Chapter 2

Probability Space

2.1 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 set of all possible outcomes is called the **Sample Space** for the experiment, and this set will be denoted by Ω.

Example 2.1.1 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

$$\Omega = \{H, T\} \tag{2.1.1}$$

Example 2.1.2 Suppose we toss a fair coin thrice and note the sequence of outcomes. Then for this experiment we have

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\} \tag{2.1.2}$$

Example 2.1.3 For the experiment of throwing a six faced die, with its faces numbered $1, 2, 3, 4, 5, 6$ we have

$$\Omega = \{1, 2, 3, 4, 5, 6\} \tag{2.1.3}$$

Example 2.1.4 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

$$\Omega = \{(x, y) : a \leq x \leq b, \ c \leq y \leq d\} \tag{2.1.4}$$

Thus the first ingredient in Probability Theory is the Sample Space of a random experiment.
The next important ingredient in Probability theory is the notion of **Events**. 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 Ω, 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,

$$A \in \mathcal{B} \implies A' \in \mathcal{B} \tag{2.1.5}$$

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

$$A, B \in \mathcal{B} \implies A \cup B \in \mathcal{B} \tag{2.1.6}$$

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

$$A_1, A_2, \cdots, A_N \in \mathcal{B} \implies \bigcup_{j=1}^{N} A_j \in \mathcal{B} \tag{2.1.7}$$

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

$$A, B \in \mathcal{B} \implies A \cap B \in \mathcal{B} \tag{2.1.8}$$

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

$$A_1, A_2, \cdots, A_N \in \mathcal{B} \implies \bigcap_{j=1}^{N} A_j \in \mathcal{B} \tag{2.1.9}$$

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

$$\lim_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n \tag{2.1.10}$$

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

$$\lim_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n \in \mathcal{B} \tag{2.1.11}$$

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

$$\lim_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} A_n \tag{2.1.12}$$

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

$$\lim_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} A_n \in \mathcal{B} \tag{2.1.13}$$

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

$$\mathcal{B} \text{ is a nonempty collection} \tag{2.1.14}$$

$$A \in \mathcal{B} \Longrightarrow A' \in \mathcal{B} \tag{2.1.15}$$

$$A, B \in \mathcal{B} \Longrightarrow A \cup B \in \mathcal{B} \tag{2.1.16}$$

$$\{A_n\}_{n=1,2,\cdots} \in \mathcal{B}, \text{ and } A_n \subseteq A_{n+1} \text{ for } n = 1, 2, \cdots \Longrightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{B} \tag{2.1.17}$$

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

$$A_1 = B_1 \tag{2.1.18}$$

$$A_n = \bigcup_{j=1}^{n} B_j \ , \ n = 2, 3, \cdots \tag{2.1.19}$$

Then clearly we have

$$A_n = \bigcup_{j=1}^{n} B_j \quad \in \quad \mathcal{B} \text{ by equation 2.1.16} \tag{2.1.20}$$

$$\bigcup_{j=1}^{n} B_j = \bigcup_{j=1}^{n} A_j \text{ for every } n \tag{2.1.21}$$

$$\bigcup_{j=1}^{\infty} B_j = \bigcup_{j=1}^{\infty} A_j \tag{2.1.22}$$

Clearly $\{A_n\}_{n=1,2,\cdots}$ is a non decreasing sequence in $\mathcal{B}$. Hence by equation 2.1.17 we get

$$\bigcup_{n=1}^{\infty} A_n \quad \in \quad \mathcal{B} \tag{2.1.23}$$

Hence by equation 2.1.22 we get

$$\bigcup_{n=1}^{\infty} B_n \quad \in \quad \mathcal{B} \tag{2.1.24}$$

Thus we have

$$B_n \in \mathcal{B} \implies \bigcup_{n=1}^{\infty} B_n \in \mathcal{B} \tag{2.1.25}$$

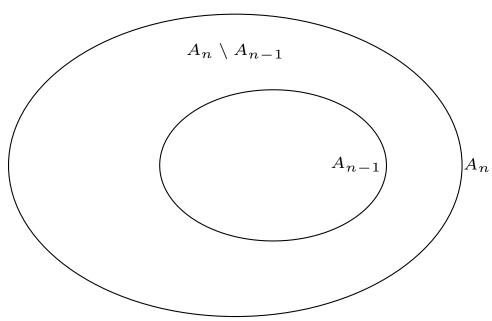
Whenever the above is true we say that $\mathcal{B}$ is closed under countable union. Thus we have

$$\begin{aligned} 2.1.14, 2.1.15, 2.1.16, 2.1.17 \implies & \text{(2.1.25), that is ,} \\ & \mathcal{B} \text{ is closed under countable union} \end{aligned} \tag{2.1.26}$$

Conversely suppose we have 2.1.14,2.1.15,2.1.25 Then clearly 2.1.16 is satisfied. We shall now see that 2.1.17 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

$$B_1 = A_1 \tag{2.1.27}$$

$$B_n = A_n \setminus A_{n-1} \text{ for } n = 2, 3 \cdots \tag{2.1.28}$$



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

Using 2.1.25 we see that $\displaystyle\bigcup_{j=1}^{\infty} B_j \in \mathcal{B}$ and hence $\displaystyle\bigcup_{j=1}^{\infty} A_j \in \mathcal{B}$. This means that $\mathcal{B}$ is closed under monotone non decreasing limits. Thus we get that

$$\begin{aligned} 2.1.14, 2.1.15, 2.1.16 \text{ and } 2.1.17 &\iff \\ 2.1.14, 2.1.15, 2.1.25 & \end{aligned} \qquad (2.1.29)$$

This leads us to the following definition:

Definition 2.1.1 A collection $\mathcal{B}$, of subsets of a set Ω, is said to be a $\boldsymbol{\sigma}$**-algebra** of subsets of Ω 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 σ-algebra Σ of subsets of Ω has the following additional properties:

1. By (2.1.16) we have Σ is closed under finite union and closed under monotone non decreasing limits
2. By DeMorgan's laws we have Σ 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 2.1.1 The smallest σ-algebra is the collection containing only the two sets Ω and ϕ, and the largest σ-algebra is the collection of all subsets of Ω, which is called the **Power Set** of Ω and denoted by either $\mathcal{P}(\Omega)$ or 2^Ω.

We next introduce the notion of the "Smallest σ-algebra Containing a Collection of Subsets"'
Consider a collection $\mathcal{S}$ of subsets of Ω. This collection $\mathcal{S}$ may or may not be a σ-algebra. For example, if

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

then let

$$\mathcal{S} = \{A_{12}, A_{34}, A_{56}\} \tag{2.1.30}$$

where

$$A_{12} = \{1, 2\} \tag{2.1.31}$$
$$A_{34} = \{3, 4\} \tag{2.1.32}$$
$$A_{56} = \{5, 6\} \tag{2.1.33}$$

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

$$\Sigma = \{\Omega, \phi, A_{12}, A_{34}, A_{56}, A_{1234}, A_{3456}, A_{1256}\} \tag{2.1.34}$$

where

$$A_{1234} = \{1, 2, 3, 4\} \tag{2.1.35}$$
$$A_{3456} = \{3, 4, 5, 6\} \tag{2.1.36}$$
$$A_{1256} = \{1, 2, 5, 6\} \tag{2.1.37}$$

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

1. $\mathcal{S} \subseteq \Sigma$ and
2. If Σ_1 is any σ-algebra that contains $\mathcal{S}$, that is $\mathcal{S} \subseteq \Sigma_1$, then $\Sigma \subseteq \Sigma_1$

Can we find such an optimal σ-algebra? It can be shown that this is possible and this smallest σ-algebra containig $\mathcal{S}$ is called the **σ-algebra generated by $\mathcal{S}$**, and is denoted by $\Sigma(\mathcal{S})$. In the above example the Σ given in 2.1.34 is the smallest σ-algebra generated by the $\mathcal{S}$ in 2.1.30.

Remark 2.1.2 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 σ-algebra generated by this collection of all intervals is called the **Borel σ-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 σ-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 σ-algebra of subsets of Ω. Thus we have now two main ingredients for Probability theory, namely the Sample Space and a σ-algebra of events.
We next look at the third important ingredient in Probability Theory, namely, the notion of a **Probability Measure**.

2.2 Probability

Let Ω be the sample space of a random experiment and $\mathcal{B}$ the σ-algebra of events. With each event we associate an "index" with certain protocols and call this index as the probability measure. More precisely we have

Definition 2.2.1 A map

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

is called a "**Probability Measure**" on Ω if it satisfies the following properties:

$$P(\Omega) = 1 \tag{2.2.1}$$

$$P(\phi) = 0 \tag{2.2.2}$$

$$0 \leq P(A) \leq 1 \text{ for all } A \in \mathcal{B} \tag{2.2.3}$$

$$\left.\begin{array}{l} \{A_n\}_n \in \mathbb{N} \text{ disjoint sets in } \mathcal{B} \\ \Longrightarrow \\ P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n) \end{array}\right\} \tag{2.2.4}$$

Remark 2.2.1 The property expressed in equation 2.2.4 is called "**Countable Additivity**" of the probability measure

We observe the following properties of a probability measure:

1. A Probability measure is "**Finitely Additive**, that is

$$\left.\begin{array}{l} \{A_j\}_{j=1}^{N} \text{ finite number of disjoint sets in } \mathcal{B} \\ \Longrightarrow \\ P\left(\bigcup_{j=1}^{N} A_j\right) = \sum_{j=1}^{N} P(A_j) \end{array}\right\} \tag{2.2.5}$$

   This is obtained by applying the countable additivity property 2.2.4 by taking $A_n = \phi$ for $n > N$ and using 2.2.2

2. Suppose $A \in \mathcal{B}$. Since $\mathcal{B}$ is a σ-algebra we have $A' \in \mathcal{B}$. Further these two sets are disjoint and $\Omega = A \cup A'$. Hence using finite additivity property above we get

$$P(A) + P(A') = P(\Omega) = 1$$

   and hence

$$P(A') = 1 - P(A) \text{ for every } A \in \mathcal{B} \tag{2.2.6}$$

3. Let $A \subseteq B \in \mathcal{B}$ such that $A \subseteq B$. Then we have

$$B = A \cup (B \setminus A)$$

$$\begin{aligned}
&\Longrightarrow \\
P(B) &= P\left(A \cup (B \setminus A)\right) \\
&= P(A) + P(B \setminus A) \text{ (using finite additivity)} \\
&\Longrightarrow \\
P(B \setminus A) &= P(B) - P(A)
\end{aligned}$$

Thus we have

$$A, B \in \mathcal{B} \text{ and } A \subseteq B \Longrightarrow P(B \setminus A) = P(B) - P(A) \tag{2.2.7}$$

4. An immediate consequence of the above is the "**Monotonicity**" of the probability measure. We have, from above

$$\begin{aligned}
A \subseteq B &\Longrightarrow P(B \setminus A) = P(B) - P(A) \\
&\Longrightarrow P(B) = P(A) + P(B \setminus A) \\
&\Longrightarrow P(B) \geq P(A) \text{ since } P(B \setminus A) \geq 0
\end{aligned}$$

Thus we have

$$\left.\begin{array}{l}
A, B \in \mathcal{B} \text{ and } A \subseteq B \\
\Longrightarrow \\
P(A) \leq P(B)
\end{array}\right\} \tag{2.2.8}$$

5. The next property is what is known as "**countable subadditivity**" of the probability measure. We have seen in 2.2.4 that the probability of a union of a sequence of sets is the sum of the individual probabilities when the sets are disjoint. We shall now see what happens when the sets are not necessarily disjoint. Let $\{A_n\}_{n \in \mathbb{N}}$ be a sequence of sets in $\mathcal{B}$ which may or may not be disjoint. We define a new sequence $\{B_n\}_{n \in \mathbb{N}}$ in $\mathcal{B}$ which are disjoint as follows:

$$\begin{aligned}
B_1 &= A_1 \\
B_n &= A_n \setminus \bigcup_{j=1}^{n-1} A_j \text{ for } n \geq 2
\end{aligned}$$

We see that

(a) $\displaystyle\bigcup_{j=1}^{n} B_j = \bigcup_{j=1}^{n} A_j$

(b) $\bigcup_{j=1}^{\infty} B_j = \bigcup_{j=1}^{\infty} A_j$

(c) B_n are disjoint

(d) $B_n \subseteq A_n$ for all $n \in \mathbb{N}$ and hence by 2.2.8 we get $P(B_n) \leq P(A_n)$ for all $n \in \mathbb{N}$

By property (b) above we have

$$\begin{aligned} P\left(\bigcup_{n=1}^{\infty} A_n\right) &= P\left(\bigcup_{n=1}^{\infty} B_n\right) \\ &= \sum_{n=1}^{\infty} P(B_n) \\ &\quad \text{(using countable additivity and the fact that } B_n \text{ are disjoint)} \\ &\leq \sum_{n=1}^{\infty} P(A_n) \\ &\quad \text{(by property (d) above)} \end{aligned}$$

Thus we have

$$\left.\begin{array}{l} \{A_n\}_{n\in\mathbb{N}} \in \mathcal{B} \\ \Longrightarrow \\ P\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} P(A_n) \end{array}\right\} \tag{2.2.9}$$

6. The next property we shall look at is a continuity property of the probability measure mown as "**Continuity from below"** property of the probability measure. Consider a sequence of sets $\{A_n\}_{n\in\mathbb{N}}$ in $\mathcal{B}$ which is nondecreasing, that is, $A_n \subseteq A_{n+1}$ for every $n \in \mathbb{N}$. For such a sequence we have

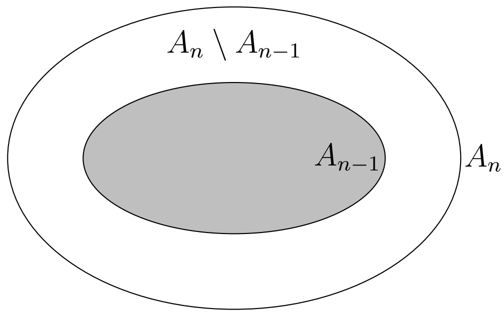
$$\lim_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n$$

