 2, \cdots, N$$

Since we must have the sum of all the $p_k$ to be one we get

$$p_k = \frac{1}{N}$$

Such random variables are called Uniform Discrete Random Variables.

**Uniform Discrete Random Variable**

$$\mathcal{R}_X = \{x_1, \cdots, x_N\}$$

**PMF:**

$$p_k = P(X = k) = \tfrac{1}{N}$$

**Remark 3.4.1** **Some Generic Ideas**:
In general we can choose $N$ positive real numbers $p_1, p_2, \cdots, p_N$ such that

$$0 \leq p_k \leq 1 \text{ for } 1 \leq k \leq N \text{ and}$$
$$\sum_{k=1}^{N} p_k = N$$

Then get a PMF as

$$P(X = k) \;=\; p_k$$

Different choices of $p_k$ as above give rise to different random variables. Some popular such choices are given in the sections below

### 3.4.3 Binomial Distribution

Consider a random variable which takes the values $0, 1, 2, \cdots, N$. We then have

$$\mathcal{R}_X \;=\; \{0, 1, 2, \cdots, N\} \tag{3.4.6}$$

Let $0 < p < 1$. We have by Binomial Theorem,

$$1 = (p + (1-p))^N \;=\; \sum_{k=0}^{N} \binom{N}{k} p^k (1-p)^{N-k} \tag{3.4.7}$$

We then can define the PMF as

$$p_k = P(X = k) \;=\; \binom{N}{k} p^k (1-p)^{N-k} \text{ for } k = 0, 1, 2 \cdots, N \tag{3.4.8}$$

Such random variables are called Binomial Random Variables.

**Remark 3.4.2** We can replace $\mathcal{R}_X$ above by

$$\mathcal{R}_X = \{x_0, x_1, x_2, \cdots, x_{_N}\} \tag{3.4.9}$$

where $x_j$ are real numbers.

**Binomial Random Variable**

$$\mathcal{R}_X = \{x_0, x_1, \cdots, x_N\}$$

**PMF:**

$$p_k = P(X = x_k) = \binom{N}{k} p^k (1-p)^{N-k}$$

**for** $k = 0, 1, 2 \cdots, N$

### 3.4.4 Zipf RV

Consider a random variable $X$ for which again

$$\mathcal{R}_X = \{x_1, x_2, \cdots, x_N\} \tag{3.4.10}$$

where $x_1 < x_2 < \cdots < x_N$ are real numbers. (We can for example take $x_1 = 1, x_2 = 2, \cdots, x_N = N$). Suppose the random variable is such that the lower values are attained with higher probability and higher values are attained with less probability. In particulr, for example, suppose $P(X = x_k)$ is proportional to $\frac{1}{k}$, that is

$$p_k = P(X = k) \propto \frac{1}{k} \tag{3.4.11}$$

Let $C$ be the constant of proportionality. Then we have

$$p_k = P(X = k) = C\frac{1}{k} \tag{3.4.12}$$

Since the total probability must be one we get

$$\sum_{k=1}^{N} C\frac{1}{k} = 1$$

which gives us

$$C = \frac{1}{s_{_N}}$$

where

$$s_{_N} = \sum_{k=1}^{N} \frac{1}{k} = 1 = \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{N}$$

Thus we have

$$p_k = P(X = k) = \frac{1}{s_{_N}} \frac{1}{k} \tag{3.4.13}$$

**Zipf Random Variable**

$\mathcal{R}_X = \{x_1, \cdots, x_N\}$

**PMF:**

$p_k = P(X = x_k) = \frac{1}{s_N} \frac{1}{k}$

(**for** $k = 0, 1, 2 \cdots, N$)

$s_{_N} = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{N}$

We can reverse the situation above and get a random variable which takes higher values with higher probabilities. Let us define

$$\mathcal{R}_X = \{x_1, x_2, \cdots, x_N\} \tag{3.4.14}$$

where as before, $x_1 < x_2 < \cdots < x_N$. We can define the PMF as

$$P(X = x_k) = \frac{1}{s_N} \frac{1}{N + 1 - k} \tag{3.4.15}$$

so that we get

$$P(X = x_1) = \tfrac{1}{s_N} \tfrac{1}{N},$$

$$P(X = x_2) = \tfrac{1}{s_N} \tfrac{1}{N-1}, \cdots,$$

$$P(X = x_{_{N-1}}) = \tfrac{1}{s_N} \tfrac{1}{2}, \text{and}$$

$$P(X = x_{_N}) = \tfrac{1}{s_N} \tfrac{1}{1}$$

We can generalize this further as follows:
Let $a_1, a_2, \cdots, a_N$ be a sequence of positive real numbers, such that

$$a_1 < a_2 < \cdots < a_N \tag{3.4.16}$$

Let

$$C = \sum_{k=1}^{N} a_k \tag{3.4.17}$$

Then for a random variable $X$ for which

$$\mathcal{R}_X = \{x_1, x_2, \cdots, x_N\} \tag{3.4.18}$$

where $x_1 < x_2 < \cdots < x_N$ we can define the PMF as

$$p_k = P(X = x_k) = \frac{a_k}{C} \tag{3.4.19}$$

Thus we get the probability that $X$ attains lower values is higher than that of attaining higher values. We can again reverse the situation and define

$$p_k = P(X = x_k) = \frac{a_{(N+1-k)}}{C} \tag{3.4.20}$$

Now the higher values are attained with higher probabilities.
We shall next look at some discrete random variables which take an infinite sequence of values.

### 3.4.5 Discrete Random Variables taking an infinite sequence of values

We shall look at some generic ideas. Let $X$ be a random variable taking an infinite sequence of values. We look at the following situations:

$$\mathcal{R}_X = \{1, 2, 3 \cdots\} \tag{3.4.21}$$

(This is the random variable which counts the number of tosses when we toss a coin until we get a Head). We also look at

$$\mathcal{R}_X = \{0, 1, 2, 3 \cdots\} \tag{3.4.22}$$

(This is the random variable which counts the number of tails when we toss a coin until we get a Head).
For the first case we consider an infinite sequence $a_1, a_2, \cdots, a_k, \cdots$ of positive real numbers such that the infinite series

$$\sum_{k=1}^{\infty} a_k$$

converges. Let the sum of the series be $S$. Now if we let

$$p_k = \frac{a_k}{S}, \ k = 1, 2, \cdots$$

we get that

$$\sum_{k=1}^{\infty} p_k = 1$$

This gives rise to the following PMF:

$$p_k = P(X = k) \ = \ \frac{a_k}{S}, \ k = 1, 2, \cdots \qquad (3.4.23)$$

In the second case where the random variable takes the values $0, 1, 2, \cdots$, we consider an infinite sequence $a_0, a_1, a_2, \cdots, a_k, \cdots$ of positive real numbers such that the infinite series

$$\sum_{k=0}^{\infty} a_k$$

converges. Let the sum of the series be $S$. Now if we let

$$p_k = \frac{a_k}{S}, \ k = 0, 1, 2, \cdots$$

we get that

$$\sum_{k=0}^{\infty} p_k = 1$$

This gives rise to the following PMF:

$$p_k = P(X = k) \ = \ \frac{a_k}{S}, \ k = 0, 1, 2, \cdots \qquad (3.4.24)$$

By choosing suitable $a_k$ we get different random variables. We now look at some standard such models

### 3.4.6 Geometric Random Variable

Let $q$ is a real number such that $0 < q < 1$. Let $a_k = q^k,\ k = 0, 1, 2, \cdots$. Then we have

$$\sum_{k=0}^{\infty} a_k = \sum_{k=0}^{\infty} q^k = \frac{1}{1-q} \tag{3.4.25}$$

We shall denote $p = 1 - q$. Then we have

$$\sum_{k=0}^{\infty} a_k = \frac{1}{p} \tag{3.4.26}$$

Hence we define

$$p_k = pa_k,\ k = 0, 1, 2, \cdots \tag{3.4.27}$$

Then we get

$$\sum_{k=0}^{\infty} p_k = 1 \tag{3.4.28}$$

Thus for a random variable $X$ taking values $0, 1, 2, \cdots$ this gives rise to the PMF

$$\begin{aligned} p_k = P(X = k) &= pa_k = pq^k \\ &= p(1-p)^k \end{aligned}$$

We also have if we define

$$a_k = q^k,\ k = 1, 2, \cdots$$

we get

$$\begin{aligned} \sum_{k=1}^{\infty} a_k &= \sum_{k=1}^{\infty} q^k \\ &= \frac{q}{1-q} \\ &= \frac{1-p}{p} \end{aligned}$$

hence if we define

$$p_k = \frac{p}{1-p} a_k$$

we get

$$\sum_{k=1}^{\infty} p_k = 1$$

Thus for arandom variable $X$ taking values $1, 2, \cdots$ this gives rise to the PMF

$$\begin{aligned} p_k = P(X = k) &= \frac{p}{1-p} a_k \\ &= \frac{p}{1-p} q^k \\ &= \frac{p}{1-p} (1-p)^k \\ &= p(1-p)^{(k-1)} \end{aligned}$$

These are called the Geometric Random Variables

**Geometric Random Variable TYPE 1**

$$\mathcal{R}_X = \{1, 2, \cdots\}$$

**PMF:**

$$p_k = P(X = k) = p(1-p)^{(k-1)},\ k = 1, 2, \cdots$$
$$(0 < p < 1)$$

**Geometric Random Variable TYPE 2**

$$\mathcal{R}_X = \{0, 1, 2, \cdots\}$$

**PMF:**

$$p_k = P(X = k) = p(1-p)^{k},\ k = 1, 2, \cdots$$
$$(0 < p < 1)$$

### 3.4.7 Exponential Random Varible

For any real number $\lambda$ let

$$a_k = \frac{\lambda^k}{k!},\ k = 0, 1, 2, \cdots$$

Then we have

$$\sum_{k=0}^{\infty} a_k = e^{\lambda}$$

Hence if we define

$$p_k = e^{-\lambda} a_k,\ k = 0, 1, 2, \cdots$$

then we have

$$\sum_{k=0}^{\infty} p_k = 1$$

Thus for a random variable $X$ taking values $0, 1, 2, \cdots$ tis gives rise to the PMF

$$\begin{aligned} p_k = P(X = k) &= e^{-\lambda} a_k \\ &= e^{-\lambda} \frac{\lambda^k}{k!},\ k = 0, 1, 2, \cdots \end{aligned}$$

Such a random variable is called Exponential Random Variable.