We would like the probability measure reasonably continuous in the sense that for such sequences we must have

$$P(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} P(A_n)$$

We shall now see that this is indeed true. We define a sequence of sets $\{B_n\}_{n\in\mathbb{N}}$ as follows:

$$\begin{aligned} B_1 &= A_1 \\ B_n &= A_n \setminus A_{n-1} \text{ for } n \geq 2 \end{aligned}$$



It is easy to see that

(a) B_n are all in $\mathcal{B}$

(b) B_n are all disjoint

(c) $\displaystyle\bigcup_{j=1}^{n} A_j = \bigcup_{j=1}^{n} B_j$ for every $n \in \mathbb{N}$

(d) $\displaystyle\bigcup_{n=1}^{\infty} A_j = \bigcup_{j=1}^{\infty} B_j$

(e) $P(B_n) = P(A_n) - P(A_{n-1})$ (by equation 2.2.7)

We have

$$\begin{aligned} P\left(\bigcup_{j=1}^{\infty} A_j\right) &= P\left(\bigcup_{j=1}^{\infty} B_j\right) \text{ (by property (d) above)} \\ &= \sum_{n=1}^{\infty} P(B_n) \\ &\quad \text{(by the fact that } B_n \text{ are disjoint and } P \text{ is countably additive)} \\ &= \lim_{n\to\infty} \sum_{j=1}^{n} P(B_j) \end{aligned}$$

$$
\begin{aligned}
&= \lim_{n\to\infty} \{P(B_1) + P(B_2) + \cdots + P(B_n)\} \\
&= \lim_{n\to\infty} \{P(A_1) + (P(A_2) - P(A_1)) + \cdots + (P(A_n) - P(A_{n-1}))\} \\
&= \lim_{n\to\infty} P(A_n)
\end{aligned}
$$

Since $\bigcup_{n=1}^{\infty} A_n = \lim_{n\to\infty} A_n$ we get

$$
\left.
\begin{array}{l}
\{A_n\}_{n\in\mathbb{N}} \text{ nondecreasing sequence of sets in } \mathcal{B} \\
\Longrightarrow \\
P\left(\lim_{n\to\infty} A_n\right) = \lim_{n\to\infty} P(A_n) \text{ ,that is} \\
P\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} P(A_n)
\end{array}
\right\} \tag{2.2.10}
$$

7. The next property we shall look at is the dual of the above property, namely continuity from above, that is for non increasing sequences. Let $\{A_n\}_{n\in\mathbb{N}}$ be a non increasing sequence of sets in $\mathcal{B}$, that is , $A_{n+1} \subseteq A_n$ for every $n \in \mathbb{N}$. In this case we have

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

We define a new sequence of sets $\{B_n\}_{n\in\mathbb{N}}$ as $B_n = A_n'$. Then $\{B_n\}_{n\in\mathbb{N}}$ is a nondecreasing sequence of sets in $\mathcal{B}$ and hence by the above continuity from below property we get

$$
\begin{aligned}
P\left(\bigcup_{n=1}^{\infty} B_n\right) &= \lim_{n\to\infty} P(B_n) \\
&\Longrightarrow \\
P\left(\bigcup_{n=1}^{\infty} A_n'\right) &= \lim_{n\to\infty} P(A_n') \\
&\Longrightarrow \\
P\left(\left\{\bigcap_{n=1}^{\infty} A_n\right\}'\right) &= \lim_{n\to\infty} P(A_n') \\
&\Longrightarrow
\end{aligned}
$$

$$
\begin{aligned}
1 - P\left(\bigcap_{n=1}^{\infty} A_n\right) &= 1 - \lim_{n\to\infty} P(A_n) \\
&\Longrightarrow \\
P\left(\bigcap_{n=1}^{\infty} A_n\right) &= \lim_{n\to\infty} P(A_n)
\end{aligned}
$$

Thus we get

$$
\left.
\begin{array}{l}
\{A_n\}_{n\in\mathbb{N}} \text{ nonincreasing sequence of sets in } \mathcal{B} \\
\Longrightarrow \\
P\left(\lim_{n\to\infty} A_n\right) = \lim_{n\to\infty} P(A_n) \text{ ,that is} \\
P\left(\bigcap_{jn=1}^{\infty} A_j\right) = \lim_{n\to\infty} P(A_n)
\end{array}
\right\} \tag{2.2.11}
$$

8. Next we shall look at a general sequence of sets in $\mathcal{B}$. A general sequence of sets may not have a limit. However the limsup and liminf exist. We shall now observe a property through the notion of limsup of a sequence of sets. This property is known as the "**Borel-Cantelli**" Lemma. Consider a sequence $\{A_n\}_{n\in\mathbb{N}} \in \mathcal{B}$. Recall that we defined

$$
\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k
$$

Let

$$
B_n = \bigcup_{k=n}^{\infty} A_n = \sup_{k\geq n} A_k
$$

Then

$$
\lim_{n\to\infty} B_n = \limsup_{n\to\infty} A_n
$$

The sequence $\{B_n\}_{n\in\mathbb{N}}$ is a non increasing sequence of sets in $\mathcal{B}$ and hence by the property of continuity from above applied to the sequence $\{B_n\}_{n\in\mathbb{N}}$ (2.2.11 applied for B_n) we get

$$
\lim_{n\to\infty} P(B_n) = P\left(\lim_{n\to\infty} B_n\right)
$$

Substituting for B_n and using the fact that $\lim\limits_{n\to\infty} B_n = \limsup\limits_{n\to\infty} A_n$ we get

$$\begin{aligned}
P(\limsup_{n\to\infty} A_n) &= P\left(\lim_{n\to\infty}\bigcup_{k=n}^{\infty} A_k\right) \\
&= \lim_{n\to\infty} P\left(\bigcup_{k=n}^{\infty} A_k\right) \\
&\leq \lim_{n\to\infty}\sum_{k=n}^{\infty} A_k \text{ (by countable subadditivity property)} \\
&= 0 \text{ if } A_n \text{ are such that } \sum_{n=1}^{\infty} P(A_n) < \infty
\end{aligned}$$

Thus we have

Lemma 2.2.1 Borel-Cantelli Lemma

$$\{A_n\}_{n\in\mathbb{N}} \in \mathcal{B} \text{ and } \sum_{n=1}^{\infty} P(A_n) < \infty \Longrightarrow P(\limsup_{n\to\infty} A_n) = 0 \quad (2.2.12)$$

Remark 2.2.2 Since $\limsup\limits_{n\to\infty} A_n$ is the set of all those points in Ω which belong to an infinite number of the A_n sets we can write the Borel-Cantelli Lemma as

$$\left.\begin{array}{l}
\{A_n\}_{n\in\mathbb{N}} \in \mathcal{B} \text{ and } \sum_{n=1}^{\infty} P(A_n) < \infty \\
\Longrightarrow \\
P(\omega\in\Omega : \omega \in \text{ infinitely many of the } A_n) = 0
\end{array}\right\} \quad (2.2.13)$$

2.3 Random Variables

We shall next introduce the notion of a Random Variable. In most random experiments we are not directly interested in the outcome but certain consequences of the outcome.

Example 2.3.1 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 this pay off 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 Ω.
Let $(\Omega, \mathcal{B}, P)$ be a Probability space. Let

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

be a function defined on the sample space Ω. Then for each $\omega \in \Omega$ we have that $X(\omega)$ is a real number. We are interested in looking at the values of ω for which the values of the function $X(\omega)$ lie within a threshold value $x \in \mathbb{R}$, that is we are interested in the set

$$\{\omega \in \Omega : -\infty < X(\omega) \leq x\}$$

For any $x \in \mathbb{R}$ let us denote by I_x the interval $I_x = (-\infty, x]$. Thus we are interested in the set

$$X^{-1}(I_x) = \{\omega \in \Omega : X(\omega) \in I_x\}$$

Note that $X^{-1}(I_x)$ is a subset of Ω for every $x \in \mathbb{R}$. However this subset $X^{-1}(I_x)$ may not be an event, that is, $X^{-}(I_x)$ may not be in $\mathcal{B}$. If this set $X^{-1}(I_x)$ were in $\mathcal{B}$ then its probability $P(X^{-1}(x))$ is defined and hence we get the probability (or the chances) that the value of the function X lies within the threshold value of x, and we can do this for every $x \in \mathbb{R}$. We shall, therefore consider only such functions and call such a function as a "**Real Valued Random Variable**" defined on the probability space $(\Omega, \mathcal{B}, P)$. We therefore give the following definition:

Definition 2.3.1 A function $\boldsymbol{X} : \boldsymbol{\Omega} \longrightarrow \mathbb{R}$ is said to be a Real Valued Random Variable on a probability space $(\boldsymbol{\Omega}, \boldsymbol{\mathcal{B}}, \boldsymbol{P})$ if

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

Example 2.3.2 Let us again consider the random experiment of rolling a fair die as in Example 2.3.1. In that example we had taken $\mathcal{B}$ to be the power set and hence for every x in $\mathbb{R}$ the set $X^{-1}(I_x)$ is in $\mathcal{B}$ thereby maing every function $X : \Omega \longrightarrow \mathbb{R}$ a Real Valued Random Variable in this case.
Let us now consider $\mathcal{B}$ to be the following σ-algebra:

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

Consider the random variable X as in Example 2.3.1. Let $x = 2$ and consider the interval

$$I_2 = (-\infty, 2]$$

We have

$$\begin{aligned} X^{-1}(I_2) &= \{\omega \in \Omega : X(\omega) = -5 \text{ or } -3 \text{ or } -1 \text{ or } 2\} \\ &= \{5,3,1,2\} \\ &\notin \mathcal{B} \end{aligned}$$

Note that we have an $x \in \mathbb{R}$ namely $x = 2$ such that $X^{-1}(I_x) \notin \mathcal{B}$. Hence this X is not a Real Valued Random Variable on the probability space $(\Omega, \mathcal{B}, P)$ where $\mathcal{B}$ is as defined as above. Note that the same function on Ω may or may not be a random variable depending on the σ-algebra of events that is under consideration.

Example 2.3.3 Consider the random experiment of rolling a fair die repeatedly until a 6 appears. The 6 may appear in the first roll or it does not appear in the first $(n-1)$ rolls and appears in the nth roll, (for $n = 2, 3, \cdots$). Thus the sample space Ω can be written down as follows:
Let $A = \{1,2,3,4,5\}$ and define

$$S_1 = \{6\}$$

and for $n \geq 2$ define

$$S_n = \{k_1 k_2 \cdots k_{n-1} 6 : k_j \in A\}$$

S_1 denotes the outcome in which the 6 appears in the first roll itself. For $n \geq 2$ the set S_n denotes the outcomes in which the first time 6 appears is in the nth roll. For each $n \geq 1$ the set S_n has 5^{n-1} elements. Then we have

$$\Omega = \bigcup_{n=1}^{\infty} S_n$$

Let us take $\mathcal{B}$, the σ-algebra of events, to be the power set $\mathcal{P}(\Omega)$. The probability measure is defined by

$$\begin{aligned} P(\omega) &= \frac{1}{6} \text{ for } \omega \in S_1 \\ P(\omega) &= \frac{1}{6^n} \text{ for any } \omega \in S_n \end{aligned}$$

Since $\mathcal{B}$ is the power set of Ω, every function $X : \Omega \longrightarrow \mathbb{R}$ is a Real Valued Random Variable on this probability space $(\Omega, \mathcal{B}, P)$. Consider the following Real Valued Random Variable:

$$X(\omega) = \textit{Number of Rolls in } \omega$$

For example

$$\begin{aligned} X(6) &= 1 \\ X(36) &= 2 \\ X(56) &= 2 \\ X(436) &= 3 \end{aligned}$$

We see that for $n \geq 1$,

$$P(S_n) = \frac{5^{n-1}}{6^n}$$

Note that the set of possible values that the random variable X can take is

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

Let us consider $I_3 = (-\infty, 3]$. Then

$$\begin{aligned} X^{-1}(I_3) &= \{\omega \in \Omega : X(\omega) \leq 3\} \\ &= S_1 \cup S_2 \cup S_3 \\ \Longrightarrow \\ P(X^{-1}(I_3)) &= P(S_1) + P(S_2) + P(S_3) \\ &= \frac{1}{6} + \frac{5}{6^2} + \frac{5^2}{6^3} \end{aligned}$$

Note that for any x such that $3 \leq x < 4$ we have

$$\begin{aligned} P(X^{-1}(I_3)) &= P(S_1) + P(S_2) + P(S_3) \\ &= \frac{1}{6} + \frac{5}{6^2} + \frac{5^2}{6^3} \end{aligned}$$

Example 2.3.4 Let us consider the random experiment of rolling a fair die 6 times. Then the sample space can be written as

$$\Omega = \{k_1k_2k_3k_4k_5k_6 : k_j \in \{1,2,3,4,5,6\}\}$$

There are 36 elements in Ω. Let $\mathcal{B}$ be again the power set of Ω so that every function $X : \Omega \longrightarrow \mathbb{R}$ is a Real Valued Random Variable. Let the probability measure be defined as

$$P(\omega) = \frac{1}{36} \text{ for every } \omega \in \Omega$$

Consider the random variable X defined as

$$X(\omega) = Number\ of\ sixes\ in\ \omega$$

For example

$$X(122636) = 2$$

The set of all possible values this random variable can take is

$$\mathcal{R}_X = \{0,1,2,3,4,5,6\}$$

Let $x = 3$ and consider the interval $I_3 = (-\infty, 3]$. Then

$$\begin{aligned} X^{-1}(I_3) &= \{\omega \in \Omega : \text{there are at most 3 sixes in } \omega\} \\ &= S_0 \cup S_1 \cup S_2 \cup S_3 \text{ where} \\ S_k &= \{\omega \in \Omega : \text{there are exactly } k \text{ sixes in } \omega\} \end{aligned}$$

Hence

$$\begin{aligned} P(X^{-1}(I_3)) &= P(S_0) + P(S_1) + P(S_2) + P(S_3) \\ &= \left(\frac{5}{6}\right)^6 + \binom{6}{1}\frac{5^5}{6^6} + \binom{6}{2}\frac{5^4}{6^2} + \binom{6}{3}\frac{5^3}{6^3} \\ &= \sum_{k=0}^{3} \binom{6}{k}\frac{5^{6-k}}{6^6} \end{aligned}$$

If we had rolled the die N times instead of 6 times then we have

$$\begin{aligned} P(X^{-1}(I_3)) &= P(S_0) + P(S_1) + P(S_2) + P(S_3) \\ &= \frac{5^N}{6^N} + \binom{N}{1}\frac{5^{N-1}}{6^N} + \binom{N}{2}\frac{5^{N-2}}{6^N} + \binom{N}{3}\frac{5^{N-3}}{6^N} \\ &= \sum_{k=0}^{3}\binom{N}{k}\frac{5^{N-k}}{6^N} \end{aligned}$$

Remark 2.3.1 Suppose now X is a Real Valued Random Variable on the probability space $(\Omega, \mathcal{B}, P)$. Let us consider the following collection of subsets of $\mathbb{R}$:

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

We observe the following:

1. Clearly $\mathcal{C}$ is a nonempty collection because all sets of the form $A = I_x$, for any $x \in \mathbb{R}$, are in $\mathcal{C}$ since X is a Real Valued Random Variable.

2. We have

$$\begin{aligned} A \in \mathcal{C} &\implies X^{-1}(A) \in \mathcal{B} \\ &\implies (X^{-1}(A))' \in \mathcal{B} \text{ (since } \mathcal{B} \text{ is a } \sigma\text{-algebra)} \\ &\implies X^{-1}(A') \in \mathcal{B} \text{ (since } X^{-1}(A') = (X^{-1}(A))') \\ &\implies A' \in \mathcal{B} \end{aligned}$$

   Thus $\mathcal{C}$ is closed under complementation

3. Further

$$\begin{aligned} \{A_n\}_{\mathbb{N}} \in \mathcal{C} &\implies X^{-1}(A_n) \in \mathcal{B} \\ &\implies \bigcup_{n=1}^{\infty} X^{-1}(A_n) \in \mathcal{B} \text{ (since } \mathcal{B} \text{ is a } \sigma\text{-algebra)} \\ &\implies X^{-1}\left(\bigcup_{n=1}^{\infty} A_n\right) \in \mathcal{B} \\ &\implies \bigcup_{n=1}^{\infty} A_n \in \mathcal{C} \end{aligned}$$

   Thus $\mathcal{C}$ is closed under countable union

From the above three properties we see that $\mathcal{C}$ is a σ-algebra of subsets of $\mathbb{R}$. Further all intervals of the form $I_x = (-\infty, x]$ are in $\mathcal{C}$. Thus $\mathcal{C}$ is a σ-algebra that contains all intervals of this form. But the Borel σ-algebra $\mathcal{B}_{\mathbb{R}}$ is the smallest σ-algebra that contains all these intervals. Hence $\mathcal{B}_{\mathbb{R}} \subseteq \mathcal{C}$. Thus we have

$$\left.\begin{array}{l} X \text{ is a Real Valued Random Variable} \\ \text{on the probability space } (\Omega, \mathcal{B}, P) \\ \Longrightarrow \\ X^{-1}(A) \in \mathcal{B} \text{ for every Borel set in } \mathbb{R} \end{array}\right\} \tag{2.3.1}$$

In particular since every interval is a Borel set we see that

$$\left.\begin{array}{l} X \text{ is a Real Valued Random Variable} \\ \text{on the probability space } (\Omega, \mathcal{B}, P) \\ \Longrightarrow \\ X^{-1}(\mathcal{I}) \in \mathcal{B} \text{ for every interval } \mathcal{I} \text{ in } \mathbb{R} \end{array}\right\} \tag{2.3.2}$$

We shall next see some examples of random variables. We shall consider two types of random variables, namely

- Discrete Random Variables
- Continuous Random Variables

2.4 Discrete Random variables

We shall first consider Discrete random variables. We introduce two types of discrete random variables

1. Random Variables which take a finite number of real values, that is, the Range $\mathcal{R}_X$ of X is a finite set in $\mathbb{R}$,

$$\mathcal{R}_X = \{x_1, x_2, x_3, \cdots, x_N\}$$

where we arrange the values as

$$x_1 < x_2 < x_3 < \cdots < x_N$$

2. Random variables which take an infinite sequence of values, that is, the Range $\mathcal{R}_X$ of X is a sequence,

$$\mathcal{R}_X = \{x_1, x_2, x_3, \cdots, x_n, \cdots\}$$

where we arrange the values as

$$x_1 < x_2 < x_3 < \cdots < x_n < x_{N+1} < \cdots$$

Random Variables Taking A Finite Number Of Values

We shall first consider those random variables which take a finite number of values. Let

$$\mathcal{R}_X = \{x_1, x_2, \cdots, x_N : x_j \in \mathbb{R} \text{ for } 1 \leq j \leq N\}$$

where

$$x_1 < x_2 < \cdots < x_N$$

Since there are only finite number of values for X we are basically interested in the values $P(\omega \in \Omega : X(\omega) = k)$ for $k = 1, 2, \cdots, N$. (From now on we shall write $P(\omega \in \Omega : X(\omega) = k)$ as $P(X = k)$ and $P(\omega \in \Omega : X(\omega) \leq x)$ as $P(X \leq x)$). We shall denote $P(X = k)$ as p_k. Then we have a function

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

defined as

$$p_X(k) = p_k = P(X = k)$$

This function is called the "**Probability Mass Function**" (or pmf in short) of the random variable X. Once we know the pmf we have

$$P(X \in A) = \sum_{\{k : x_k \in A\}} p_k \text{ for any Borel set } A \text{ in } \mathbb{R}$$

Thus the pmf is the basic function that gives us the distribution of the values of the random variable X. We shall now see some typical examples:

Example 2.4.1 The simplest example is the random variable which takes only one value - say C. Then we have

$$\mathcal{R}_X = \{C\}$$

and the pmf is given by

$$p_X(X = C) = 1 \text{ and}$$

For any $x \neq C$ we have $P(X = x) = 0$ and for any Borel set A in $\mathbb{R}$ we have

$$P(X \in A) = \begin{cases} 1 & \text{if} \quad C \in A \\ 0 & \text{if} \quad C \notin A \end{cases}$$

Such random variables are called constant random variables

Example 2.4.2 The next example is that of a random variable that takes exactly two values - which are referred to as Success or Failure. We shall denote success as 1 and Failure as 0. Then we have

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

Then the pmf is known the moment $P(X = 1) = p$ is known, (referred to as the probability of "success"). Then we have $P(X = 0) = 1 - p = q$,say - (the probability of "failure"). For any Borel set A in $\mathbb{R}$ we have

$$P(A) = \begin{cases} 1 & \text{if} \quad 0 \text{ and } 1 \in A \\ p & \text{if} \quad 1 \in A \text{ and } 0 \notin A \\ q = 1 - p & \text{if} \quad 0 \in A \text{ and } 1 \notin A \\ 0 & \text{if} \quad 1 \text{ and } 0 \notin A \end{cases}$$

Such a random variable is called a "**Bernoulli Random Variable**" with success probability p. We write such a random variable a as $\boldsymbol{Ber(p)}$ random variable.

As an illustration we shall consider the following two Bernoulli random variables:

1. Consider tossing a coin with probability of getting a Head as p (where $0 < p < 1$). Define
$$X : \Omega \longrightarrow \mathbb{R}$$
as
$$X(H) = 1 \text{ and } X(T) = 0$$
Then X is a Bernoulli Random Variable with success probability p. This is a $Ber(p)$ random variable. For a fair coin we get a $Ber(0.5)$ random variable.

2. Consider the experiment of rolling a fair die. Let us consider getting 6 as a success and getting anything other than 6 a Failure. Then the random variable X becomes

$$X(\omega) = \begin{cases} 1 & \text{if} \quad \omega = 6 \\ 0 & \text{if} \quad \omega \neq 6 \end{cases}$$

   We have

$$P(X = 1) = \frac{1}{6} \text{ and } P(X = 0) = \frac{5}{6}$$

   This is a $Ber(\frac{1}{6})$ random variable.

Example 2.4.3 Let us now consider a random variable which take finite number of values. Without loss of generality let these values be $0, 1, 2, \cdots, N$. As an illustration let us consider the experiment of tossing a coin (with probability p for success) N times. Let us assume that

1. the probability of getting a Head in each toss is p and

2. the outcome in any toss is independent of the outcomes in the other tosses

Thus for instance $N = 5$ and an outcome is $HTTHT$ then $P(HTTHT) = p^2(1-p)^3$. In general

$$P(\omega) = p^k(1-p)^{N-k} \text{ where } k = \text{ Number of Heads in } \omega$$

Let us define a random variable X as follows:

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

Then X can take values $0, 1, 2, \cdots, N$, that is,

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

Then we have

$$P(X = k) = \left(\begin{array}{c} N \\ k \end{array} \right) p^k(1-p)^{N-k}$$

If the coin is fair this becomes

$$P(X = k) = \left(\begin{array}{c} N \\ k \end{array} \right) \frac{1}{2^N} \text{ since } p = \frac{1}{2} \text{ for a fair coin}$$

In the case of rolling a fair die where getting a six is treated as success we have $p = \frac{1}{6}$ and hence we get

$$P(X = k) = \left(\begin{array}{c} N \\ k \end{array} \right) \frac{1}{6^k} \times \frac{5^{N-k}}{6^{N-k}} = \left(\begin{array}{c} N \\ k \end{array} \right) \frac{5^{N-k}}{6^N}$$

Such Random Variables are said to have the "**Binomial Distribution**" (and are also called **Bernoulli Trials**). Thus we have

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

$$P(X = k) = \left(\begin{array}{c} N \\ k \end{array} \right) p^k (1-p)^{N-k}$$

Such random variables are referred to as $\boldsymbol{B(N, p)}$ random variables

Example 2.4.4 Consider a random variable X for which again

$$\mathcal{R}_X = \{x_1, x_2, \cdots, x_N\} \tag{2.4.1}$$

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

$$p_k = P(X = k) \propto \frac{1}{k} \tag{2.4.2}$$

Let C be the constant of proportionality. Then we have

$$p_k = P(X = k) = C \times \frac{1}{k} \tag{2.4.3}$$

Since the total probability must be one we get

$$\sum_{k=1}^{N} \left(C \times \frac{1}{k} \right) = 1$$

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

$$p_k = P(X = k) = \frac{1}{s_{_N}} \frac{1}{k} \tag{2.4.4}$$

Such a RV is called a **Zipf Random Variable**. Thus for a Zipf Random Variable X we have

$$\begin{aligned} \mathcal{R}_X &= \{x_1, \cdots, x_N\} \\ p_k = P(X = x_k) &= \frac{1}{s_N} \frac{1}{k} \text{ (for } k = 0, 1, 2 \cdots, N) \end{aligned}$$

where

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

Remark 2.4.1 We can reverse the situation above and get a random variable which takes higher values with higher probabilities. Let us define

$$\mathcal{R}_X = \{x_1, x_2, \cdots, x_N\} \tag{2.4.5}$$

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

$$P(X = x_k) = \frac{1}{s_N} \frac{1}{N + 1 - k} \tag{2.4.6}$$

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

$$a_1 < a_2 < \cdots < a_{_N} \tag{2.4.7}$$

Let

$$C = \sum_{k=1}^{N} a_k \tag{2.4.8}$$

Then for a random variable X for which

$$\mathcal{R}_X = \{x_1, x_2, \cdots, x_N\} \tag{2.4.9}$$

where $x_1 < x_2 < \cdots < x_N$ we can define

$$p_k = P(X = x_k) = \frac{a_k}{C} \tag{2.4.10}$$

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

$$p_k = P(X = x_k) = \frac{a_{_{(N+1-k)}}}{C} \tag{2.4.11}$$

Now the higher values are attained with higher probabilities.

Discrete Random Variables Taking An Infinite Sequence Of Values

We shall next look at some discrete random variables which take an infinite sequence of values. In such cases we have

$$\mathcal{R}_X = \{x_1, x_2, \cdots, x_n, \cdots,\}$$

where

$$x_1 < x_2 < x_3, \cdots < x_{n-1} < x_n < \cdots$$

Such random variables are typically used where we repeat an experiment until we get a "success" and count the number of "failures" before getting a success, the random veritable being the number of failures before the first success. We shall describe this first with an example

Example 2.4.5 Consider the experiment of rolling a fair die until we get a six. Getting a six is treated as a success while that of getting any other number is treated as a failure. (We assume that in each roll each number is equally likely to show up independent of the other rolls). We can describe the sample space of this experiment as follows:
Let S_n denote the event that the success occurs in the nth roll. Let

$$A = \{1, 2, 3, 4, 5\}$$

Then

$$\begin{aligned} S_1 &= \{6\} \\ S_2 &= \{a_1 6 : a_1 \in A\} \end{aligned}$$

and in general for $n \geq 2$ we have

$$S_n = \left\{a_1 a_2 \cdots a_{(n-1)} 6 : a_j \in A\right\}$$

We have the sample space

$$\Omega = \bigcup_{n=1}^{\infty} S_n$$

$$\begin{aligned} P(S_1) &= \frac{1}{6} \\ P(S_2) &= \frac{5}{6} \times \frac{1}{6} = \frac{5}{6^2} \end{aligned}$$

and, in general, for $n \geq 1$ we have

$$P(S_n) = \frac{5^{(n-1)}}{6^n}$$

Let us define the random variable $X : \Omega \longrightarrow \mathbb{R}$ as

$$X(\omega) = \#\text{ of rolls in } \omega$$

For example,

$$X(\omega) = n \text{ for every } \omega \in S_n$$

We have

$$P(X = n) \;=\; P(S_n) = \frac{5^{(n-1)}}{6^n}$$

In general we can take the probability of success as p and that of failure as $(1 - p)$ then we get above

$$P(X = n) \;=\; p(1-p)^{(n-1)}$$

Such a random variable is called a "**Geometric Random Variable**" and we denote this by $\boldsymbol{Geo(p)}$. We write $\boldsymbol{X \sim Geo(p)}$.