**Exponential Random Variable**

$\mathcal{R}_X = \{0, 1, 2, \cdots\}$

**PMF:**

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

$(\lambda \in \mathbb{R})$

## 3.5 Continuous Random Variables

### 3.5.1 Some Standard Models

Let $X$ be a continuous random variable and let $\mathcal{R}_X$ be the Range of $X$, that is $\mathcal{R}_X$ is the set of values taken by the random variable $X$. We shall consider the following three types of continuous random variables:

1. Random Variables for which the Range is a finite interval, that is,

   $$\mathcal{R}_X = [a, b] \text{ where } -\infty < a \leq x < b < \infty$$

   These are called **Bounded random Variables**

2. Random Variables for which the Range is a semi infinite interval, that is,

   $$\mathcal{R}_X = [0, \infty)$$

   These are called **Random Variables Bounded Below**. (Without loss of generality we have taken the lower bound to be 0)

3. Random Variables for which the Range is the full infinite interval, that is,

   $$\mathcal{R}_X = (-\infty, \infty)$$

   These are called **Unbounded Random Variables**

We shall look at some standard models of each of these types.

### 3.5.2 Bounded Random Variables

As above let

$$\mathcal{R}_X = [a, b] \text{ where } -\infty < a \leq x < b < \infty \tag{3.5.1}$$

For such a random variable clearly we have

$$P(X \leq x) = 0 \text{ if } x \leq a$$

since all the values of $X$ are $\geq a$. Hence we must have

$$F_X(x) = P(X \leq x) = 0 \text{ if } x \leq a \tag{3.5.2}$$

Further we must have

$$P(X \leq x) = 1 \text{ if } x > b$$

since all the values of $X$ are $\leq b$. Hence we must have

$$F_X(x) = P(X \geq x) = 1 \text{ if } x > b \tag{3.5.3}$$

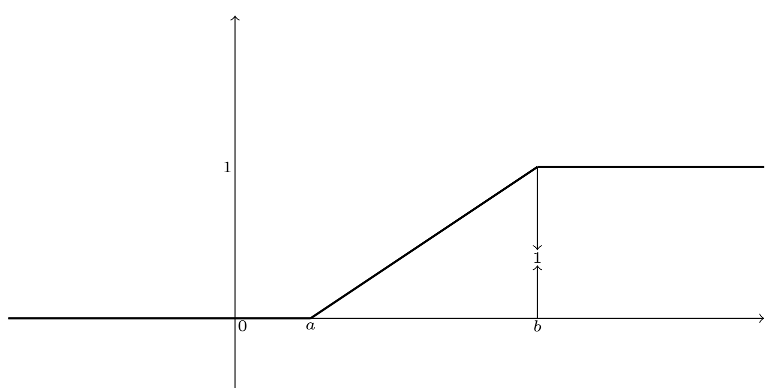
Thus different such random variables are obtained depending on how $F_X(x)$ increases from 0 at $x = a$ to to 1 at $x = b$. The simplest model is obtained by making $F_X(x)$ vary linearly from 0 at $x = a$ to 1 at $x = b$. Thus $F_X(x)$ must be given by

$$F_X(x) = \frac{x-a}{b-a} \text{ for } a \leq x \leq b \tag{3.5.4}$$

Combining all these we get

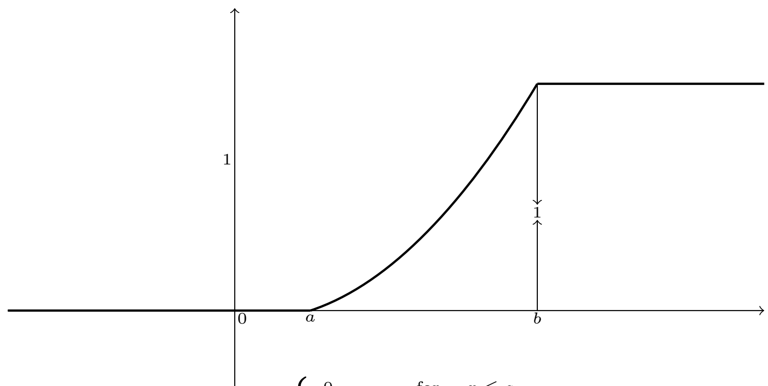
$$F_X(x) = \begin{cases} 0 & \text{for } x \leq a \\ \frac{x-a}{b-a} & \text{for } a \leq x \leq b \\ 1 & \text{for } x \geq b \end{cases} \tag{3.5.5}$$

Such a Random Variable is said to be **Uniformly Distributed** over the interval $[a, b]$. We call such RVs as **Uniform Random Variables**. The garph of the CDF is as shown below:

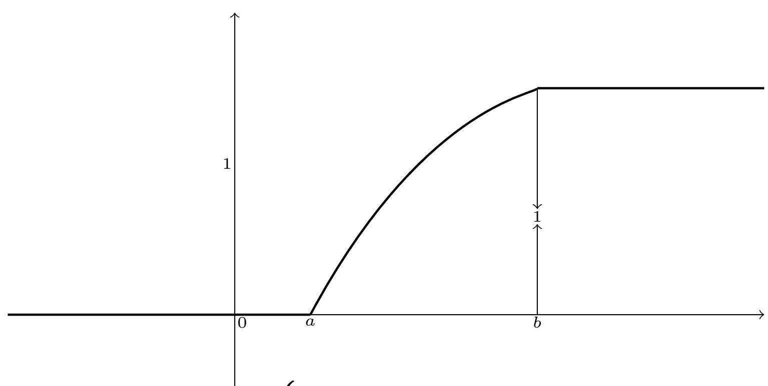



$$F_X(x) = \begin{cases} 0 & \text{for} \quad x \leq a \\ \frac{x-a}{b-a} & \text{for} \quad a \leq x \leq b \\ 1 & \text{for} \quad x \geq b \end{cases}$$

We can get various other random variables with $\mathcal{R}_X = [a, b]$ by choosing different functions that increase from the value 0 at $x = a$ to the value 1 at $x = b$. We give below some examples:




$$F_X(x) = \begin{cases} 0 & \text{for} \quad x \le a \\ \frac{(x-a)^2}{(b-a)^2} & \text{for} \quad a \le x \le b \\ 1 & \text{for} \quad x \ge b \end{cases}$$




$$F_X(x) = \begin{cases} 0 & \text{for} \quad x \le a \\ \frac{e^{-a}-e^{-x}}{e^{-a}-e^{-b}} & \text{for} \quad a \le x \le b \\ 1 & \text{for} \quad x \ge b \end{cases}$$

In general, we can take any continuous function $g(x)$ defined in $[a, b]$ and such that $g(a) = 0$ and $g(b) = 1$ and get a random variable with CDF as

$$F_X(x) = \begin{cases} 0 & \text{for} \quad x \le a \\ g(x) & \text{for} \quad a \le x \le b \\ 1 & \text{for} \quad x \ge b \end{cases} \tag{3.5.6}$$

### 3.5.3 Random Variables Bounded Below

We shall next look at Random Variables for which

$$\mathcal{R}_X \;=\; [0, \infty)$$

Clearly for such random variables we must have

$$P(X \leq x) \;=\; 0 \text{ for } x < 0$$

since $X$ does not take any negative values. Hence we must have

$$F_X(x) \;=\; 0 \text{ for } x < 0 \tag{3.5.1}$$

On $[0, \infty)$, we must have $F_X(x)$ to be a increasing, continuous function such that

$$F_X(0) \;=\; 0 \text{ and} \tag{3.5.2}$$

$$\lim_{x \to +\infty} F_X(x) \;=\; 1 \tag{3.5.3}$$

We shall now look at examples of such random variables

### 3.5.4 Exponential Random Variable

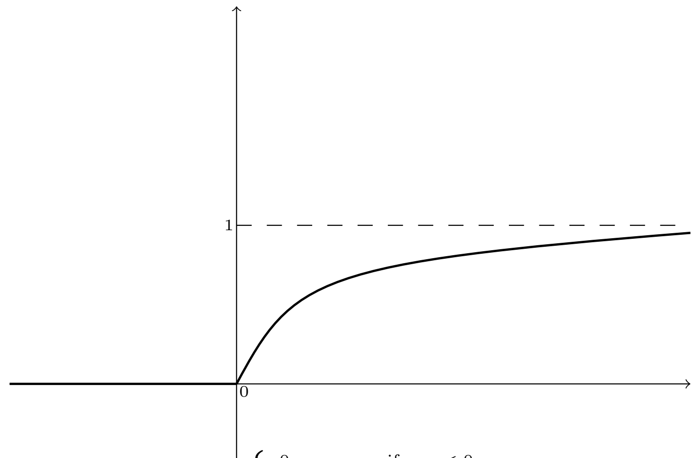
The function $g(x) = e^{-\lambda x}$, (where $\lambda$ is real and $> 0$), is a decreasing function in $[0, \infty)$ decreasing from 1 to 0. Hence the function $-g(x) = -e^{-\lambda x}$ is an increasing function in $[0, \infty)$ increasing from $-1$ to 0. Consequently the function

$$h(x) \;=\; 1 - e^{-\lambda x}$$

is an increasing function in $[0, \infty)$ increasing from 0 to 1. Thus we can $F_X(x)$ to be this function in $[0, \infty)$. Hence we can have a random variable $X$ whose CDF is of the form,

$$F_X \;=\; \begin{cases} 0 & \text{for} \quad x < 0 \\ 1 - e^{-\lambda x} & \text{for} \quad x \geq 0 \end{cases} \tag{3.5.1}$$

The graph of such a function is sketched below:




$$F_X(x) = \begin{cases} 0 & \text{if} \quad x < 0 \\ 1 - e^{-\lambda x} & \text{if} \quad 0 \leq x < \infty \end{cases}$$

Such a random variable is called a **Exponential Random Variable** (with parameter $\lambda > 0$).