In general we take

$$\mathcal{R}_X = \{x_0, x_1, x_2, \cdots, x_n, \cdots\}$$

where

$$x_0 < x_1 < x_2 < x_3 < \cdots < x_{(n-1)} < x_n < \cdots$$

We then choose a sequence of positive real numbers $p_0, p_1, p_2, \cdots, p_n, \cdots$ such that

$$0 < p_n < 1 \text{ for } n = 0, 1, 2, \cdots$$
$$\sum_{n=0}^{\infty} p_n = C < \infty$$

We then define

$$P(X = x_n) \;=\; \frac{p_n}{C} \text{ for } n = 0, 1, 2, \cdots$$

We next look at an example of this type of random variable by choosing suitable p_j.

Example 2.4.6 Without loss of generality we assume

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

Let λ be a fixed positive real number. Consider the infinite series

$$\sum_{n=0}^{\infty} \frac{\lambda^n}{n!} = e^{\lambda}$$

($C = e^{\lambda}$ in this case) Then we can choose

$$P(X = n) = p_n \;=\; \frac{\lambda^n}{n!} e^{-\lambda} \text{ for } n = 0, 1, 2, \cdots$$

Such a random variable is called "**Exponential random Variable**" and is denoted by $\boldsymbol{Exp(\lambda)}$. We write $\boldsymbol{X \sim Exp(\lambda)}$

2.5 Continuous Random Variables

We shall next consider Random Variables which take a continuum of real values. 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. Before we introduce these random variables we make the following observations:
In the case of discrete random variables the basic pieces of information we provide are

1. The Range of the Random Variable, $\mathcal{R}_X$ (which is a discrete set) and

2. The probability $p_k = P(X = x_k)$ for every $x_k \in \mathcal{R}_X$

We can then define a function

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

as

$$p_X(x_k) = p_k = P(X = x_k)$$

This function is called the "**Probability Mass Function**" (in short we write pmf) of the random variable X. Thus the pmf is the basic piece of information we need to prescribe a discrete random variable.
However this gets a little involved in the case of random variables which take a continuum of values. We shall briefly describe the process involved in describing a random variable, in general. We shall now see that the random variable X and the probability measure on the probability space $(\Omega, \mathcal{B}, P)$ together induce a probability measure P_X on the Borel sets in $\mathbb{R}$ as follows:
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 $\mathcal{B}_1 = \mathcal{B}_{\mathbb{R}}$. For any Borel set $B \in \mathcal{B}_{\mathbb{R}}$ we have $X^{-1}(B) \in \mathcal{B}$ and hence we can define $P(X^{-1}(B))$. We denote this by $P_X(B)$. Thus we have a function

$$P_X : \mathcal{B}_{\mathbb{R}} \longrightarrow \mathbb{R} \tag{2.5.1}$$

We observe the following properties of the function P_X:

1. We have

$$P_X(B) \;=\; P(X^{-1}(B))$$

   Since the right hand side is a probability we get

$$0 \leq P_X(B) \leq 1 \text{ for every } B \in \mathcal{B}_{\mathbb{R}} \tag{2.5.2}$$

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

$$P_X(\mathbb{R}) = P(\Omega) \;=\; 1 \tag{2.5.3}$$
$$P_X(\phi) = P(\phi) \;=\; 0 \tag{2.5.4}$$

3. 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,

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

$$P_X\left(\overrightarrow{\bigcup_{n=1}^{\infty} B_n}\right) = \sum_{n=1}^{\infty} P_X(B_n)$$

Hence we get that
P_X is countably additive additive

The above properties show that P_X is a probability measure on the Borel sets of $\mathbb{R}$. We call this the probability measure P_X on the Borel σ-algebra $\mathcal{B}_{\mathbb{R}}$ in $\mathbb{R}$, as the measure induced by P and X. We shall now see that this induced probability measure P_X can be analysed using a real valued function called the CDF (**Cumulative Distribution Function**) of the random variable X. Let X be a random variable and x be any real number. Then consider the interval,

$$I_x = (-\infty, x] \tag{2.5.5}$$

Then since I_x is a Borel set,

$$X^{-1}(I_x) \in \mathcal{B} \tag{2.5.6}$$

This means

$$\{\omega : X(\omega) \in I_x\} \in \mathcal{B} \tag{2.5.7}$$

which gives

$$\{\omega : X(\omega) \leq x\} \in \mathcal{B} \quad \forall x \in \mathbb{R} \tag{2.5.8}$$

Hence we can define

$$P_X(I_x) = P(-\infty < X(\omega) \leq x) \text{ for every } x \in \mathbb{R} \tag{2.5.9}$$

Thus we have a function

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

defined as

$$F_X(x) = P(X \leq x) \tag{2.5.10}$$

This function is called the "**Cumulative Distribution Function**" (in short we write cdf) of the random variable X. It is through this function we

describe a random variable which takes a continuum of values. We shall first look at the properties of this function so that we understand what sort of functions can be cdf of random variables.
Let us first look at discrete random variables and their cdf. (We have already seen what is meant by the pmf for such random variables). .

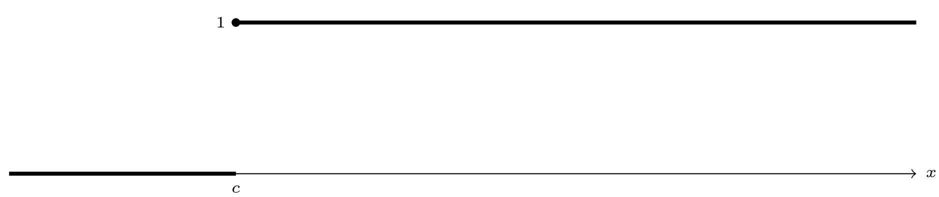
Example 2.5.1 Consider a random variable $X(\omega)$ which takes only one value say c on the probability space $(\Omega, \mathcal{E}, P)$.
Then we have

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

is given by

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

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

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

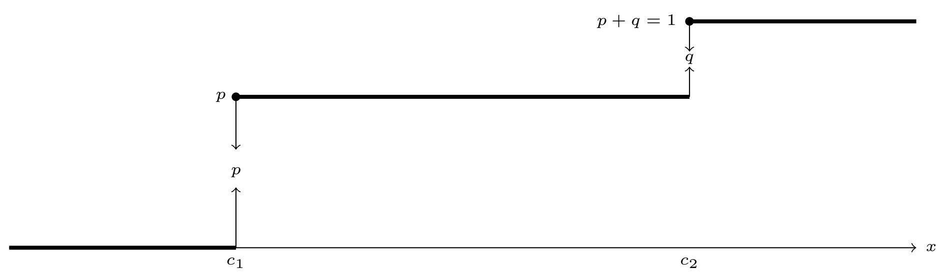
where

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

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

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

$F_X(x)$ 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

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

where

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

then

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

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 fundamental properties of the cumulative distribution function $F_X(x)$ of a random variable X.

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

2. $\boldsymbol{x_1 < x_2 \Longrightarrow F_X(x_1) \le F_X(x_2)}$,**that is, $\boldsymbol{F_X(x)}$ is a nondecreasing function**.
   This follows from the fact

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

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

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

   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 $\lim\limits_{x \longrightarrow -\infty} F_X(x) = 0$

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

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

   is a sequence of events decreasing to the event

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

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

$$\begin{aligned} \mathcal{P}(E) &= \lim_{n\to\infty} \mathcal{P}(E_n) \\ \Longrightarrow & \end{aligned}$$

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

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. We can,using the properties of the probability measure, easily find the probabilities of the sets $\{X(\omega) \in I\}$ where I is any interval. We observe the following:

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

$$\begin{aligned}
\mathcal{P}(\{X(\omega) \in (x, \infty\}) &= \mathcal{P}(\{X(\omega) \in (-\infty, x]\}') \\
&= 1 - \mathcal{P}(\{X(\omega) \in (-\infty, x]\}) \\
&= 1 - F_X(x)
\end{aligned}$$

Thus

$$\mathcal{P}(\{X(\omega) \in (x, \infty)\}) = 1 - F_X(x) \tag{2.5.11}$$

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

$$\mathcal{P}(\{X(\omega) \in (a, b]\}) = F_X(b) - F_X(a) \tag{2.5.12}$$

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

$$\mathcal{P}(\{X(\omega) \in (-\infty, x)\}) = F_X(x-) \qquad (2.5.13)$$

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

$$\begin{aligned}
\mathcal{P}(\{X(\omega) \in (a, b)\}) &= F_X(b-) - F_X(a) && (2.5.14) \\
\mathcal{P}(\{X(\omega) \in [a, b)\}) &= F_X(b-) - F_X(a-) && (2.5.15) \\
\mathcal{P}(\{X(\omega) \in [a, b]\}) &= F_X(b) - F_X(a-) && (2.5.16)
\end{aligned}$$

5. We can write

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

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

$$\mathcal{P}(\{X = x\}) = F_X(x) - F_X(x-) \qquad (2.5.17)$$

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 by 2.5.17

$$\mathcal{P}(\{X = x\}) = 0 \text{ for any continuous random variable } X \quad (2.5.18)$$

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,

$$\mathcal{P}(\{X \in I\}) = F_X(b) - F_X(a) \quad (2.5.19)$$

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 derivative exists 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:

$$f_X(x) = \begin{cases} \frac{d}{dx}F_X(x) \text{ whenever the derivative exists at } x \text{ and,} \\ \text{any arbitrary nonnegative real value at other points} \end{cases} \quad (2.5.20)$$

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

$$f_X(x) \geq 0 \text{ for all } x \in \mathbb{R} \quad (2.5.21)$$

Moreover, we have

$$F_X(x) = \int_{-\infty}^{x} f_X(s)ds \quad (2.5.22)$$

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

$$\int_{-\infty}^{\infty} f_X(x)dx = 1 \quad (2.5.23)$$

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

$$\mathcal{P}(\{X \in I\}) = \int_a^b f_X(x)dx \tag{2.5.24}$$

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

Remark 2.5.1 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

$$f_X(x) = \sum_j p_j \delta(x - x_j) \tag{2.5.25}$$

We shall next see some examples of continuous random variables

2.6 Examples Of Continuous Random Variables

We shall consider the following three types of random variables:

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

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

   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.

Bounded Random Variables

Let

$$\mathcal{R}_X = [a, b] \text{ where } -\infty < a \leq x \leq b < \infty \tag{2.6.1}$$

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

$$F_X(x) = P(X \leq x) = 0 \text{ if } x \leq a \tag{2.6.2}$$

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

$$F_X(x) = P(X \geq x) = 1 \text{ if } x > b \tag{2.6.3}$$

Thus different such random variables are obtained depending on how $F_X(x)$ increases from 0 at $x = a$ to to 1 at $x = b$. We do this as follows:
Let $g(x)$ be a nondecreasing, nonnegative, continuous real valued function defined over the interval $[a, b]$. If we define $h(x) = g(x) - g(a)$ then $h(x)$ is a nondecreasing, nonnegative, continuous real valued function defined over the interval $[a, b]$ such that $h(a) = g(a) - g(a) = 0$. If we now define

$$\varphi(x) = \frac{h(x)}{h(b)} = \frac{g(x) - g(a)}{g(b) - g(a)}$$

then we get $\varphi(x)$ is a nondecreasing, nonnegative, continuous real valued function defined over the interval $[a, b]$ such that $\varphi(a) = 0$ and $\varphi(b) = 1$. Hence we can define a random variable X by the cdf

$$F_X(x) = \begin{cases} 0 & \text{for } \ x < a \\ \dfrac{g(x) - g(a)}{g(b) - g(a)} & \text{for } \ a \leq x \leq b \\ 1 & \text{for } \ x > a \end{cases}$$

For such random variables, we get by differentiating the cdf, the corresponding pdf as

$$f_X(x) = \begin{cases} 0 & \text{for} \quad x < a \\ \dfrac{g'(x)}{g(b) - g(a)} & \text{for} \quad a \leq x \leq b \\ 1 & \text{for} \quad x > a \end{cases}$$

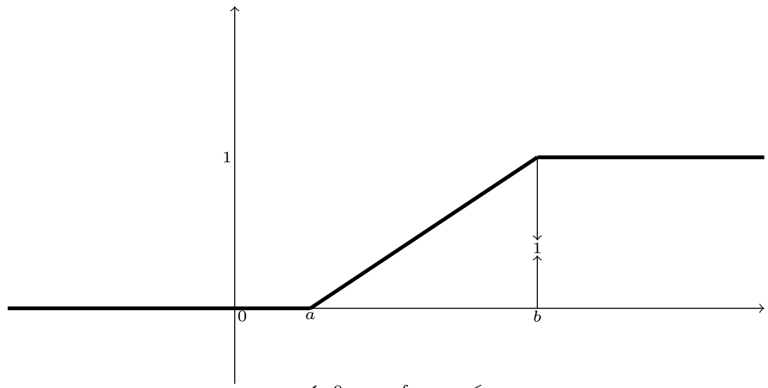
By varying a, b and g we get different Bounded Random Variables.
The simplest model is obtained by making $F_X(x)$ vary linearly from 0 at $x = a$ to 1 at $x = b$, that is, by taking $g(x) = x$. Thus $F_X(x)$ is be given by

$$F_X(x) = \begin{cases} 0 & \text{for} \quad x < a \\ \dfrac{x-a}{b-a} & \text{for} \quad a \leq x \leq b \\ 1 & \text{for} \quad x > a \end{cases} \tag{2.6.4}$$

The corresponding pdf is obtained as the derivative of the cdf as,

$$f_X(x) = \begin{cases} 0 & \text{for} \quad x < a \\ \dfrac{1}{b-a} & \text{for} \quad a \leq x \leq b \\ 1 & \text{for} \quad x > a \end{cases} \tag{2.6.5}$$

Such a Random Variable is said to be **Uniformly Distributed** over the interval $[a, b]$. We call such Random Variables as **Uniform Random Variables** and write $\boldsymbol{X \sim Uni[a, b]}$. The graph of the CDF is as shown below:



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

If we choose $g(x) = x^2$ then $F_X(x)$ varies quadratically over $[a, b]$ from the value 0 at $x = a$ to the value 1 at $x = b$ and we have a Random Variable X with cdf $F_x(x)$ given by

$$F_X(x) = \begin{cases} 0 & \text{for} \quad x < a \\ \dfrac{(x-a)^2}{(b-a)^2} & \text{for} \quad a \le x \le b \\ 1 & \text{for} \quad x > a \end{cases} \tag{2.6.6}$$

By differentiating the cdf we get the corresponding pdf as

$$f_X(x) = \begin{cases} 0 & \text{for} \quad x < a \\ \dfrac{2(x-a)}{(b-a)^2} & \text{for} \quad a \le x \le b \\ 1 & \text{for} \quad x > a \end{cases} \tag{2.6.7}$$

If we choose $g(x) = -e^{-x}$ then $F_X(x)$ varies exponentially over $[a, b]$ from the value 0 at $x = a$ to the value 1 at $x = b$ and we have a Random Variable

X with cdf $F_x(x)$ given by

$$F_X(x) = \begin{cases} 0 & \text{for} \quad x < a \\ \\ \dfrac{e^{-a} - e^{-x}}{e^{-a} - e^{-b}} & \text{for} \quad a \leq x \leq b \\ \\ 1 & \text{for} \quad x > a \end{cases} \tag{2.6.8}$$

By differentiating the cdf we get the corresponding pdf as

$$F_X(x) = \begin{cases} 0 & \text{for} \quad x < a \\ \\ \dfrac{e^{-x}}{e^{-a} - e^{-b}} & \text{for} \quad a \leq x \leq b \\ \\ 1 & \text{for} \quad x > a \end{cases} \tag{2.6.9}$$

__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

$$F_{_X}(x) = 0 \text{ for } x < 0 \tag{2.6.10}$$

On $[0, \infty)$, we must have $F_{_X}(x)$ to be a nondecreasing, continuous function such that

$$F_{_X}(0) = 0 \text{ and} \tag{2.6.11}$$

$$\lim_{x \to +\infty} F_{_X}(x) = 1 \tag{2.6.12}$$

We shall now look at examples of such random variables:

Exponential Random Variable:

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

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

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

$$F_X \;=\; \begin{cases} 0 & \text{for} \quad x < 0 \\ 1 - e^{-\lambda x} & \text{for} \quad x \geq 0 \end{cases} \tag{2.6.13}$$

The corresponding pdf is given by

$$f_X \;=\; \begin{cases} 0 & \text{for} \quad x < 0 \\ \lambda e^{-\lambda x} & \text{for} \quad x \geq 0 \end{cases} \tag{2.6.14}$$

Such a RV is called "**Exponential Random Variable**" and for any such random variable we write $\boldsymbol{X \sim Exp(\lambda)}$.

Rayleigh Random Variable

The function

$$h(x) \;=\; 1 - e^{-\beta^2 x^2} \text{ (where } \beta \text{ is real and nonzero)} \tag{2.6.15}$$

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

$$F_X \;=\; \begin{cases} 0 & \text{for} \quad x < 0 \\ 1 - e^{-\beta^2 x^2} & \text{for} \quad x \geq 0 \end{cases} \tag{2.6.16}$$

The corresponding pdf is given by

$$f_X \;=\; \begin{cases} 0 & \text{for} \quad x < 0 \\ 2\beta^2 x e^{-\beta^2 x^2} & \text{for} \quad x \geq 0 \end{cases} \tag{2.6.17}$$

Such a random variable is called **Rayleigh Random Variable** (with parameter β) and we write such a Random Variable as $\boldsymbol{X \sim Ray(\beta)}$.

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 ∞. We can modify the Exponential CDF above as follows:

$$F_X(x) = \begin{cases} 0 & \text{for} \quad x < 0 \\ \\ \dfrac{e^{-\lambda a} - e^{-\lambda x}}{e^{-\lambda a}} & \text{for} \quad x \geq a (\text{where } \lambda > 0) \end{cases} \tag{2.6.18}$$

The corresponding pdf is given by

$$F_X(x) = \begin{cases} 0 & \text{for} \quad x < 0 \\ \\ \dfrac{\lambda e^{-\lambda x}}{e^{-\lambda a}} & \text{for} \quad x \geq a (\text{where } \lambda > 0) \end{cases} \tag{2.6.19}$$

Such a random variable is called **Pareto Random Variable** (with parameter β) and we write such a Random Variable as $\boldsymbol{X \sim Par(\beta)}$.

<u>Modified Rayleigh Random Variable</u>

We can also modify the Rayleigh distribution as follows:

$$F_X(x) = \begin{cases} 0 & \text{for} \quad x < a \\ \\ \dfrac{e^{-\beta^2 a^2} - e^{-\beta^2 x^2}}{e^{-\beta^2 a^2}} & \text{for} \quad x \geq a \text{ (where } \beta \text{ is real)} \end{cases} \tag{2.6.20}$$

with the corresponding pdf as

$$F_X(x) = \begin{cases} 0 & \text{for} \quad x < a \\ \\ \dfrac{2\beta^2 x e^{-\beta^2 x^2}}{e^{-\beta^2 a^2}} & \text{for} \quad x \geq a \text{ (where } \beta \text{ is real)} \end{cases} \tag{2.6.21}$$

For such a Random Variable we have $\mathcal{R}_X = [a, \infty)$. We write such Random Variables as $\boldsymbol{Ray(a; \beta)}$

<u>Unbounded Random Variables</u>

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 $\lim_{x\to-\infty} F_x(x) = 0$ and $\lim_{x\to+\infty} F_x(x) = 1$. We shall look at such models below:

Laplace Random Variable

The function,

$$F_X(x) = \begin{cases} \frac{1}{2}e^{\lambda x} & \text{if} \quad x < 0 \\ \\ 1 - \frac{1}{2}e^{-\lambda x} & \text{if} \quad 0 \le x < \infty \end{cases} \tag{2.6.22}$$

(where $\lambda > 0$), satisfies all the requirements above. A random variable with the above CDF is called a **Laplace Random Variable** (with parameter λ) and is denoted as $\boldsymbol{X \sim Lap(\lambda)}$. The corresponding pdf is given by

$$f_X(x) = \begin{cases} \frac{1}{2}\lambda e^{\lambda x} & \text{if} \quad x < 0 \\ \\ \frac{1}{2}\lambda e^{-\lambda x} & \text{if} \quad 0 \le x < \infty \end{cases} \tag{2.6.23}$$

or we can write this as

$$f_X(x) \;=\; \frac{1}{2}\lambda e^{-|x|} \tag{2.6.24}$$

For the Laplace Random Variable we observe the following:

$$P(X \le 0) \;=\; F_x(0) = \frac{1}{2} \tag{2.6.25}$$

Hence we see that

$$P(X \ge 0) \;=\; 1 - P(X \le 0) \tag{2.6.26}$$

$$\;=\; 1 - \frac{1}{2} = \frac{1}{2} \tag{2.6.27}$$

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 α be a real number such that $0 < \alpha < 1$. Then we can take

$$F_x(x) \;=\; \begin{cases} \alpha e^{\lambda x} & \text{if} \quad x < 0 \\ \\ 1 - (1-\alpha)e^{-\lambda x} & \text{if} \quad 0 \le x < \infty \end{cases} \tag{2.6.28}$$

satisfies all the requirements for a CDF with the corresponding pdf as

$$f_X(x) = \begin{cases} \alpha\lambda e^{\lambda x} & \text{if} \quad x < 0 \\ \\ (1-\alpha)\lambda e^{-\lambda x} & \text{if} \quad 0 \leq x < \infty \end{cases} \tag{2.6.29}$$

A Random Variable X with the above CDF satisfies

$$P(X \leq 0) = \alpha \tag{2.6.30}$$

$$P(X > 0) = 1 - \alpha \tag{2.6.31}$$

When $\alpha = \frac{1}{2}$ this reduces to the Laplace Random Variable.

Cauchy Random Variable:

The function

$$F_X(x) = \frac{1}{2} + \frac{1}{\pi} tan^{-1}\left(\frac{x}{\alpha}\right) \tag{2.6.32}$$

(where α is a positive real constant), satisfies the requirements of a CDF, with the corresponding pdf as

$$f_X(x) = \frac{\alpha}{\pi(x^2 + \alpha^2)} \tag{2.6.33}$$

A random variable with this CDF is called a **Cauchy Random Variable**, with parameter α. We denote such ramdom variables as $\boldsymbol{X \sim Cauchy(\alpha)}$. 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 $0 < \beta < 1$

$$F_X(x) = \begin{cases} \beta + \frac{2\beta}{\pi} tan^{-1}\left(\frac{x}{\alpha}\right) & \text{for} \quad x < 0 \\ \\ \beta + \frac{2(1-\beta)}{\pi} tan^{-1}\left(\frac{x}{\alpha}\right) & \text{for} \quad x \geq 0 \end{cases} \tag{2.6.34}$$

with the corresponding pdf as

$$f_X(x) = \begin{cases} 2\dfrac{\beta\alpha}{\pi(x^2+\alpha^2)} & \text{for} \quad x < 0 \\ \\ 2\dfrac{(1-\beta)\alpha}{\pi(x^2+\alpha^2)} & \text{for} \quad x \geq 0 \end{cases} \tag{2.6.35}$$

For this random variable we have

$$P(X < 0) = \beta \quad (2.6.36)$$
$$P(X > 0) = 1 - \beta \quad (2.6.37)$$

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.

2.7 Conditional Probability

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

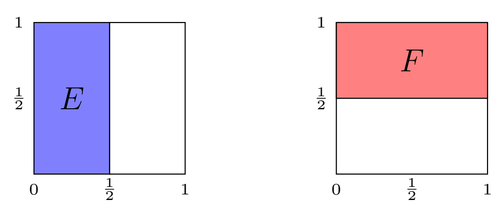
Example 2.7.1 Consider the random experiment of choosing a point at random in the unit square $0 \leq x \leq 1, 0 \leq y \leq 1$ and noting its coordinates. We then have

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

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:

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

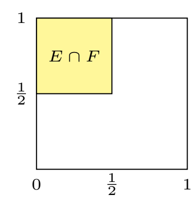
These events are sketched below:



We have

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

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



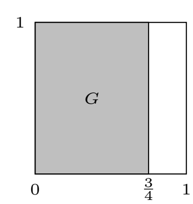
Then we have

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

Hence the proportion of F in E is given by

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

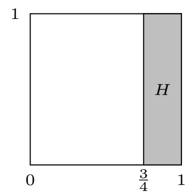
Thus the knowledge of the fact that the event E has occurred did not alter the Probability of occurrence of F. Now consider the event G sketched below:



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

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

In this case we see that 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

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

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

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

$$P(F|E) = \frac{P(F \cap E)}{P(E)} \tag{2.7.1}$$

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

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

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

Definition 2.7.2 Two events E and F are said to be **Independent** if

$$P(E \cap F) \;=\; P(E)P(F) \tag{2.7.2}$$

(This is called the product rule of the probabilities for independence)

Example 2.7.2 In the Example 2.7.1 we have E, F are independent but E, G are not independent and also E, H are not independent.

Remark 2.7.1 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 2.7.2 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:

$$\mathcal{P}\left(\bigcap_{j=1}^{n} E_j\right) \;=\; \mathcal{P}(E_1)\mathcal{P}(E_2)\cdots\mathcal{P}(E_n) \tag{2.7.3}$$

Remark 2.7.3 Note that a for collection to be independent we need that for every finite subcollection the product rule of the probabilities hold. If it fails even for one subcollection we do not have independence of the collection. For example a collection $\mathcal{C}$ of three events may be such that every pair in the collection may be independent but the collection may not be independent

2.8 Bayes' Rule

We shall next introduce an important concept useful in conditional probability computations. Consider a finite or infinite sequence of "nonempty" events

$$\Pi_1 = \{E_n\}_n$$

in the probability space $(\Omega, \mathcal{B}, P)$, such that they give a partition of Ω, that is

$$E_i \cap E_j = \phi \text{ for } i \neq j \text{ and} \tag{2.8.1}$$

$$\Omega = \bigcup_n E_n \tag{2.8.2}$$

(Such a collection of events are also referred to as "**collectively exhaustive**").
For any set $A \in \mathcal{B}$ we have

$$\begin{aligned}
A &= A \cap \Omega \\
&= A \cap \left(\bigcup_n E_n \right) \\
&= \bigcup_n (A \cap E_n) \\
\Longrightarrow & \\
P(A) &= P\left(\bigcup_n (A \cap E_n) \right) \\
&= \sum_n P\left(A \cap E_n\right) \text{ since } E_n \text{ are all disjoint} \\
&= \sum_n P(A|E_n)P(E_n)
\end{aligned}$$

Thus we have

$$P(A) = \sum_n P(A|E_n)P(E_n) \tag{2.8.3}$$

This is called the "**Law of Total Probability**".
We further have

$$P(E_n|A) = \frac{P(E_n \cap A)}{P(A)}$$

$$
\begin{aligned}
&= \frac{P(A|E_n)P(E_n)}{P(A)} \\
&= \frac{P(A|E_n)P(E_n)}{\sum_k P(A|E_k)P(E_k)} \text{ (using 2.8.3)}
\end{aligned}
$$

This is known as Bayes' Rule. Thus we have

Theorem 2.8.1 Bayes' Rule
Let $\{E_n\}_n$ be collectively exhaustive events in a probability space $(\Omega, \mathcal{B}, P)$. For any $A \in \mathcal{B}$ we have

$$
P(E_n|A) = \frac{P(A|E_n)P(E_n)}{\sum_k P(A|E_k)P(E_k)} \tag{2.8.4}
$$

Remark 2.8.1 Suppose $\{E_n\}_n$ and $\{F_n\}_n$ are two collectively exhaustive sets of events then we have from 2.8.4

$$
P(E_n|F_m) = \frac{P(F_m|E_n)P(E_n)}{\sum_k P(F_m|E_k)P(E_k)} \text{ for every } m \text{ and } n \tag{2.8.5}
$$

It is in this form that Bayes' Rule is used often in the computation of conditional probabilities.