### 3.5.5 Rayleigh Random Variable

The function

$$h(x) \quad = \quad 1 - e^{-\beta^2 x^2} \text{ where } \beta \text{ is real} \tag{3.5.1}$$

increases from 0 to 1 in the interval $[0, \infty)$. Thus we can have a random variable with the CDF as

$$F_X \quad = \quad \begin{cases} 0 & \text{for} \quad x < 0 \\ 1 - e^{-\beta^2 x^2} & \text{for} \quad x \geq 0 \end{cases} \tag{3.5.2}$$

Such a random variable is called **Rayleigh Random Variable** (with parameter $\beta$).

### 3.5.6 Pareto Random Variable

We can also have a random variable $X$ for which $\mathcal{R}_X = [a, \infty)$ for some $a > 0$. Then the CDF will be 0 for $x < a$ and a continuous function in $[a, \infty)$ increasing from the value 0 at $a$ to the value 1 at $\infty$. We can modify the Exponential CDF above as follows:

$$F_X(x) \quad = \quad \begin{cases} 0 & \text{for} \quad x < 0 \\ \frac{e^{-\lambda a} - e^{-\lambda x}}{e^{-\lambda a}} & \text{for} \quad x \geq a (\text{where } \lambda > 0) \end{cases} \tag{3.5.1}$$

We can also modify the Rayleigh distribution as follows:

$$F_X(x) = \begin{cases} 0 & \text{for} \quad x < a \\ \frac{e^{-\beta^2 a^2} - e^{-\beta^2 x^2}}{e^{-\beta^2 a^2}} & \text{for} \quad x \geq a \text{ (where } \beta \text{ is real)} \end{cases} \tag{3.5.2}$$

We can also make the CDF vary at rate different from the exponentials. For example let $N$ be any positive integer. Then the function $g(x) = \frac{1}{x^N}$ is a decreasing function in the interval $(a, \infty)$. Hence the function $-g(x)$ is increasing from $-\frac{1}{a^N}$ at $x = a$ to the value 0 as $x \to \infty$. Thus the function

$$-\frac{a^N}{x^N}$$

increases from the value $-1$ at $x = a$ to the value 0 as $x \to \infty$. Finally we get, therefore, the function

$$h(x) = 1 - \frac{a^N}{x^N}$$

increases from the value 0 at $x = a$ to the value 1 as $x \to \infty$. Hence we can have a random variable $X$ for which the CDF is given by

$$F_X(x) = \begin{cases} 0 & \text{for} \quad x < a \\ 1 - \frac{a^N}{x^N} & \text{for} \quad x \geq a \end{cases} \tag{3.5.3}$$

The increase from 0 at $x = a$ to 1 as $x \to \infty$ is now at a polynomial rate. Such a random variable is called a **Pareto Random Variable**.

### 3.5.7 Unbounded Random Variables

We shall next consider some examples of unbounded random variables, that is , random variables $X$ for which $\mathcal{R}_X = (-\infty, \infty)$. The CDF of such random variables must be continuous functions defined on $(-\infty, \infty)$ , and such that $\lim\limits_{x \to -\infty} F_X(x) = 0$ and $\lim\limits_{x \to +\infty} F_X(x) = 1$. We shall look at such models below:

**Laplace Random Variable**

The function,

$$F_X(x) = \begin{cases} \frac{1}{2} e^{\lambda x} & \text{if} \quad x < 0 \\ \\ 1 - \frac{1}{2} e^{-\lambda x} & \text{if} \quad 0 \leq x < \infty \end{cases} \tag{3.5.1}$$

(where $\lambda > 0$), satisfies all the requirements above. A random variable with the above CDF is called a **Laplace Random Variable** (with parameter $\lambda$). For the Laplace Random Variable we observe the following:

$$P(X \leq 0) \;=\; F_X(0) = \frac{1}{2} \tag{3.5.2}$$

Hence we see that

$$\begin{aligned} P(X \geq 0) &= 1 - P(X \leq 0) && (3.5.3) \\ &= 1 - \frac{1}{2} = \frac{1}{2} && (3.5.4) \end{aligned}$$

Hence the random variable takes negative values and positive values with equal probability. We can also have Random Variables for which these two probabilities are not equal. For example, let $\alpha$ be a real number such that $0 < \alpha < 1$. Then we can take

$$F_X(x) \;=\; F_X(x) = \begin{cases} \alpha e^{\lambda x} & \text{if} \quad x < 0 \\ & \\ 1 - (1-\alpha)e^{-\lambda x} & \text{if} \quad 0 \leq x < \infty \end{cases} \tag{3.5.5}$$

satisfies all the requirements for a CDF. A Random Variable $X$ with the above CDF satisfies

$$\begin{aligned} P(X \leq 0) &= \alpha && (3.5.6) \\ P(X > 0) &= 1 - \alpha && (3.5.7) \end{aligned}$$

When $\alpha = \frac{1}{2}$ this reduces to the Laplace Random Variable.
<u>**Cauchy Random Variable**</u>:
The function

$$F_X(x) \;=\; \frac{1}{2} + \frac{1}{\pi} tan^{-1}\left(\frac{x}{\alpha}\right) \tag{3.5.8}$$

satisfies the requirements of a CDF. A random variable with this CDF is called a **Cauchy Random Variable**, with parameter $\alpha$. We again observe that the Cauchy Random Variable takes negative values with the same probability as it takes positive values. We can alter this by considering the following CDF: Let $0 < \beta < 1$

$$F_X(x) \;=\; \begin{cases} \beta + \frac{2\beta}{\pi} tan^{-1}\left(\frac{x}{\alpha}\right) & \text{for} \quad x < 0 \\ \beta + \frac{2(1-\beta)}{\pi} tan^{-1}\left(\frac{x}{\alpha}\right) & \text{for} \quad x \geq 0 \end{cases} \tag{3.5.9}$$

For this random variable we have

$$
\begin{aligned}
P(X < 0) &= \beta \qquad (3.5.10) \\
P(X > 0) &= 1 - \beta \qquad (3.5.11)
\end{aligned}
$$

If $\beta < \frac{1}{2}$ it takes positive values with higher probability than negative vales, and vice versa if $\beta > \frac{1}{2}$. If $\beta = \frac{1}{2}$ we get the Cauchy Random Variable which takes both positive and negative values with equal probabilities.

## 3.6 Conditional Probability

Consider a random experiment and the associated probability space $(\Omega, \mathcal{B}, P)$. Let us now begin with a simple example

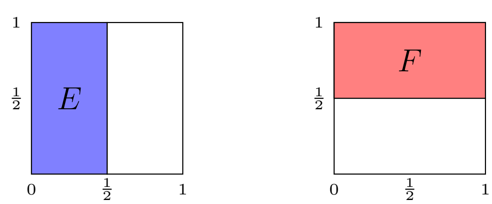
**Example 3.6.1** Consider the random experiment of choosing a point at random in the unit square $0 \leq x \leq 1, 0 \leq y \leq 1$ and noting its coordinates. We then have

$$\Omega = \{(x, y) : 0 \leq x \leq 1, 0 \leq y \leq 1\}$$

We shall consider subrectangles of the unit square as the elementary events and the probability of such events as the area of the rectangle. We shall as usual take the Borel subsets of this unit square as the collection $\mathcal{B}$ of all events and extended concept of area to the Borel sets as the Probability measure $P$. Consider the following two events in this experiment:

$$
\begin{aligned}
E &= \left\{(x, y) : 0 \leq x \leq \frac{1}{2}, 0 \leq y \leq 1\right\} \\
F &= \left\{(x, y) : 0 \leq x \leq 1, \frac{1}{2} \leq y \leq 1\right\}
\end{aligned}
$$

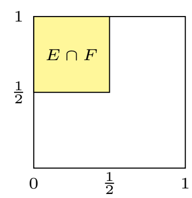
These events are sketched below:



We have

$$P(E) = \frac{1}{2}$$
$$P(F) = \frac{1}{2}$$

Let us now look at the proportion of $F$ in $E$. We have $F \cap E$ as shown in Figure below:



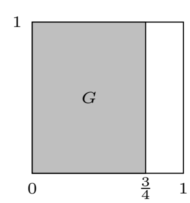

Then we have

$$P(E \cap F) = \frac{1}{4}$$

Hence the proportion of $F$ in $E$ is given by

$$\begin{aligned} \frac{P(E \cap F)}{P(E)} &= \frac{(\frac{1}{4})}{(\frac{1}{2})} \\ &= \frac{1}{2} \\ &= P(F) \end{aligned}$$

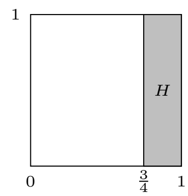
Thus the knowledge of the fact that the event $E$ has occurred did not alter the Probability of occurrence of $F$. Now consider the event $G$ sketched below:




Then we see that $E \cap G = E$ and hence

$$\frac{P(E \cap G)}{P(E)} = 1$$

In this case we see that since the knowledge of occurrence of the event has enhanced completely the probability of occurrence of $G$. On the other hand consider the event $H$ sketched below:




Then we see that $H \cap E = \phi$ and hence

$$\frac{P(G \cap E)}{P(E)} = 0$$

Hence in this case the knowledge of the occurrence of $E$ has reduced the probability of occurrence of $G$.

From the above example we see that the knowledge of occurrence of an event may or may not affect the probability of occurrence of another event, and in the case where it afftects it may either increase or decrease the probability of occurrence of the second event. This leads us to the following definitions:

**Definition 3.6.1** If $E$ and $F$ are two events then the **conditional probability of $F$ given $E$** is denoted by $P(F|E)$ and is defined as

$$P(F|E) = \frac{P(F \cap E)}{P(E)} \tag{3.6.1}$$

In the case that conditional probability of $F$ given $E$ is the same as the probability of $F$ it means that the knowledge of occurrence of $E$ does not affect the probability of occurrence of $F$. We then say $F$ is independent of $E$. From above we have

$$\begin{aligned} P(F|E) &= P(F) \\ \Longrightarrow & \\ \frac{P(F \cap E)}{P(E)} &= P(F) \\ \Longrightarrow & \\ P(F \cap E) &= P(E)P(F) \end{aligned}$$

Note that the above is unaffected if we interchange $E$ and $F$. Thus we have

**Definition 3.6.2** Two events $E$ and $F$ are said to be **Independent** if

$$P(E \cap F) = P(E)P(F) \tag{3.6.2}$$

**Remark 3.6.1** Note that the notion of independence of events arises from conditional probability and the definition of conditional probability is dependent on the probability measure on the random experiment.