2.9 Independence of Random Variables

The notion of independence of events induces a notion of independence of random variables. We shall first look at two discrete random variables X and Y on a probability space $(\Omega, \mathcal{B}, P)$ which take a finite number of values. Let the values taken by these two random variables be

$$
\begin{aligned}
\mathcal{R}_X &= \{x_1, x_2, x_3, \cdots, x_m\} \\
\mathcal{R}_Y &= \{y_1, y_2, y_3, \cdots, y_n\}
\end{aligned}
$$

Let us now consider the sets

$$
\begin{aligned}
E_i &= \{\omega \in \Omega : X(\omega) = x_i\} \\
F_j &= \{\omega \in \Omega : Y(\omega) = y_j\}
\end{aligned}
$$

Since X and Y are random variables the sets E_i and F_j are events. These events may or may not be independent. If these two events are independent for every i and j we say that the two random variables X and Y are independent. We can do the same thing even if the sets $\mathcal{R}_X$ or/and $\mathcal{R}_Y$ are infinite sequences. Thus we have

Definition 2.9.1 Two discrete random variables X, Y on a probability space $(\Omega, \mathcal{B}, P)$ are said to be independent if the events $\{\omega \in \Omega : X(\omega) = x_i\}$ and $\{\omega \in \Omega : Y(\omega) = y_j\}$ are independent for every $x_i \in \mathcal{R}_X$ and every $y_j \in \mathcal{R}_Y$. Using 2.7.2 we can also write this as follows:
Two discrete random variables X, Y on a probability space $(\Omega, \mathcal{B}, P)$ are said to be independent if

$$\begin{aligned} P(X = x_i \text{ and } Y = y_j) &= P(\{X = x_i\} \cap \{Y = y_j\}) \\ &= P(X = x_i) \times P(Y = y_j) \qquad (2.9.1) \end{aligned}$$

If p_X and p_Y are the pmfs of X and Y respectively then we can write the above definition as

Definition 2.9.2 Two discrete random variables X, Y on a probability space $(\Omega, \mathcal{B}, P)$ are said to be independent if

$$P(X = x_i \text{ and } Y = y_j) = p_X(x_i) \times p_Y(y_j) \qquad (2.9.2)$$

Analogously for any two general random variables we can define independence through the independence of the basic events $\{X \leq x\}$ and $\{Y \leq y\}$. We have the following

Definition 2.9.3 Two random variables X, Y on a probability space $(\Omega, \mathcal{B}, P)$ are said to be independent if

$$\begin{aligned} P(X \leq x \text{ and } Y \leq y) &= P(X \leq x) \times P(Y \leq y) \\ &= F_X(x) F_Y(y) \text{ for every } x, y \in \mathbb{R} \qquad (2.9.3) \end{aligned}$$

(where F_X and F_Y are the cdfs of X and Y respectively).

Remark 2.9.1 We say a collection $\{X_\alpha\}_{\alpha \in \mathcal{I}}$ of random variables on a probability space $(\Omega, \mathcal{B}, P)$ is independent if

$$P\left(\bigcap_{j=1}^{n} \{X_{\alpha_j} \leq x_j\}\right) = \prod_{j=1}^{n} F_{X_{\alpha_j}}(x_j)$$

for every positive integer $k \geq 2$, and for every $\alpha_j \in \mathcal{I}$ and every $x_j \in \mathbb{R}$.

2.10 Joint Distribution

In this section we shall study two or more random variables together. We shall first consider two discrete random variables. Consider two discrete random variables X and Y on a probability space $(\Omega, \mathcal{B}, \mathcal{P})$. Let

$$\begin{aligned}\mathcal{R}_X &= \{x_1, x_2, x_3 \cdots, x_m\}\\ \mathcal{R}_Y &= \{y_1, y_2, y_3, \cdots, y_n\}\end{aligned}$$

be the values taken by these random variables. Let p_X and p_Y be the pmfs of X and Y. Let

$$\begin{aligned}p_X(x_i) &= p_i \text{ for } 1 \le i \le m \text{ and}\\ p_Y(y_j) &= q_j \text{ for } 1 \le j \le n\end{aligned}$$

We now want to look at every point $\omega \in \Omega$ and analyse simulataneously X and Y at that point and we want to do this at every $\omega \in \Omega$. We do this as follows:
Let

$$\begin{aligned}E_i &= \{\omega \in \Omega : X(\omega) = x_i\}\\ F_j &= \{\omega \in \Omega : Y(\omega) = y_j\}\end{aligned}$$

Since X and Y are random variables, for every i and j, the sets E_i and F_j are events, (that is $E_i, F_j \in \mathcal{B}$), and hence the set $E_{ij} = E_i \cap F_j$ is also an event, (that is $E_{ij} \in \mathcal{B}$). We can therefore define its probability. We let

$$p_{ij} = P(E_{ij}) = P(E_i \cap F_j) = P(\omega \in \Omega : X(\omega) = x_i \text{ and } Y(\omega) = y_j)$$

Thus we get an $m \times n$ matrix

$$P_{XY} = (p_{ij})_{m\times n}$$

whose (i, j)th entry is the probability of the set of all those points where simultaneously X and Y take the values x_i and y_j respectively. Note that there may not be any point at which this simultaneous event takes place. In such a case $E_{ij} = \phi$ and $p_{ij} = P(E_{ij}) = 0$.
We shall now look at some simple examples:

Example 2.10.1 Let us consider the experiment of rolling a fair die., (with $\mathcal{B} = \mathcal{P}(\Omega)$). We have

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

Conisder the random variables X and Y defined as follows:

$$X(\omega) = \begin{cases} 1 & \text{if } \omega \text{ is even} \\ -1 & \text{if } \omega \text{ is odd} \end{cases}$$

$$Y(\omega) = \begin{cases} 1 & \text{if } \omega \text{ is a prime number} \\ -1 & \text{if } \omega \text{ is not a prime number} \end{cases}$$

The corresponding pmfs are given below:

X	p_X
1	$\frac{1}{2}$
-1	$\frac{1}{2}$

Y	p_Y
1	$\frac{1}{2}$
-1	$\frac{1}{2}$

Note that X and Y take the same set of values with the same probabilities. (Such random variables are said to be identically distributed random variables). Let us now find the joint pmf. We have

$Y \rightarrow$ $X \downarrow$	$y_1 = -1$	$y_2 = 1$
$x_1 = -1$	$\frac{1}{6}$	$\frac{2}{6}$
$x_2 = 1$	$\frac{2}{6}$	$\frac{1}{6}$

Thus the joint pmf matrix is given by

$$P_{XY} = \begin{pmatrix} \frac{1}{6} & \frac{2}{6} \\ \frac{2}{6} & \frac{1}{6} \end{pmatrix}$$

Let us look at the row sums and column sums of this matrix. We have

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

These are called the "**marginal distributions**" of X and Y. Note that the Row Sums give the pmf P_X of X and the column sums give the pmf p_Y of Y

Example 2.10.2 Let us again consider the random experiment of rolling a fair die. Consider the random variables X and Y defined as follows:

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

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:

$Y \rightarrow$ $X \downarrow$	$y_1 = -2$	$y_2 = 2$	*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	$\frac{1}{3}$	$\frac{2}{3}$	1

The joint pmf matrix P_{XY} is given by

$$P_{XY} = \begin{pmatrix} \frac{1}{6} & \frac{1}{3} \\ \frac{1}{6} & \frac{1}{3} \end{pmatrix}$$

Example 2.10.3 Consider two rvs X and Y on a probability space $(\Omega, \mathcal{B}, P)$ such that

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

with the joint pmf given by

$Y \rightarrow$ $X \downarrow$	$y_1 = -1$	$y_2 = 0$	$y_3 = 1$
$x_1 = -1$	0	$\frac{1}{4}$	0
$x_2 = 0$	$\frac{1}{4}$	0	$\frac{1}{4}$
$x_3 = 1$	0	$\frac{1}{4}$	0

From the joint pmf we can find the pmfs p_X of X and p_y of Y using the marginal distributions. For the marginal distributions we find the row and column sums. We get

$Y \rightarrow$ $X \downarrow$	$y_1 = -1$	$y_2 = 0$	$y_3 = 1$	*Row Sums*
$x_1 = -1$	0	$\frac{1}{4}$	0	$\frac{1}{4}$
$x_2 = 0$	$\frac{1}{4}$	0	$\frac{1}{4}$	$\frac{1}{2}$
$x_3 = 1$	0	$\frac{1}{4}$	0	$\frac{1}{4}$
Column Sums	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	1

Hence we get

X	p_X
$x_1 = -1$	$\frac{1}{4}$
$x_2 = 0$	$\frac{1}{2}$
$x_3 = 1$	$\frac{1}{4}$

and

Y	p_Y
$y_1 = -1$	$\frac{1}{4}$
$y_2 = 0$	$\frac{1}{2}$
$y_3 = 1$	$\frac{1}{4}$

Thus we see that we can easily find the individual pmfs from the joint pmf using the marginal distributions. However, it is not easy to get the joint pmf knowing only the individual pmfs.

We shall now see that the joint pmf is much easier to compute from the individual pmfs if the random variables are independent. Suppose X and Y are independent random variables on a probability space $(\Omega.\mathcal{B}, P)$, and

$$\begin{aligned}\mathcal{R}_X &= \{x_1, x_2, \cdots, x_m\}\\ \mathcal{R}_Y &= \{y_1, y_2, \cdots, y_n\}\end{aligned}$$

Let p_X and p_Y be the pmfs of X and Y respectively.

$$\begin{aligned}p_X(x_i) &= p_i \text{ for } 1 \leq i \leq m\\ p_Y(y_j) &= q_j \text{ for } 1 \leq j \leq n\end{aligned}$$

Then we have the joint pmf $P_{XY} = (p_{ij})_{m\times n}$ where

$$\begin{aligned}p_{ij} &= P(\omega \in \Omega : X(\omega) = x_i \text{ and } Y(\omega) = y_j)\\ &= P(\{\omega \in \Omega : X(\omega) = x_i\} \cap \{\omega \in \Omega : Y(\omega_{=} y_j\})\\ &= P(\{\omega \in \Omega : X(\omega) = x_i\}) \times P(\{\omega \in \Omega : Y(\omega_{=} y_j\})\\ &\quad \text{(by using independence)}\\ &= p_X(x_i) \times p_Y(y_j)\\ &= p_i q_j\end{aligned}$$

By marginal distribution, we have

$$\begin{aligned}p_i &= i - th\ Row\ Sum\\ q_j &= j - th\ Column\ Sum\end{aligned}$$

Thus we see that if X and Y are independent then (for every i and j), the (i, j)-th entry in the joint pmf matrix is the product of the i-th Row Sum with the j-th Column Sum. Conversely if the (i, j)-th entry in the joint pmf matrix is the product of the i-th Row Sum with the j-th Column Sum (for every i and j) then the two random variables are independent.
The random variables in Example 2.10.2 the two random variables are independent since every entry in the joint pmf matrix is the product of the corresponding Row Sum and Column Sum. The random variables of Examples 2.10.1 and 2.10.3 are not independent.
While we used the pmfs to describe the joint pmf in the case of discrete random variables, we use the cdfs to describe the joint distribution of continuous

random variables. If X and Y are continuous random variables whose cdfs are respectively

$$\begin{aligned} F_X(x) &= P(X \le x) \text{ and} \\ F_Y(y) &= P(Y \le y) \end{aligned}$$

we define their joint distribution as

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

We define the two marginal distributions as $\lim_{y\to\infty} F_{XY}(x,y)$ and $\lim_{x\to\infty} F_{XY}(x,y)$. These are the individual cdfs. Thus

$$\begin{aligned} F_X(x) &= \lim_{y\to\infty} F_{XY}(x,y) \\ F_Y(y) &= \lim_{x\to\infty} F_{XY}(x,y) \end{aligned}$$

The corresponding joint PDF, $f_{XY}(x,y)$ is defined as

$$f_{XY} = \frac{\partial^2}{\partial x \partial y} F_{XY}(x,y) \tag{2.10.1}$$

We then have

$$\begin{aligned} F_{XY}(x,y) &= \int_{-\infty}^{y}\int_{-\infty}^{x} f_{XY}(x,y)dxdy &\quad (2.10.2) \\ &= \int_{-\infty}^{x}\int_{-\infty}^{y} f_{XY}(x,y)dydx &\quad (2.10.3) \end{aligned}$$

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,

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

The marginal pdfs are given by

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)dy \tag{2.10.4}$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)dx \tag{2.10.5}$$

2.11 Expectation and Variance

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.

Consider a discrete random variable X on a Probability Space $(\Omega, \mathcal{B}, P)$, with

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

and with pmf p_X given by $p_X(x_i) = p_i$ for $1 \le i \le n$. The "**Expectation of *X***" is denoted by $E(X)$ (or $\mu_{_X}$) and is defined as the weighted average,

$$\begin{aligned} E(X) &= \sum_{i=1}^{N} x_i P(X = x_i) \\ &= \sum_{i=1}^{N} x_i p_i \end{aligned}$$

(The Expectation of a random variable X is also referred to as the "**mean**" or "**Expected value of *X***").