**Remark 3.6.2** It is easy to see that if $E, F$ is an independent pair of events then the following pairs are also independent:

$$\{E, F'\}\,,\ \{E', F'\}\,,\ \{E', F\}$$

**Remark 3.6.3** Note that the notion of independence of events arises from conditional probability and the definition of conditional probability is dependent on the probability measure on the random experiment. Hence the notion of independence of events is dependent on the probability measure on the probability space. Consequently two events may be independent with respect to one probability measure and not independent with respect to another probability measure.

**Independence of a Collection of Events**:
Let $\mathcal{C} = \{E_\alpha\}_{\alpha\in\mathcal{I}}$ be a collection of events, (where the index set $\mathcal{I}$ can be finite or an infinite sequence or any continuum also). Then we say that the collection $\mathcal{C}$ is independent if for **every** finite subcollection $\{E_1, E_2, \cdots, E_n\}$ in $\mathcal{C}$, the following holds:

$$\mathcal{P}\left(\bigcap_{j=1}^{n} E_j\right) = \mathcal{P}(E_1)\mathcal{P}(E_2)\cdots\mathcal{P}(E_n) \tag{3.6.3}$$

**Remark 3.6.4** A collection $\mathcal{C}$ of events may be such that every pair in the collection may be independent but the collection may not be independent

**Bayes' Rule**
Let us next consider a finite number of nonempty events $\Pi_1 = \{E_1, E_2, \cdots, E_n\}$ in the probability space $(\Omega, \mathcal{B}, \mathcal{P})$, such that they give a partition of $\Omega$, that is

$$E_i \cap E_j = \phi \text{ for } i \neq j \text{ and} \tag{3.6.4}$$

$$\Omega = \bigcup_{j=1}^{n} E_j \tag{3.6.5}$$

Then for any set $A \in \mathcal{B}$ we have

$$\begin{aligned}
A &= A \cap \Omega \\
&= A \cap \left( \bigcup_{j=1}^{n} E_j \right) \\
&= \bigcup_{j=1}^{n} (A \cap E_j) \\
\Longrightarrow & \\
\mathcal{P}(A) &= \mathcal{P}\left( \bigcup_{j=1}^{n} (A \cap E_j) \right) \\
&= \sum_{j=1}^{n} \mathcal{P}\left(A \cap E_j\right) \text{ since } E_j \text{ are all disjoint} \\
&= \sum_{j=1}^{n} \mathcal{P}(A|E_j)\mathcal{P}(E_j)
\end{aligned}$$

Thus we have

$$\mathcal{P}(A) = \sum_{j=1}^{n} \mathcal{P}(A|E_j)\mathcal{P}(E_j) \tag{3.6.6}$$

Now let $\Pi_1 = \{E_1, E_2, \cdots, E_n\}$ and $\Pi_2 = \{F_1, F_2, \cdots, F_m\}$ be any two partitions of $\Omega$ where $E_j$ and $F_k$ are all in $\mathcal{B}$.

$$\begin{aligned}
\mathcal{P}(E_r|F_k) &= \frac{\mathcal{P}(E_r \cap F_k)}{\mathcal{P}(F_k)} \\
&= \frac{\mathcal{P}(F_k|E_r)\mathcal{P}(E_r)}{\mathcal{P}(F_k)} \\
&= \frac{\mathcal{P}(F_k|E_r)\mathcal{P}(E_r)}{\sum_{j=1}^{n} \mathcal{P}\left(F_k|E_j\right)\mathcal{P}(E_j)} \\
& \text{(by applying (3.2.6) with } A = F_k\text{)}
\end{aligned}$$

This is called Bayes' Rule. We have
**Bayes' Rule**:
Let $\Pi_1 = \{E_1, E_2, \cdots, E_n\}$ and $\Pi_2 = \{F_1, F_2, \cdots, F_m\}$ be any two partitions

of $\Omega$ where $E_j$ and $F_k$ are all in $\mathcal{B}$. Then

$$\mathcal{P}(E_r|F_k) \quad = \quad \frac{\mathcal{P}(F_k|E_r)\mathcal{P}(E_r)}{\sum_{j=1}^{n} \mathcal{P}\left(F_k|E_j\right) \mathcal{P}(E_j)} \text{ for } 1 \leq r \leq n \tag{3.6.7}$$

## 3.7 Joint Distribution

Consider two random variables $X$ and $Y$ on a probability space $(\Omega, \mathcal{B}, \mathcal{P})$. We now define a vector valued random variable $Z : \Omega \longrightarrow \mathbb{R}^2$ as

$$Z(\omega) \quad = \quad (X(\omega), Y(\omega)) \tag{3.7.1}$$

For any $x, y \in \mathbb{R}$ let $I_x = (-\infty, x]$ and $I_y = (-\infty, y]$ and consider, in $\mathbb{R}^2$ the rectangle

$$R_{xy} = I_x \times I_y = \left\{(\xi, \eta) \in \mathbb{R}^2 : -\infty < \xi \leq x \text{ and } -\infty < \eta \leq y\right\}$$

We now collect all those $\omega \in \Omega$ for which $Z(\omega)$ is in this rectangle $R_{xy}$, that is

$$\begin{aligned} Z^{-1}(R_{xy}) &= \{\omega \in \Omega : Z(\omega) \in R_{xy}\} \\ &= \{\omega \in \Omega : -\infty < X(\omega) \leq x \text{ and } -\infty < Y(\omega) \leq y\} \\ &= X^{-1}(I_x) \bigcap Y^{-1}(I_y) \end{aligned}$$

Since $X$ and $Y$ are random variables the sets $X^{-1}(I_x)$ and $Y^{-1}(I_y)$ are in $\mathcal{B}$ and since $\mathcal{B}$ is a $\sigma$-algebra their intersection is also in $\mathcal{B}$. Thus $Z^{-1}(R_{xy}) \in \mathcal{B}$. Hence $\mathcal{P}\left(Z^{-1}(R_{xy})\right)$ is defined. This probability is a function of $x$ and $y$. We call this function of $x$ and $y$ as the **Joint CDF** of $X$ and $Y$ and denote it by $F_{XY}$. Thus we have

$$F_{XY}(x, y) \quad = \quad \mathcal{P}\left(Z^{-1}(R_{xy})\right) \tag{3.7.2}$$

We can write this as

$$= \quad \mathcal{P}\left(\{\omega \in \Omega : -\infty < X(\omega) \leq x \text{ and } -\infty < Y(\omega) \leq y\}\right) \tag{3.7.3}$$

We define

$$\lim_{y \to \infty} F_{XY}(x, y) \tag{3.7.4}$$

as the **Marginal Distribution of** $X$ (which is equal to $F_X(x)$ the CDF of $X$), and similarly

$$\lim_{x \to \infty} F_{XY}(x, y) \tag{3.7.5}$$

as the **Marginal Distribution of** $Y$, (and this is the same as the CDF $F_Y(y)$ of $Y$).

Consider two discrete random variables $X$ and $Y$, on a probability space $(\Omega, \mathcal{B}, \mathcal{P})$ such that

$$\begin{aligned} \mathcal{R}_X &= \{x_1, x_2, \cdots, x_M\} & (3.7.6) \\ \mathcal{R}_Y &= \{y_1, y_2, \cdots, y_N\} & (3.7.7) \end{aligned}$$

(that is, $X$ takes the values $x_1, x_2, \cdots, x_N$ and $Y$ takes the values $y_1, y_2, \cdots, y_M$). Let $p_X$ and $p_Y$ be the PMFs of $X$ and $Y$ respectively. As above we now define the vector valued random variable $Z : \Omega \longrightarrow \mathbb{R}^2$ as

$$Z(\omega) = (X(\omega), Y(\omega)) \tag{3.7.8}$$

The values taken by $Z$ is given by the set,

$$\mathcal{R}_Z = \{(x_i, y_j\}_{1 \le i \le M, 1 \le j \le N}$$

We define the **Joint PMF**, $p_{XY}$ as

$$p_{XY}(x_i, y_j) = \mathcal{P}\left(\{\omega \in \Omega : X(\omega) = x_i \text{ and } Y(\omega) = y_j\}\right) \tag{3.7.9}$$

The Marginal PMFs are given by

$$\begin{aligned} p_X(x_i) &= \sum_{j=1}^{N} p_{XY}(x_i, y_j) & (3.7.10) \\ p_Y(y_j) &= \sum_{i=1}^{M} p_{XY}(x_i, y_j) & (3.7.11) \end{aligned}$$

We can represent $p_{XY}(x_i, y_j)$ as an $M \times N$ matrix $(p_{ij})$ where

$$p_{ij} = p_{XY}(x_i, y_j) \tag{3.7.12}$$

The sum of the entries in the $i$th row of this matrix is the Probability $p_X(x_i)$ and the sum of the entries in the $j$th column gives the probability $p_Y(y_j)$.

If $X$ and $Y$ are continuous random variables and $F_X(x, y)$ is their Joint CDFs then their joint PDF, $f_{XY}(x, y)$ is defined as

$$f_{XY} = \frac{\partial^2}{\partial x \partial y} F_{XY}(x, y) \tag{3.7.13}$$

We then have

$$\begin{aligned} F_{XY}(x, y) &= \int_{-\infty}^{y} \int_{-\infty}^{x} f_{XY}(x, y) dx dy & (3.7.14) \\ &= \int_{-\infty}^{x} \int_{-\infty}^{y} f_{XY}(x, y) dy dx & (3.7.15) \end{aligned}$$

Let $a < b$ and $c < d$. For any rectangle $I \times J$ where $I$ is an interval with left and right end points as $a$ and $b$ respectively, and $J$ is an interval with left and right end points as $c$ and $d$ respectively, we have,

$$\begin{aligned} \mathcal{P}\left(\{\omega : X(\omega) \in I \text{ and } Y(\omega) \in J\}\right) &= \int_{I \times J} f_{XY}(x, y) dx dy \\ &= \int_a^b \left( \int_c^d f_{XY}(x, y) dy \right) dx \\ &= \int_c^d \left( \int_a^b f_{XY}(x, y) dx \right) dy \end{aligned}$$

The marginal pdfs are given by

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{XY}(x, y) dy & (3.7.16) \\ f_Y(y) &= \int_{-\infty}^{\infty} f_{XY}(x, y) dx & (3.7.17) \end{aligned}$$

**Example 3.7.1** Let us consider the experiment of rolling a fair die. Let $X$ and $Y$ be the random variables defined as follows:

$$\begin{aligned} X(\omega) &= \begin{cases} 1 & \text{if } \omega \textit{ is even} \\ -1 & \text{if } \omega \textit{ is odd} \end{cases} \\ Y(\omega) &= \begin{cases} 1 & \text{if } \omega \textit{ is prime} \\ -1 & \text{if } \omega \textit{ is not prime} \end{cases} \end{aligned}$$

The pmf of $X$ and $Y$ are as follows:

| $\omega$ | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $X(\omega)$ | $-1$ | 1 | $-1$ | 1 | $-1$ | 1 |
| $Y(\omega)$ | $-1$ | 1 | 1 | $-1$ | 1 | $-1$ |

Let $Z = (X, Y)$. Since $X$ and $Y$ take values $-1, 1$ the possible values of $Z$ are $(-1, -1).(-1, 1), (1, -1), (1, 1)$. The joint pmt is given below

| $Y \rightarrow$ <br> $X \downarrow$ | $-1$ | $1$ |
|---|---|---|
| $-1$ | $\frac{1}{6}$ | $\frac{1}{3}$ |
| $1$ | $\frac{1}{3}$ | $\frac{1}{6}$ |

Note that the first row sum is 0.5 which is $p_1 = p(X = -1)$ and the second row sum is 0.5 which is $p_2 = p(X = 1$. Thus the marginal pmt, namely the row sums give the pmt of $X$. Similarly the column sums give the pmt of $Y$, that is $q_1 = P(Y = -1) = 0.5$ and $q_2 = P(Y = 1) = 0.5$. Notice that the $(1, 1)$th entry, namely $\frac{1}{6}$ is not equal to the product $p_1 q_1$.

**Example 3.7.2** Consider two discrete random variables $X$ and $Y$ whose joint pmt is given below:

| $Y \rightarrow$ <br> $X \downarrow$ | $y_1$ | $y_2$ | *Row Sum* |
|---|---|---|---|
| $x_1$ | 0.12 | 0.28 | 0.40 |
| $x_2$ | 0.18 | 0.42 | 0.6 |
| *Column Sum* | 0.3 | 0.7 | |

Note that every entry is equal to the product of its corresponding row sum and column sum. Hence the two random variables are independent.

**Example 3.7.3** Let us again consider the random experiment of rolling a fair die. Consider the random variables $X$ and $Y$ defined as follows:

$$
\begin{aligned}
X(\omega) &= \begin{cases} -1 & \text{if} \quad \omega \text{ is odd} \\ \phantom{-}1 & \text{if} \quad \omega \text{ is even} \end{cases} \\
Y(\omega) &= \begin{cases} -2 & \text{if} \quad \omega \leq 2 \\ \phantom{-}2 & \text{if} \quad \omega > 2 \end{cases}
\end{aligned}
$$

The pmfs of $X$ and $Y$ are as given below:

| $X(\omega)$ | $x_1 = -1$ | $x_2 = 1$ |
|---|---|---|
| $p(x_i)$ | $\frac{1}{2}$ | $\frac{1}{2}$ |

| $Y(\omega)$ | $y_1 = -2$ | $x_2 = 2$ |
|---|---|---|
| $p(x_i)$ | $\frac{1}{3}$ | $\frac{2}{3}$ |

The joint pmt is as given below:

| $Y \rightarrow$ <br> $X \downarrow$ | $y_1 = -2$ | $y_2 = 2$ | *Row Sum* |
|---|---|---|---|
| $x_1 = -1$ | $\frac{1}{6}$ | $\frac{1}{3}$ | $\frac{1}{2}$ |
| $x_2 = 1$ | $\frac{1}{6}$ | $\frac{1}{3}$ | $\frac{1}{2}$ |
| *Column Sum* | $\frac{1}{3}$ | $\frac{2}{3}$ | |

## 3.8 Independence of Random variables

Consider two random variables $X$ and $Y$ on a probability space $(\Omega, \mathcal{E}, \mathcal{P})$. For any Borel set $B$ in $\mathbb{R}$ we must have $X^{-1}(B)$ and $Y^{-1}(B)$ in $\mathcal{E}$. If these two events are independent for every Borel set $B$ then we say that the two random variables are independent. In particular, if the generic events $X^{-1}(I_x)$ and $Y^{-1}(y)$ are independent for every $x$ and $y$ in $\mathbb{R}$ then the two random variables are independent. Hence we get in such a case

$$\begin{aligned} \mathcal{P}\left(X^{-}(I_x)\right) \bigcap Y^{-1}(I_y)) &= \mathcal{P}(X^{-}(I_x)) \times \mathcal{P}(Y^{-1}(I_y)) \\ &\Longrightarrow \\ F_{XY}(x,y) &= F_X(x)F_X(y) \qquad (3.8.1) \end{aligned}$$

Thus we have the definition of independence of random variables as follows:

**Definition 3.8.1** Two random variables $X$ and $Y$ are said to be independent if

$$F_{XY}(x,y) = F_X(x)F_Y(y) \text{ for all } x, y \in \mathbb{R} \qquad (3.8.2)$$

In particular, for discrete random variables we have the following:
Two discrete random variables $X$ and $Y$ taking values $x_1, x_2, \cdots$ and $y_1, y_2, \cdots$ respectively are independent if

$$p_{XY}(x_i, y_j) = p_X(x_i)p_Y(y_j) \qquad (3.8.3)$$

We observe that in this case the matrix $(p_{ij})$, (where $p_{ij} = p_{XY}(x_i, y_j)$) is such that

$$p_{ij} = p_i q_j \text{ where } p_i = p_X(x_i) \text{ and } q_j = p_Y(y_j) \tag{3.8.4}$$

The $(i, j)$-th entry $p_{ij}$ is, therefore, the product of the sum of the entries in the $i$-th row and the sum of the entries in the $j$-th column.

**Example 3.8.1** In Examples 3.7.2 and 3.7.3 we see that every entry in the joint pmt matrix is the product of the corresponding row and column sums. Hence in both these examples two random variables are independent. However in Example 3.7.1 this is not true since in this example each row sum and each column sum is $\frac{1}{2}$ and hence if the two random variables are to be independent each entry should have been their product, namely, $\frac{1}{4}$, which is not the case. Thus the the two random variables of Example 3.7.1 are NOT independent.

## 3.9 Expectation and Variance of a Random Variable

We next study two important parameters associated with a random variable. These measure the mean value of the random variable and its deviation from the mean value.

Discrete Random Variable:

Consider a discrete random variable $X$ on a probability space $(\Omega, \mathcal{B}, P)$, taking values $x_1, x_2, \cdots, x_N$ with probabilities $p_1, p_2, \cdots, p_N$. the "Expectation" of $X$ is denoted by $E(X)$ and is defined as the weighted average,

$$E(X) = \sum_{j=1}^{N} x_j p_j \tag{3.9.1}$$

Continuous Random variable

Consider a continuous random variable $X$ on a probability space $(\Omega, \mathcal{B}, P)$, with PDF $f_X(x)$. We define

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx \tag{3.9.2}$$

Let $\mathcal{R}_X$ be the Range of $X$, that is, $\mathcal{R}_X$ is the set of values taken by $X$. If $g : \mathcal{R}_X \longrightarrow \mathbb{R}$ is a "reasonably smooth" real valued function defined on the Range $\mathcal{R}_X$ of $X$ and $Y = g(X)$ then we define

$$E(Y) = E(g(X)) \quad = \quad \sum_{j=1}^{N} g(x_j)p_j \tag{3.9.3}$$

in the case of a discrete random variable as above, and

$$E(Y) = E(g(X)) \quad = \quad \int_{-\infty}^{\infty} g(x)f_X(x)dx \tag{3.9.4}$$

in the case of a continuous random variable with pdf $f_X(x)$.
It is to easy that the Expectation satisfies the following properties:

1. $E(\alpha X) = \alpha E(X)$ for any $\alpha \in \mathbb{R}$

2. If $X$ is a constant random variable $X = C$ then $E(X) = E(C) = C$

3. If $X$ and $Y$ are any two random variables then $E(X + Y) = E(X) + E(Y)$

4. Combining the first and third properties above we get that $E$ is a linear function, that is, if $X_1, X_2, \cdots, X_n$ are a finite number of real valued random variables on a probability space $(\Omega, \mathcal{B}, P)$ and $\alpha_1, \alpha_2, \cdots, \alpha_n \in \mathbb{R}$ then

$$E(\alpha_1 X_1 + \alpha_2 X_2 + \cdots + \alpha_n X_n) = \alpha_1 E(X_1) + \alpha_2 E(X_2) + \cdots + \alpha_n E(X_n)$$

**Example 3.9.1** Consider the sample space $\Omega = \{H, T\}$ of the experiment of tossing a fair coin. Let $X$ and $Y$ be the random variables defined as follows:

| $\omega$ | $H$ | $T$ |
|---|---|---|
| $Random\ Variable\ X$ | 1 | $-1$ |
| $Random\ Variable\ Y$ | $-1$ | 1 |

Then we have

$$p_X(1) = 0.5 = p_Y(1) \text{ and } p_X(-1) = 0.5 = p_Y(-1)$$

Hence we have

$$E(X) = (1)(0.5) + (-1)(0.5) = 0$$
$$E(Y) = (1)(0.5) + (-1)(0.5) = 0$$

We also have $X + Y$ is the random variable which takes the value 0 for both outcomes and hence $X + Y$ is the Zero random variable and we have $E(0) = 0$. Thus we have

$$E(X + Y) \;=\; E(X) + E(Y)$$

**Example 3.9.2** Let us pick a card at random from a pack of cards. Let the random variables $X$ and $Y$ be defined as follows:

$$X = \begin{cases} 1 & \text{if the picked card is RED} \\ -1 & \text{if the picked card is BLACK} \end{cases}$$

$$Y = \begin{cases} 1 & \text{if the picked card is HEARTS} \\ -1 & \text{if the picked card is DIAMOND} \\ 0 & \text{if the picked card is BLACK} \end{cases}$$

Then we have

$$\begin{aligned} E(X) &= 1 \times (0.5) + (-1) \times (0.5) = 0 \\ E(Y) &= 1 \times (0.25) + (-1) \times (0.25) = 0 \end{aligned}$$