Analogously for a continuous random variable X the Expectation is defined through the pdf $f_{_X}(x)$ as

$$E(X) = \int_{-\infty}^{\infty} x f_{_X}(x)dx$$

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

$$E(Y) = E(g(X)) = \sum_{j=1}^{N} g(x_j)p_j \tag{2.11.1}$$

in the case of a discrete random variable X (where $\mathcal{R}_X = \{x_1, x_2, \cdots, x_N\}$), and

$$E(Y) = E(g(X)) \quad = \quad \int_{-\infty}^{\infty} g(x) f_X(x) dx \qquad (2.11.2)$$

in the case of a continuous random variable X 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 $E(X) = E(C) = C$. In particular, $E(1) = 1$

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 on the collection of all random variables on $(\Omega, \mathcal{B}, P)$, that is, if $X_1, X_2, \cdots, X_n$ are a finite number of real valued random variables on $(\Omega, \mathcal{B}, P)$ 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 2.11.1 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:

ω	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

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

Consider the random variable $Z = XY$. Then $Z(\omega) = -1$ for both $\omega = H$ and $\omega = T$. Hence Z is the constant random variable $Z = -1$ and we have $E(Z) = -1$. Note that in this case $E(XY) \neq E(X)E(Y)$

Example 2.11.2 Consider the random variables X and Y of Example 2.10.2. We had the joint pmf given by

$Y \rightarrow$ $X \downarrow$	$y_1 = -2$	$y_2 = 2$	*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	$\frac{1}{3}$	$\frac{2}{3}$	1

We have

$$\begin{aligned} E(X) &= (-1) \times (\frac{1}{2} + 1 \times (\frac{1}{2}) \\ &= 0 \\ E(Y) &= (-2) \times (\frac{1}{3}) + 2 \times (\frac{2}{3}) \\ &= \frac{2}{3} \end{aligned}$$

Now consider the random variable $U = X+Y$. We have $\mathcal{R}_U = \{-3, -1, 1, 3\}$ and the pmf of U given as

U	-3	-1	1	3
p_U	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{3}$

Hence we get

$$\begin{aligned} E(X+Y) = E(U) &= (-3) \times (\frac{1}{6}) + (-1) \times (\frac{1}{6}) + 1 \times (\frac{1}{3}) + 3 \times (\frac{1}{3}) \\ &= \frac{2}{3} \\ &= E(X) + E(Y) \end{aligned}$$

Next consider the random variable $Z = XY$. We have

$$\mathcal{R}_Z = \{-2, 2\}$$

The pmf of Z is given by

Z	-2	2
p_Z	$\frac{1}{2}$	$\frac{1}{2}$

Hence we get

$$\begin{aligned} E(XY) = E(Z) &= (-2) \times (\frac{1}{2}) + 2 \times (\frac{1}{2}) \\ &= 0 \end{aligned}$$

We see that in this case we get $E(XY) = E(X)E(Y)$ whereas in the above Example 2.11.1 we had $E(XY) \neq E(X)E(Y)$

Example 2.11.3 Consider the random variables X, Y of Example 2.10.3.

We have the joint pmf

$Y \rightarrow$ $X \downarrow$	$y_1 = -1$	$y_2 = 0$	$y_3 = 1$	*Row Sums*
$x_1 = -1$	0	$\frac{1}{4}$	0	$\frac{1}{4}$
$x_2 = 0$	$\frac{1}{4}$	0	$\frac{1}{4}$	$\frac{1}{2}$
$x_3 = 1$	0	$\frac{1}{4}$	0	$\frac{1}{4}$
Column Sums	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	1

Hence we get

$$\begin{aligned} E(X) = E(Y) &= \frac{1}{4} \times (-1) + \frac{1}{2} \times 0 + \frac{1}{4} \times 1 \\ &= 0 \end{aligned}$$

Now consider the random variable $U = X + Y$. The possible values of Z is given by

$$\mathcal{R}_U = \{-2, -1, 0, 1, 2\}$$

We have the pmf of U given as below:

U	-2	-1	0	1	2
p_U	0	$\frac{1}{2}$	0	$\frac{1}{2}$	0

We have

$$\begin{aligned} E(X+Y) = E(U) &= (-2) \times 0 + (-1) \times (\frac{1}{2}) + 0 \times 0 + 1 \times \frac{1}{2}) + 2 \times 0 \\ &= 0 \\ &= E(X) + E(Y) \end{aligned}$$

Now consider the random variable $Z = XY$. We have

$$\mathcal{R}_Z = \{-1, 0, 1\}$$

The pmf of Z is given by

Z	-1	0	-1
p_Z	0	1	0

Hence we get

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

In this example we have

$$E(XY) = E(X)E(Y)$$

We can also compute the Expectation $E(X+Y)$ and $E(XY)$ using the joint pmf as follows:
If X and Y are discrete random variables with $\mathcal{R}_X = \{x_1, x_2, \cdots, x_m\}$ and $\mathcal{R}_Y = \{y_1, y_2, \cdots, y_n\}$ with joint pmf p_{XY} as $p_{XY}(x_i, y_j) = p_{ij}$ then

$$\begin{aligned} E(X+Y) &= \sum_{i=m}^{N} \sum_{j=1}^{n} (x_i + y_j) p_{XY}(x_i, y_j) \\ &= \sum_{i=m}^{N} \sum_{j=1}^{n} (x_i + y_j) p_{ij} \end{aligned}$$

$$\begin{aligned} E(XY) &= \sum_{i=m}^{N} \sum_{j=1}^{n} (x_i y_j) p_{XY}(x_i, y_j) \\ &= \sum_{i=m}^{N} \sum_{j=1}^{n} (x_i y_j) p_{ij} \end{aligned}$$

Analogously, if X and Y are continuous random variables with joint pdf $f_{XY}(x,y)$ then

$$\begin{aligned}
E(X+Y) &= \int_0^\infty \int_0^\infty (x+y) f_{XY}(x,y)dxdy \\
E(XY) &= \int_0^\infty \int_0^\infty (xy) f_{XY}(x,y)dxdy
\end{aligned}$$

Using the linearity property of the Expectation we see that for any random variable

$$\begin{aligned}
E(X-\mu_X) &= E(X) - \mu_X \\
&= \mu_X - \mu_X \\
&= 0
\end{aligned}$$

Thus the random variable Y has $E(Y) = 0$. Hence give any random variable we can always standardize it to have mean zero by introducing the random variable $Y = X - \mu_X$.
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 2.11.1 Let X be a random variable with expectation $E(X) = \mu_X$. We then define the variation of X as the expectation of the random variable $(X - \mu_X)^2$

$$Var(X) = E((X-\mu_X)^2) \tag{2.11.3}$$

We have

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

Thus we have

$$Var(X) = E(X^2) - (E(X)^2) \tag{2.11.4}$$

We define "**standard deviation**" (denoted by σ) of the random variable as

$$\sigma_X = \sqrt{Var(X)} \tag{2.11.5}$$

Example 2.11.4 For the random variable X and Y of Example 2.11.1 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

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

Hence $\sigma_{_X} = 1 = \sigma_{_Y}$

Example 2.11.5 For the uniform random variable $X \sim Uni[a, b]$ we have the pdf

$$f_{_X}(x) = \begin{cases} \frac{1}{b-a} & \text{for} \quad a \leq x \leq b \\ 0 & \text{other} \quad \text{values of } x \end{cases}$$

Hence we get

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f_{_X}(x) dx \\ &= \int_a^b x(\frac{1}{b-a}) dx \\ &= \frac{b^2 - a^2}{2(b-a)} \\ &= \frac{b+a}{2} \end{aligned}$$

Hence we get

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

$$
\begin{aligned}
&= \frac{b^2 + a^2 - 2ab}{12} \\
&= \frac{(b-a)^2}{12}
\end{aligned}
$$

Hence we get

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

We observe the following properties of the variance:

1. We have, for any real number α,

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

   Thus we have

$$
Var(\alpha X) = \alpha^2 Var(X) \text{ for every } \alpha \in \mathbb{R} \qquad (2.11.6)
$$

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

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

   where

$$
Cov(X,Y) = E(XY) - E(X)E(Y) \qquad (2.11.7)
$$

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

$$
Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y) \qquad (2.11.8)
$$

Thus we see that in general $Var(X+Y)$ need not be equal to $Var(X)+Var(Y)$. This is because $Cov(X,Y)$ need not be zero in general. However, a "**sufficient condition**" under which this happens is that of independence of X and Y. We have for two discrete random variables X and Y, using joint pmf, $p_{XY}(x_i, y_j) = p_{ij}$,

$$\begin{aligned} E(XY) &= \sum_{i=1}^{m}\sum_{j=1}^{n} x_i y_j p_{ij} \\ &= \sum_{i=1}^{m}\sum_{j=1}^{n} x_i y_j p_i p_j \text{ (if } X \text{ and } Y \text{ are independent)} \\ &= \sum_{i=1}^{m} x_i p_i \times \sum_{j=1}^{n} y_j p_j \\ &= E(X)E(Y) \end{aligned}$$

(We can give a similar proof for the continuous case also), Thus we have

Theorem 2.11.1

X, Y are independent random variables
$\Longrightarrow$
$E(XY) = E(X)E(Y)$

Remark 2.11.1 In Example 2.11.2 we obtained $E(XY) = E(X)E(Y)$ as the two random variables are independent and hence

$$cov(X,Y) = E(XY) - E(X)E(Y) = 0$$

Remark 2.11.2 It must be stressed that the requirement of independence is only a sufficient condition for $cov(X,Y)$ to be zero. We may get

$$E(XY) - E(X)E(Y) = 0$$

even without the two random variables being independent. For instance in Example 2.11.3 we had $E(XY) = E(X)E(Y)$ but the two random variables are not independent.

2.12 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

$$F_X^T(x) \quad = \quad P(X \geq x) \tag{2.12.1}$$

If X is real valued random variable with expectation μ_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

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

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

Theorem 2.12.1 **Markov's Inequality**

If $\boldsymbol{X}$ is any nonnegative random variable, on a Probability Space $(\boldsymbol{\Omega}, \boldsymbol{\mathcal{B}}, \boldsymbol{P})$, with expectation $\boldsymbol{E(X)}$ then

$$\boldsymbol{P(X \geq k) \quad \leq \quad \frac{E(X)}{k}} \textbf{ for any } \boldsymbol{k > 0} \tag{2.12.2}$$

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 $k > 0$,

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

thus proving the inequality. Analogously we prove in the case of continuous random variables as follows:
Case 2: Continuous random variable

Let $f(x)$ be the pdf of the random variable X. Then we have

$$
\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} x f_X(x)dx \\
&= \int_{0}^{\infty} x f_X(x)dx \text{ (since } X \text{ is nonnegative)} \\
&= \int_{0}^{k} x f_X(x)dx + \int_{k}^{\infty} x f_X(x)dx \\
&\geq \int_{k}^{\infty} x f_X(x)dx \\
&\geq k\int_{k}^{\infty} f_X(x)dx \text{ (since } x \geq k \text{ in the domain of integration)} \\
&\geq kP(X \geq k) \\
\Longrightarrow & \\
P(X \geq k) &\leq \frac{E(X)}{k}
\end{aligned}
$$

Remark 2.12.1 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 2.12.2 The inequality may give some times some bizarre results even if $E(X) < \infty$, as shown in the following example.

Example 2.12.1 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 ≤ 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 2.12.3 Despite the above Remark, it must be noted that the inequality is of a very general nature in the sense that it holds 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 2.12.2 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.

II Chebychev's Inequality:
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 positive increasing function $f(t) : [0, \infty) \longrightarrow \mathbb{R}$,

$$\begin{aligned}
Y(\omega) \geq k \iff & f(Y(\omega)) \geq f(k) \\
\implies & \\
P(Y \geq k) \quad = & \quad P(f(Y) \geq f(k)) \\
\leq & \quad \frac{E(f(Y))}{f(k)} \text{ by Markov's inequality 2.12.2} \\
& \quad \text{applied to the random variable } f(X) \text{ and with } k \text{ as } f(k))
\end{aligned}$$

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

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

But

$$E(|X - \mu|^2) = E((X - \mu)^2) = Var(X) \text{ (since } X \text{ and } \mu \text{ are real)}$$

Substituting above we get

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

Thus we have

Theorem 2.12.2 Chebychev's Inequality

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

$$\boldsymbol{P(|X - \mu|) \leq \frac{Var(X)}{k^2}} \textbf{ for any real number } \boldsymbol{k > 0} \quad (2.12.3)$$

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 α 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

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

Hence we get

$$P(X \geq a) \;=\; P(Y \geq e^{\alpha k}) \tag{2.12.4}$$

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

$$P(Y \geq e^{\alpha k}) \;\leq\; \frac{E(Y)}{e^{\alpha k}} \tag{2.12.5}$$

From the above two equations we get

$$P(X \geq k) \;\leq\; \frac{E(Y)}{e^{\alpha k}} = e^{-\alpha k} E(e^{\alpha X}) \tag{2.12.6}$$

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

$$P(X \leq k) \;\leq\; \frac{E(e^{-\alpha X})}{e^{-\alpha k}} = e^{\alpha k} E(e^{-\alpha X}) \tag{2.12.7}$$

Since 2.12.6 and 2.12.7 hold whatever $\alpha > 0$ we choose we get

$$P(X \geq k) \quad \leq \quad \min_{\alpha>0} \left\{ e^{-\alpha k} E(e^{\alpha X}) \right\} \tag{2.12.8}$$

$$P(X \leq k) \quad \leq \quad \min_{\alpha>0} \left\{ e^{\alpha k} E(e^{-\alpha X}) \right\} \tag{2.12.9}$$

Equations 2.12.8 and 2.12.9 are known as Chernoff bounds.