Now consider the random variable $Z = XY$. Then $Z$ takes the values 1 and $-1$ with probabilities 0.25 each and the value 0 with probability 0.5 and hence we get $E(XY) = 0$

**Example 3.9.3** Let $X$ be a random variable uniformly distributed over the interval $[a, b]$. Then its pdf is given by

$$f_X(x) \;=\; \frac{1}{b-a} \tag{3.9.5}$$

Hence we get

$$E(X) \;=\; \int_a^b x \times \frac{1}{b-a} dx$$

$$
\begin{aligned}
&= \frac{[x^2]_{x=a}^{x=b}}{2(b-a)} \\
&= \frac{b^2-a^2}{2(b-a)} \\
&= \frac{b+a}{2}
\end{aligned}
$$

We next introduce the notion of Variance of a random variable. The variance measures the average of the square of the deviation of $X$ from its mean. We have

**Definition 3.9.1** Let $X$ be a random variable with expectation $E(X) = \mu$. We then define the variation of $X$ as the expectation of the random variable $(X-\mu)^2$

$$
Var(X) = E((X-\mu)^2) \tag{3.9.6}
$$

We have

$$
\begin{aligned}
Var(X) &= E((X-\mu)^2) \\
&= E(X^2 - 2\mu X + \mu^2) \\
&= E(X^2) - 2\mu E(X) + \mu^2 \\
&\quad \text{(using the properties of expectation observed above)} \\
&= E(X^2) - \mu^2 \\
&= E(X^2) - (E(X))^2
\end{aligned}
$$

Thus we have

$$
Var(X) = E(X^2) - (E(X)^2) \tag{3.9.7}
$$

We define "**standard deviation**" (denoted by $\sigma$) of the random variable as

$$
\sigma = \sqrt{Var(X)} \tag{3.9.8}
$$

**Example 3.9.4** For the random variable $X$ and $Y$ of Example 3.9.1 we have both $X^2$ and $Y^2$ are the constant random variables $X^2 = 1$ and $Y^2 = 1$ and both have expectation 0. Hence we have

$$
Var(X) = E(X^2) - E(X)
$$

$$\begin{aligned} &= 1 - 0 = 1 \\ Var(Y) &= E(Y^2) - E(Y) \\ &= 1 - 0 = 1 \end{aligned}$$

Hence $\sigma = 1$

**Example 3.9.5** For the random variables $X$ and $Y$ of Example 3.9.2 we have $X^2$ is the constant random variable $X^2 = 1$ with expectation 0 and hence we have $Var(X) = E(X^2) - E(X) = 1 - 0 = 1$ and hence $sigma = 1$. For the random variable $Y$ of Example 3.9.2 we have $Y^2$ takes the values 1 and 0 each with probability 0.5. Hence we have

$$\begin{aligned} Var(Y) &= E(Y^2) - (E(Y))^2 \\ &= 0.5 - 0 = 0.5 \end{aligned}$$

Hence $\sigma = \frac{1}{\sqrt{2}}$

**Example 3.9.6** For the uniform random variable of Example 3.9.3 we have

$$\begin{aligned} Var(X) &= E(X^2) - (E(X))^2 \\ &= \int_a^b x^2 \frac{1}{b-a} dx - \frac{(b+a)^2}{4} \\ &= \frac{b^3 - a^3}{3(b-a)} - \frac{(b+a)^2}{4} \\ &= \frac{b^2 + a^2 - 2ab}{12} \\ &= \frac{(b-a)^2}{12} \end{aligned}$$

Hence we get

$$\sigma = \frac{b-a}{2\sqrt{3}}$$

We observe the following properties of the variance:

1. We have, for any real number $\alpha$,

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

   Thus we have

$$Var(\alpha X) = \alpha^2 Var(X) \text{ for every } \alpha \in \mathbb{R} \tag{3.9.9}$$

2. Let $X$ and $Y$ be two random variables. We have

$$\begin{aligned} Var(X+Y) &= E((X+Y)^2) - ()^2 E(X+Y) \\ &= \left\{E(X^2 + Y^2 + 2XY)\right\} \\ &\quad - \{E(X) + E(Y)\}^2 \\ &= \left\{E(X^2) + E(Y^2) + 2E(XY)\right\} \\ &\quad - \left\{(E(X))^2 + (E(Y))^2 + 2E(X)E(Y)\right\} \\ &= \left\{E(X^2 - (E(X))^2\right\} \\ &\quad + \left\{E(Y^2) - (E(Y))^2\right\} + 2\{E(XY) - E(X)E(Y)\} \\ &= Var(X) + Var(Y) + 2Cov(X,Y) \end{aligned}$$

   where

$$Cov(X,Y) = E(XY) - E(X)E(Y) \tag{3.9.10}$$

   is called the Covariance of $X$ and $Y$. Thus we have

$$Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y) \tag{3.9.11}$$

   Suppose $X$ and $Y$ are continuous random variables with pdfs $f_X(x)$ and $f_Y(y)$ and joint pdf as $f_{XY}(x,y)$. Then we have

$$E(XY) = \int_{-\infty}^{\infty} xy f_{XY}(x,y)dxdy$$

If $X$ and $Y$ are independent then we have $f_{XY}(x,y) = f_X(x)f_Y(y)$ and hence we get

$$\begin{aligned}
E(XY) &= \int_{-\infty}^{\infty} xy f_{XY}(x,y)dxdy \\
&= \int_{-\infty}^{\infty} xy f_X(x)f_Y(y)dxdy \\
&= \left\{\int_{-\infty}^{\infty} x f_X(x)dx\right\} \times \left\{\int_{-\infty}^{\infty} y f_Y(y)dx\right\} \\
&= E(X)E(Y)
\end{aligned}$$

(We can prove this analogously in the case of discrete random variables). Thus we have

$$X,Y \text{ independent random variables} \implies E(XY) = E(X)E(Y) \tag{3.9.12}$$

and hence

$$X,Y \text{ independent random variables} \implies Cov(X,Y) = 0 \tag{3.9.13}$$

Using this in (3.9.11) we get

$$X,Y \text{ independent random variables} \implies V + Y) = Var(X) + Var(Y) \tag{3.9.14}$$

**Remark 3.9.1** It must be noted that the converse of (3.9.13) is not true in general, that is we can have random variables $X, Y$ which are not independent for which we can have $Cov(X,Y) = 0$. For instance, for the random variables $X$ and $Y$ of Example 3.9.2 we hade $E(X) = E(Y) = E(XY) = 0$ and hence $Cov(X,Y) = 0$, but it can be easily verified that these random variables are not independent.

## 3.10 Tail Distribution

Let $X$ be a nonnegative random variable on a probability space $(\Omega, \mathcal{B}, P)$. The probability that $X$ takes values beyond a certain threshold value is what is known as the Tail Distribution. More precisely we define $F_X^T(x)$, the tail distribution of $X$ as

$$F_X^T(x) = P(X \geq x) \tag{3.10.1}$$

If $X$ is real valued random variable with expectation $\mu_X$, then we are interested in the probability that the $X$ does not deviate from its mean beyond a certain threshold value and hence we are interesting in the tail distribution of $|X - \mu|$, that is, we are interested in $P(|X - \mu| \geq x)$. We now look at some inequalities that give certain estimates for this tail distribution.

**I Markov's Inequality**

Suppose we have a class of 90 students whose average score in a test is 20. Let us say we are interested in the probability that a randomly chosen student"s score is 60 or more. We have,

$$\begin{aligned}
\text{the number of students who score 60 or more} &= k \\
&\Longrightarrow \\
\text{The total score of these } k \text{ students} &\geq 60k \\
&\Longrightarrow \\
\text{The total score of the class} &\geq 60k \\
&\Longrightarrow \\
\text{The average score must be} &\geq \frac{60k}{90} \\
&\Longrightarrow \\
20 &\geq \frac{60k}{90} \\
&\Longrightarrow \\
\frac{k}{90} &\leq \frac{20}{60} \\
P(Score \geq 60) &\leq \frac{20}{60}
\end{aligned}$$

(since $\frac{k}{90}$ is the proportion of students getting a score of at least 60)
If we now replace the scores by a general nonnegative random variable $X$, the average by its expectation $E(X)$, and the threshold score 60 by a general positive real number $k$, we should get

$$P(X \geq k) \leq \frac{E(X)}{k}$$

That this is true is what Markov's inequality establishes. We have

**Theorem 3.10.1** **Markov's Inequality**

**If $X$ is any nonnegative random variable, on a probability space**

$\boldsymbol{(\Omega, \mathcal{B}, P)}$**, with expectation** $\boldsymbol{E(X)}$ **then**

$$\boldsymbol{P(X \geq k) \;\; \leq \;\; \frac{E(X)}{k}} \textbf{ for any } \boldsymbol{k > 0} \qquad (3.10.2)$$

Proof:
Case 1: Discrete random variable
Let $X$ be a nonnegative random variable taking values $x_1, x_2, \cdots, x_n$ with respective probabilities $p_1, p_2, \cdots, p_n$. Then we have for any $k > 0$,

$$\begin{aligned}
E(X) &= \sum_{j=1}^{n} p_j x_j \\
&= \sum_{\{j:x_j<k\}} p_j x_j + \sum_{\{j:x_j \geq k\}} p_j x_j \\
&\geq \sum_{\{j:x_j \geq k\}} p_j x_j \\
&\geq k \sum_{\{j:x_j<k\}} p_j \\
\Longrightarrow & \\
E(X) &\geq kP(X \geq k) \\
\Longrightarrow & \\
P(X \geq k) &\leq \frac{E(X)}{k}
\end{aligned}$$

thus proving the inequality. Analogously we prove in the case of continuous random variables as follows:
<u>Case 2</u>: Continuous random variable
Let $f(x)$ be the pdf of the random variable X. Then we have

$$\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} x f(x) dx \\
&= \int_{-\infty}^{k} x f(x) dx + \int_{k}^{\infty} x f(x) dx \\
&\geq \int_{k}^{\infty} x f(x) dx \\
&\geq k \int_{k}^{\infty} f(x) dx \\
&\geq kP(X \geq k) \\
\Longrightarrow & \\
P(X \geq k) &\leq \frac{E(X)}{k}
\end{aligned}$$

**Remark 3.10.1** The inequality will not be of any use at all if $E(X) = \infty$. Hence without loss of generality we can assume in the above statement that $E(X) < \infty$

**Remark 3.10.2** The inequality may give some times some bizarre results even if $E(X) < \infty$, as shown in the following example.

**Example 3.10.1** Consider the random experiment of throwing a fair die. We have, in this case,

$$\Omega = \{1, 2, 3, 4, 5, 6\} \text{ and } p(j) = \frac{1}{6} \text{ for } 1 \leq j \leq 6$$

Let $X$ be the random variable defined as $X(j) = j$.
We have

$$\begin{aligned} E(X) &= \frac{1+2+3+4+5+6}{6} \\ &= 3.5 \end{aligned}$$

Hence by Markov's inequality we get

$$P(X \geq 3) \leq \frac{E(X)}{3} = \frac{3.5}{3}$$

Since the rhs above is $> 1$, in this case, we get no information from Markov's inequality since we already know that the probability is anyway $\leq 1$.
We also get from Markov's inequality

$$P(X \geq 5) \leq \frac{E(X)}{5} = \frac{3.5}{5} = 0.7$$

The actual value is

$$P(X \geq 5) = p(5) + p(6) = \frac{2}{6} = \frac{1}{3} = 0.33$$

Thus we see that the estimate we get from Markov's inequality is a highly exaggerated overestimate. Thus we see from this example that the Markov's inequality may give sometimes highly exaggerated overestimates.

**Remark 3.10.3** Despite the above Remark, it must be noted that the inequality is of a very general nature for all nonnegative random variables and for all positive $k$. If we keep this in mind, this inequality is "tight", that is there will be at least one nonnegative random variable $X$ and one positive real number $k$ for which the inequality becomes an equality as shown in the following example. Hence we cannot make the lhs any smaller if the inequality has to hold fro all nonnegative random variables and all $k > 0$

**Example 3.10.2** Let $a > 1$ be any positive real number. Consider a random variable $X$ which takes only two values $a$ and 0 with respective probabilities $\frac{1}{a}$ and $\frac{a-1}{a}$. For this nonnegative random variable we get

$$E(x) = a \times \frac{1}{a} = 1$$

$$P(X \geq a) = P(X = a) = \frac{1}{a}$$

On the other hand we get from Markov's inequality,

$$P(X \geq a) \leq \frac{E(X)}{a} = \frac{1}{a}$$

Comparing with the exact value of this probability found above we see that equality holds in the Markov's inequality for this $X$ and for this $a$.

**<u>II Chebychev's Inequality</u>**:

Let $X$ be a real valued random variable with $E(X) = \mu < \infty$. Then $Y = |X - \mu|$ is a nonnegative random variable and we have for any positive real number $k$ and any increasing function $f(t) : [0, \infty) \longrightarrow \mathbb{R}$,

$$\begin{aligned}
Y(\omega) \geq k &\iff f(Y(\omega)) \geq f(k) \\
&\Longrightarrow \\
P(Y \geq k) &= P(f(Y) \geq f(k) \\
&\leq \frac{E(f(Y))}{f(k)} \text{ (by Markov's inequality)}
\end{aligned}$$

As a special case let us take $f(t) = t^2$. Then we have from above,

$$P(Y \geq k) \leq \frac{E(Y^2)}{k^2}$$

Recalling that $Y = |X - \mu|$ we get

$$P(|X - \mu| \geq k) \quad \leq \quad \frac{E(|X - \mu|^2)}{k^2}$$

But

$$E(|X - \mu|^2) \quad = \quad E((X - \mu)^2) = Var(X)$$

Substituting above we get

$$P(|X - \mu|) \quad \leq \quad \frac{Var(X)}{k^2}$$

Thus we have

**Theorem 3.10.2** **Chebychev's Inequality**

**If $X$ is any real valued random variable with expectation $E(X) = \mu < \infty$ and Variance $Var(X)$, then**

$$\boldsymbol{P(|X - \mu|) \quad \leq \quad \frac{Var(X)}{k^2}} \textbf{ for any real number } \boldsymbol{k > 0}$$

**III Chernoff Bound**

Let $X$ be any real valued random variable. Now consider the function

$$f : \mathbb{R} \longrightarrow \mathbb{R}$$

defined as $f(t) = e^{\alpha t}$, where $\alpha$ is a positive constant. Then $f(t)$ is a nonnegative increasing function. the random variable defined as $Y = f(X) = e^{\alpha X}$ is a nonnegative random variable. Further since $f(t)$ is increasing we see that for any real number $k$ we have

$$\begin{aligned} X \geq k \quad &\Longleftrightarrow \quad f(X) \geq f(k) \\ &\Longleftrightarrow \quad e^{\alpha X} \geq e^{\alpha k} \\ &\Longleftrightarrow \quad Y \geq f(k) \end{aligned}$$

Hence we get

$$P(X \geq a) \quad = \quad P(Y \geq e^{\alpha k}) \tag{3.10.3}$$

Since $Y$ is a nonnegative random variable we can apply Markov's inequality to $Y$ to get

$$P(Y \geq e^{\alpha k}) \leq \frac{E(Y)}{e^{\alpha k}} \tag{3.10.4}$$

From the above two equations we get

$$P(X \geq k) \leq \frac{E(Y)}{e^{\alpha k}} = e^{-\alpha k} E(e^{\alpha X}) \tag{3.10.5}$$

Similarly consider the decreasing function $g(t) = e^{-\alpha t}$ and let $Y$ be the nonnegative random variable defined as $Y = g(X) = e^{-\alpha X}$. Using the fact that $X \leq k$ if and only if $g(X) \geq g(k)$, we get

$$\begin{aligned} P(X \leq k) &= P(g(X) \geq g(k)) \\ &= P(Y \geq g(k)) \\ &\leq \frac{E(Y)}{g(k)} \end{aligned}$$

Substituting $Y = g(X) = e^{-\alpha X}$ and $g(k) = e^{-\alpha k}$ we get

$$P(X \leq k) \leq \frac{E(e^{-\alpha X})}{e^{-\alpha k}} = e^{\alpha k} E(e^{-\alpha X}) \tag{3.10.6}$$

Since (3.10.5) and (3.10.6) hold whatever $\alpha > 0$ we choose we get

$$P(X \geq k) \leq \min_{\alpha > 0} \left\{ e^{-\alpha k} E(e^{\alpha X}) \right\} \tag{3.10.7}$$

$$P(X \leq k) \leq \min_{\alpha > 0} \left\{ e^{\alpha k} E(e^{-\alpha X}) \right\} \tag{3.10.8}$$

Consider now a a finite number of real valued random variables $X_1, X_2, \cdots, X_n$ and let $X = X_1 + X_2 + \cdots + X_n$. Then we get by applying (3.10.7) to the random variable $X$ defined as $X = X_1 + X_2 + \cdots + X_n$ we get

$$\begin{aligned} P(X \geq k) &\leq \min_{\alpha > 0} \left\{ e^{-\alpha k} E(e^{\alpha X}) \right\} = \min_{\alpha > 0} \left\{ e^{-\alpha k} E(e^{\alpha(X_1 + X_2 + \cdots + X_n)}) \right\} \\ &\Longrightarrow \\ P(X \geq k) &\leq \min_{\alpha > 0} \left\{ e^{-\alpha k} E(\prod_{j=1}^{n} e^{\alpha X_j}) \right\} \end{aligned} \tag{3.10.9}$$

If the random variables $X_1, X_2, \cdots, X_n$ are independentt then the random variables $e^{\alpha X_1}, e^{\alpha X_2}, \cdots, e^{\alpha X_n}$ are independent and hence we get

$$E(\prod_{j=1}^{n} e^{\alpha X_j})) \;=\; \prod_{j=1}^{n} E(e^{\alpha X_j})$$

Substituting this in (3.10.9) we get

$$P(X \geq k) \;\leq\; \min_{\alpha>0} \left\{ e^{-\alpha k} \prod_{j=1}^{n} E(e^{\alpha X_j}) \right\}$$

Analogously we get from (3.10.8)

$$P(X \leq k) \;\leq\; \min_{\alpha>0} \left\{ e^{\alpha k} \prod_{j=1}^{n} E(e^{-\alpha X_j}) \right\}$$

Thus we have

**Theorem 3.10.3** **Chernoff Bounds**

**If $\boldsymbol{X_1, X_2, \cdots, X_n}$ are independent real valued random variables on a probability space $\boldsymbol{(\Omega, \mathcal{B}, P)}$, then**

$$\boldsymbol{P(X \geq k) \;\leq\; \min_{\alpha>0} \left\{ e^{-\alpha k} \prod_{j=1}^{n} E(e^{\alpha X_j}) \right\}} \qquad (3.10.10)$$

$$\boldsymbol{P(X \leq k) \;\leq\; \min_{\alpha>0} \left\{ e^{\alpha k} \prod_{j=1}^{n} E(e^{-\alpha X_j}) \right\}} \qquad (3.10.11)$$

## 3.11 Convergence of a Sequence of Random Variables

Let $\{X_n\}_{n\geq 1}$ and $X$ be real valued random variables on a probability space $(\Omega, \mathcal{B}, P)$. We shall now introduce three notions of convergence of the sequence $X_n$ to $X$.

**Almost Sure Convergence**

For every fixed $\omega \in \Omega$ the sequence $\{X_n(\omega)\}$ is a sequence of real numbers. We look at all those $\omega$ for which the sequence $X_n(\omega)$ converges to $X(\omega)$. Let

$$\mathcal{C} \;=\; \left\{ \omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega) \right\}$$

be the set of all points in $\Omega$ where $X_n$ converges to $X$. If this set has probability one, that is the set where we do not have convergence is of Probability 0, then we say that the sequence $X_n$ converges almost surely to $X$ and write $X_n \xrightarrow{a.s.} X$. We have

**Definition 3.11.1 A sequence of random variables $X_n$ on a probability space $(\Omega, \mathcal{B}, P)$ is said to converge almost surely to a random variable $X$ on this probability space if**

$$P(X_n(\omega) \longrightarrow X(\omega)) = 1$$

**or equivalently**

$$P(X_n(\omega) \not\longrightarrow X(\omega)) = 0$$

**We then write**

$$X_n \xrightarrow{a.s.} X$$

**Convergence in Probability**

In the notion of almost sure convergence, almost at every point $\omega$ we are able to control the error $|X_n(\omega) - X(\omega)|$ within any prescribed threshold beyond a certain stage, (the stage will depend on the threshold and may depend on the point $\omega$ also). Now we look at a notion of convergence where we are able to control the set where we may not be able to control the error. What we mean is the following:

First of all consider any threshold error $\epsilon > 0$. Then look at all those points in $\Omega$ where the error at the $n$th stage is $\geq$ this threshold, that is, we look at the set

$$\{\omega \in \Omega : |X_n(\omega) - X(\omega)| \geq \epsilon\}$$

Suppose this set becomes "smaller and smaller" as $n \to \infty$, that is, suppose

$$\lim_{n\to\infty} P\left(|X_n(\omega) - X(\omega)| \geq \epsilon\right) = 0$$

Then the set where the error exceeds the threshold $\epsilon$ has very low probability for large $n$. Suppose this happens for every error $\epsilon > 0$, then we say $X_n$ converges to $X$ in probability and write $X_n \xrightarrow{p} X$. We have

**Definition 3.11.2 A sequence of random variables $X_n$ on a probability space $(\Omega, \mathcal{B}, P)$ is said to converge in probability to a random variable $X$ on this probability space if**

$$\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0 \text{ for every } \epsilon > 0$$

**We then write**

$$\boldsymbol{X_n \xrightarrow{p} X}$$

**Convergence in Distribution**

Instead of looking at the values of the random variable converging or not converging, as we did in the above two definitions, we may look at their CDFs which contain the information about the distribution of the values of the random variables and see whether the CDF of $X_n$ converges to the CDF of $X$. Since the CDF of $X$ may have points of discontinuity we look at only those points where $F_X(x)$ is continuous. We have

**Definition 3.11.3 A sequence of random variables $\boldsymbol{X_n}$ on a probability space $\boldsymbol{(\Omega, \mathcal{B}, P)}$ is said to converge in distribution to a random variable $\boldsymbol{X}$ on this probability space if**

$$\boldsymbol{\lim_{n\to\infty} F_{X_n}(x) = F_X(x)} \textbf{ at every } \boldsymbol{x} \textbf{ where } \boldsymbol{F_X} \textbf{ is continuous}$$

**We then write**

$$\boldsymbol{X_n \xrightarrow{d} X}$$

**The Hierarchy**

We have the following hierarchy of these convergences

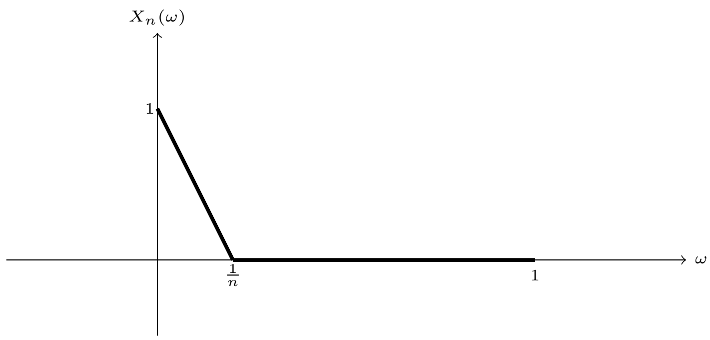
$$X_n \xrightarrow{a.s.} X \Longrightarrow X_n \xrightarrow{p} X \Longrightarrow X_n \xrightarrow{d} X$$

In general, the reverse inequalities are not true, that is

1. there are sequences of random variables which converge in probability (and hence also in distribution) but do not converge a.s. and

2. there are sequences of random variables which converge in distribution but do not converge in probability (and hence also not a.s.)

**Example 3.11.1** Let $\Omega = [0,1]$ $\mathcal{B}$, the Borel sets in $[0,1]$ and $P$ the probability measure arising from the length of an interval. Let $X_n(\omega)$ be the sequence of random variables, with the graph of $X_n(\omega)$ as shown below:




Let $\omega$ be such that $0 < \omega \leq 1$. Then we can find a positive integer $N$ such that $\dfrac{1}{N} < \omega$. Then

$$\begin{aligned} n \geq N \Longrightarrow \frac{1}{n} \leq \frac{1}{N} < \omega \\ \Longrightarrow X_n(\omega) = 0 \text{ for all } n \geq N \\ \Longrightarrow X_n(\omega) \longrightarrow 0 \end{aligned}$$

Thus for every $\omega \neq 0$ the sequence $X_n(\omega)$ converges to 0. Let $X$ be the zero random variable. Further $X_n(0) = 1$ for all $n$ and hence $X_n(0) \longrightarrow 1$. We therefore have

$$\begin{aligned} \left\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\} &= (0,1] \\ \Longrightarrow & \\ P(X_n \longrightarrow X(\omega)) &= 1 \end{aligned}$$

Hence we have

$$X_n \xrightarrow{a.s.} X$$

From the hierarchy of the types of convergence we get that the sequence also converges in probability and in distribution. We can also check convergence in probability directly as follows:
For $\epsilon > 1$ we have

$$\{\omega \in \Omega : |X_n(\omega) - X(\omega)| \geq \epsilon\} = \phi$$

since $X_n(\omega) \leq 1$ for all $\omega$ and $X(\omega) = 0$. Hence

$$\begin{aligned} P(|X_n(\omega) - X(\omega)| \geq \epsilon) &= P(\phi) = 0 \\ \Longrightarrow & \\ \lim_{n\to\infty} P(|X_n(\omega) - X(\omega)| \geq \epsilon) &= 0 \end{aligned}$$

Let $0 < \epsilon \leq 1$. We have, from the definition of $X_n$,

$$X_n(\omega) = \begin{cases} 0 & \text{for} \quad \frac{1}{n} \leq \omega \leq 1 \\ 1 - n\omega & \text{for} \quad 0 \leq \omega \leq \frac{1}{n} \end{cases}$$

Hence we get

$$\begin{aligned} \{\omega \in \Omega : |X_n(\omega) - X(\omega)| \geq \epsilon\} &= \left[0, \frac{1-\epsilon}{n}\right] \\ \Longrightarrow & \\ P(|X_n(\omega) - X(\omega)| \geq \epsilon) &= \frac{1-\epsilon}{n} \\ \Longrightarrow & \\ \lim_{n\to\infty} P(|X_n(\omega) - X(\omega)| \geq \epsilon) &= \lim_{n\to\infty} \frac{1-\epsilon}{n} = 0 \end{aligned}$$

Thus we have

$$\lim_{n\to\infty} P(|X_n(\omega) - X(\omega)| \geq \epsilon) = \lim_{n\to\infty} \frac{1-\epsilon}{n} = 0 \text{ for all } \epsilon > 0$$

Hence

$$X_n \xrightarrow{p} X$$

We can also check the convergence in distribution directly as follows: We have, since $X$ is the constant random variable $X = 0$,

$$F_X(x) = \begin{cases} 0 & \text{for} \quad x < 0 \\ 1 & \text{for} \quad x \geq 0 \end{cases}$$

From the given definition of $X_n$ we have

$$F_{X_n}(x) = \begin{cases} 0 & \text{for} \quad x < 0 \\ 1 - \frac{1-x}{n} & \text{for} \quad 0 \leq x < 1 \\ 1 & \text{for} \quad x \geq 1 \end{cases}$$

From this it follows that

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x) \text{ at all } x$$

Hence we get

$$X_n \xrightarrow{d} X$$

**Example 3.11.2** Consider the random experiment of tossing a fair coin and consider the random variables $X$ and $Y$

| $\omega$ | $H$ | $T$ |
|---|---|---|
| $Random\ Variable\ X$ | $1$ | $-1$ |
| $Random\ Variable\ Y$ | $-1$ | $1$ |

We have

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

Let $X_n$ be the sequence defined as

$$X_n = X \text{ for all } n$$

Then we have clearly

$$F_{X_n}(x) = F_X(x) = F_Y(x)$$

Hence we have

$$\lim_{n\to\infty} F_{X_n}(x) = F_Y(x) \text{ for all } x$$

Hence

$$X_n \xrightarrow{d} Y$$

On the other hand we have

$$|X_n(\omega) - Y(\omega)| = 2 \text{ for all } \omega$$

and hence

$$\{\omega : X_n(\omega) \text{ converges to } Y(\omega)\} = \phi$$

Thus

$$P(X_n(\omega) \text{ converges to } Y(\omega)) = 0 \neq 1$$

Thus $X_n$ does not converge almost surely to $Y$. Similarly we have, for $\epsilon = 1$,

$$\begin{aligned} P(|X_n - Y| \geq 1) &= P(2 \geq 1) \\ &= 1 \end{aligned}$$

Hence

$$\lim_{n\to\infty} P(|X_n - Y| \geq 1) = 1 \neq 0$$

Hence $X_n$ does not converge in probability to $X$. Thus we have convergence of $X_n$ to $Y$ in distribution but not convergence to $Y$ in probability or almost surely.

We can also introduce other types of convergence by demanding certain chosen feature(s) of the sequence to converge to the corresponding feature of the limit random variable. For example we have the following types of convergence:

**Definition 3.11.4** **Let $X_n$ and $X$ be random variables defined on a probability space $(\Omega, \mathcal{B}, P)$. We say**

1. $X_n \xrightarrow{L^1} X$ **if** $E(|X_n - X|) \longrightarrow 0$
2. $X_n \xrightarrow{L^2} X$ **if** $E(|X_n - X|^2) \longrightarrow 0$
3. **In general for** $1 \leq p < \infty$ **we say,** $X_n \xrightarrow{L^p} X$ **if** $E(|X_n - X|^p) \longrightarrow 0$

The hierarchy of convergence is as follows:

$$\begin{array}{ccc}
 & & \textbf{a.s convergence} \\
 & & \Downarrow \\
L^1 \textbf{ convergence} & \Rightarrow & \textbf{convergence in probability} \\
\Uparrow & & \Downarrow \\
L^2 \textbf{ convergence} & & \textbf{convergence in distribution} \\
\Uparrow & & \\
L^p \textbf{ convergence } (p > 2) & & \\
\Uparrow & & \\
L^q \textbf{ convergence } (q > p) & &
\end{array}$$

In general, the reverse implications in each case is not valid.