2.13 Convergence

Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of random variables defined on a Probability Space $(\Omega, \mathcal{B}, P)$. Consider a fixed $\omega \in \Omega$. Then $X_n(\omega)$ is a real number for each n. Thus we get a sequence of real numbers $\{X_n(\omega)\}_{n\in\mathbb{N}}$. This sequence of real numbers may or may not converge. If this sequence of real numbers converges then we say that the sequence of random variables $\{X_n\}_{n\in\mathbb{N}}$ converges at the point ω. We now collect all the points in Ω at which the sequence of random variables $\{X_n\}_{n\in\mathbb{N}}$ converges. Let

$$\mathcal{C} \quad = \quad \{\omega \in \Omega : X_n(\omega) \text{ converges}\} \tag{2.13.1}$$

This is the set of points of convergence for the sequence. This is a subset of Ω and it can be shown that this subset is an event, i.e, $\mathcal{C} \in \mathcal{B}$. Hence its probability, $P(\mathcal{C})$ is well defined. The complement $\mathcal{C}'$ of the set $\mathcal{C}$, is the set of points where the sequence of random variables does not converge.

$$\mathcal{C}' \quad = \quad \{\omega \in \Omega : X_n(\omega) \text{ does not converge}\} \tag{2.13.2}$$

Since $\mathcal{C} \in \mathcal{B}$ its complement $\mathcal{C}'$ also $\in \mathcal{B}$ and hence its probability $P(\mathcal{C}')$ is defined. Note that $\mathcal{C} \cup \mathcal{C}' = \Omega$ and hence

$$P(\mathcal{C}) + P(\mathcal{C}') \quad = \quad 1$$

The set $\mathcal{C}$ is the "good set" where there is no problem about convergence and the set $\mathcal{C}'$ is the "bad set" where there is problem about convergence. If the "bad set" is negligible, that is, if $P(\mathcal{C}') = 0$ which is the same as saying $P(\mathcal{C}) = 1$, then we have convergence essentially everywhere except that where there is no convergence is a negligible set. We then say that the sequence of random variables $\{X_n\}_{n\in\mathbb{N}}$ converges almost surely. We then have the formal definition,

Definition 2.13.1 A sequence of real valued random variables $\{X_n\}_{n\in\mathbb{N}}$ on a Probability Space $(\Omega, \mathcal{B}, P)$ is said to converge "almost surely" if

$$P(\omega \in \Omega : X_n(\omega) \text{ converges}) \;=\; 1$$

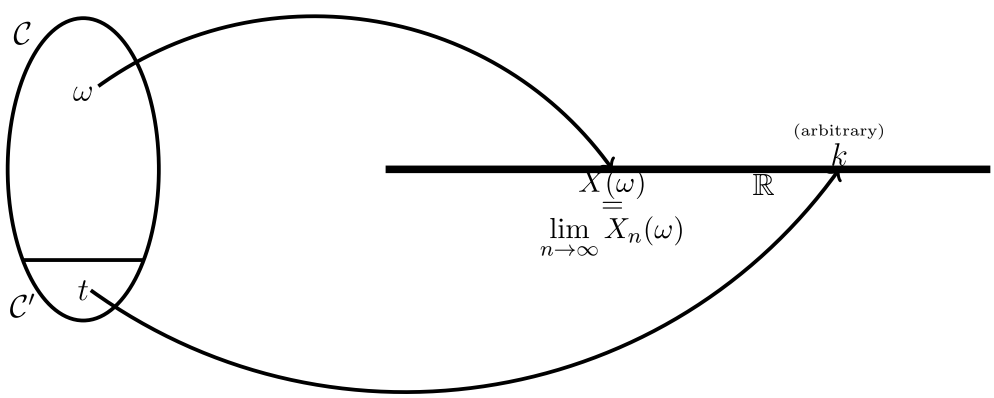
or equivalently

$$P(\omega \in \Omega : X_n(\omega) \text{ does not converge}) \;=\; 0$$

Suppose the sequence converges almost surely. Then for every $\omega \in \mathcal{C}$ the sequence converges to a real number. This real number depends on ω. We denote this limit by $X(\omega)$. We have

$$\lim_{n\to\infty} X_n(\omega) \;=\; X(\omega) \text{ for every } \omega \in \mathcal{C}$$

Thus we have a function $X(\omega)$ defined for $\omega \in \mathcal{C}$. We now define this function arbitrarily on $\mathcal{C}'$ - say- $X(\omega) = k$ for every $\omega \in \mathcal{C}'$, (where k is any fixed real number).. Then we have a function X defined from Ω to $\mathbb{R}$.



This function is a random variable on the Probability Space $(\Omega, \mathcal{B}, P)$ and we have

$$P(\omega \in \Omega : X_n(\omega) \text{ converges to } X(\omega)) \;=\; 1$$

We say that the sequence X_n converges to X almost surely. Thus we have

Definition 2.13.2 A sequence of real valued random variables $\{X_n\}_{n\in\mathbb{N}}$ on a Probability Space $(\Omega, \mathcal{B}, P)$ is said to converge "almost surely" to a random variable X on $(\Omega, \mathcal{B}, P)$ if

$$P(\omega \in \Omega : X_n(\omega) \text{ converges to } X(\omega)) = 1$$

or equivalently

$$P(\omega \in \Omega : X_n(\omega) \text{ does not converge to } X(\omega)) = 0$$

We then write

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

Remark 2.13.1 Since we have defined X arbitrarily at points in $\mathcal{C}'$ the limit is not unique at the outset. But we use the following convention:
If $f : \Omega \longrightarrow \mathbb{R}$ and $g : \Omega \longrightarrow \mathbb{R}$ are two real valued random variables on $(\Omega, \mathcal{B}, P)$ such that

$$P(\omega \in \Omega : f(\omega) \neq g(\omega)) = 0$$

or equivalently

$$P(\omega \in \Omega : f(\omega) = g(\omega)) = 1$$

we treat the two functions as same. With this convention it follows that the almost sure limit when it exists is unique (since any two limits will differ only on the set $\mathcal{C}'$ which has probability zero).

We shall illustrate this notion of almost sure convergence with a simple example.

Example 2.13.1 Consider the experiment of tossing a fair coin. We have $\Omega = \{H, T\}$ and $P(H) = P(T) = \frac{1}{2}$. Consider the sequence $\{X_n\}_{n\in\mathbb{N}}$ of random variables defined on this Probability Space as follows:

$$X_n(\omega) = \begin{cases} \dfrac{n}{n+1} & \text{if } \omega = H \\ \\ (-1)^n & \text{if } \omega = T \end{cases}$$

We see that $X_n(H)$ converges to 1 and $X_n(T)$ does not converge. Hence $\mathcal{C} = \{H\}$ and $\mathcal{C}' = \{T\}$ and consequently $P(\mathcal{C}) = P(\mathcal{C}') = \dfrac{1}{2}$. Since $P(\mathcal{C}) \neq 1$ it follows that this sequence does not converge almost surely.

Example 2.13.2 In the above experiment consider the sequence defined as follows:

$$X_n(\omega) \;=\; \begin{cases} \dfrac{n}{n+1} & \text{if} \quad \omega = H \\[2ex] \dfrac{(-1)^n}{n} & \text{if} \quad \omega = T \end{cases}$$

We see that $X_n(H)$ converges to 1 and $X_n(T)$ converges to 0. Hence $\mathcal{C} = \{H, T\}$ and consequently $P(\mathcal{C}) = 1$ and hence the sequence converges almost surely. Further, if we define X as the random variable

$$X(\omega) \;=\; \begin{cases} 1 & \text{if} \quad \omega = H \\ 0 & \text{if} \quad \omega = T \end{cases}$$

then we see that

$$\{\omega \in \Omega : X_n(\omega) \text{ converges to } X(\omega)\} \;=\; \Omega$$

and hence

$$P(\omega \in \Omega : X_n(\omega) \text{ converges to } X(\omega)) \;=\; 1$$

Hence we get

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

We shall next look at the convergence from a different viewpoint.
When we want to analyse convergence of a sequence of random variables X_n to a random variable X we look at the difference $X_n - X$ and estimate the error $|X_n - X|$. Convergence of X_n to X should mean that this error can be made as small as we want for large values of n. Let us look at a tolerance error $\varepsilon > 0$. Let us look at the n-th stage. If the error at a point $\omega \in \Omega$ at the n-th stage is $< \varepsilon$ then this point is not a problem point for the error ε at the n-th stage. So the "bad points" for the error ε at the n-th stage are those points where the error is $\geq \varepsilon$. Let $B_n(\varepsilon)$ denote the set of all such "bad" points. We have

$$B_n(\varepsilon) \;=\; \{\omega \in \Omega : |X_n(\omega) - X(\omega)| \geq \varepsilon\}$$

as the set of all bad points for the error ε at the n-th stage. $B_n(\varepsilon)$ is a subset of Ω and it is easy to see that this is an event, that is, $B_n(\varepsilon) \in \mathcal{B}$ and hence its probability $P(B_n(\varepsilon))$ is defined. This probability gives us an idea as to how "significant" is this bad set. If the significance of this set diminishes as n increases then we have some control over the bad sets. More precisely if

$$\lim_{n\to\infty} P(B_n(\varepsilon)) = 0$$

then we have control over the bad sets with respect to the error ε. We would like this control to happen for every $\varepsilon > 0$, that is, we want

$$\text{For every } \varepsilon > 0,\ \lim_{n\to\infty} P(B_n(\varepsilon)) = 0$$

If this happens we have some sort of control over the sets where the error goes beyond a prescribed $\varepsilon > 0$. This type of convergence is called convergence in probability. We have

Definition 2.13.3 A sequence, $\{X_n\}_{n\in\mathbb{N}}$, of random variables on a Probability Space $(\Omega, \mathcal{B}, P)$ is said to converge "in probability" to a random variable X on $(\Omega, \mathcal{B}, P)$ if

$$\lim_{n\to\infty} P(\omega \in \Omega : |X_n(\omega) - X(\omega)| \geq \varepsilon) \;=\; 0 \text{ for every } \varepsilon > 0$$

We then write

$$X_n \xrightarrow{p} X$$

Example 2.13.3 Consider the sequence of random variables of Example 2.13.2. We had the sequence of random variables defined as

$$X_n(\omega) \;=\; \begin{cases} \dfrac{n}{n+1} & \text{if} \quad \omega = H \\[2ex] \dfrac{(-1)^n}{n} & \text{if} \quad \omega = T \end{cases}$$

We found that this sequence converges almost surely to the random variable defined as

$$X(\omega) \;=\; \begin{cases} 1 & \text{if} \quad \omega = H \\ 0 & \text{if} \quad \omega = T \end{cases}$$

Let us now examine whether this sequence converges in probability to X. We have

$$\begin{aligned} |X_n(H) - X(H)| &= \left|\frac{n}{n+1} - 1\right| \\ &= \frac{1}{n} \end{aligned}$$

Similarly we have

$$\begin{aligned} |X_n(T) - X(T)| &= \left| \frac{(-1)^n}{n} - 0 \right| \\ &= \frac{1}{n} \end{aligned}$$

Thus we have

$$|X_n(\omega) - X(\omega)| = \frac{1}{n} \text{ for every } \omega \in \Omega$$

For any $\varepsilon > 0$ we can find N_ε such that

$$\frac{1}{N_\varepsilon} < \varepsilon$$

and hence

$$n \geq N_\varepsilon \Longrightarrow \frac{1}{n} \leq \frac{1}{N_\varepsilon} < \varepsilon$$

Thus we have for all $n \geq N_\varepsilon$

$$|X_n(\omega) - X(\omega)| < \varepsilon \text{ for all } \omega \in \Omega$$

and hence

$$\begin{aligned} \{\omega \in \Omega : |X_n(\omega) - X(\omega)| \geq \varepsilon\} &= \phi \text{ for all } n \geq N_\varepsilon \\ &\Longrightarrow \\ P(\omega \in \Omega : |X_n(\omega) - X(\omega)| \geq \varepsilon) &= 0 \text{ for all } n \geq N_\varepsilon \\ &\Longrightarrow \\ \lim_{n \to \infty} P(\omega \in \Omega : |X_n(\omega) - X(\omega)| \geq \varepsilon) &= 0 \text{ for every } \varepsilon > 0 \end{aligned}$$

Hence $X_n \xrightarrow{p} X$

Remark 2.13.2 If $X_n \xrightarrow{p} X$ then the limit is unique in the sense introduced in Remark 2.13.1

We shall next look at convergence from the point of view of certain features/parameters converging. We first introduce the notion of convergence in distribution. Since a lot of information about a random variable is contained in its cdf, we can see whether the cdfs of the random variables X_n in the sequence converge to the cdf of a random variable X. Since the cdf of X may have points of discontinuity where there may be problems of convergence we demand convergence only at points of continuity of the cdf of X. More precisely we have the following definition:

Definition 2.13.4 A sequence of real valued random variable $\{X_n\}_{n\in\mathbb{N}}$ on a Probability Space $(\Omega, \mathcal{B}, P)$ is said to converge to a random variable X on $(\Omega, \mathcal{B}, P)$ if

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

at every point of continuity of F_X

We then write $X_n \xrightarrow{d} X$

Remark 2.13.3 Since two different random variables X and Y can be identically distributed, that is, they are such that $F_X(x) = F_Y(x)$ it follows that if $X_n \xrightarrow{d} X$ then we also have $X_n \xrightarrow{d} Y$. Thus it follows that the convergence in distribution does not give unique limit.

Example 2.13.4 Let X and Y be two random variables on a Probability Space $(\Omega, \mathcal{B}, P)$ which are identically distributed. Consider the sequence X_n where

$$X_n = \begin{cases} X & \text{if} \quad n \text{ is odd} \\ Y & \text{if} \quad n \text{ is even} \end{cases}$$

The sequence becomes $X, Y, X, Y, X, Y, \cdots$. Clearly this sequence converges in distribution to both X and Y since $F_{X_n}(x) = F_X(x) = F_Y(x)$ for all x and for all n. For instance in the experiment of rolling a fair die consider the random variables,

$$X(\omega) = \begin{cases} 1 & \text{if} \quad \omega \text{ is odd} \\ -1 & \text{if} \quad n \text{ is even} \end{cases}$$
$$Y(\omega) = \begin{cases} -1 & \text{if} \quad \omega \text{ is odd} \\ 1 & \text{if} \quad n \text{ is even} \end{cases}$$

Then X and Y are identically distributed and the sequence $X, Y, X, Y, X, Y, \cdots$ converges in distribution to both X and Y. Note that the sequence $X, X, X, X, \cdots$ converges to both X and Y in distribution

We can also look at other features such as for instance the centralised moments. We have the following definitions:

Definition 2.13.5 A sequence of random variables X_n on a Probability Space $(\Omega, \mathcal{B}, P)$ is said to converge in

1. **mean to a random variable X on $(\Omega, \mathcal{B}, P)$ if**

$$\lim_{n\to\infty} E(|X_n - X|) = 0$$

   We then write

$$X_n \xrightarrow{m} X$$

2. **quadratic mean to a random variable X on $(\Omega, \mathcal{B}, P)$ if**

$$\lim_{n\to\infty} E(|X_n - X|^2) = 0$$

   We then write

$$X_n \xrightarrow{q.m} X$$

3. **k mean to a random variable X on $(\Omega, \mathcal{B}, P)$ if**

$$\lim_{n\to\infty} E(|X_n - X|^k) = 0$$

   (where k is a positive integer).
   We then write

$$X_n \xrightarrow{k.m} X$$

We shall now look at some examples of convergence.

Example 2.13.5 Consider a sequence of random variables $X_n \sim Exp(-n)$. Then we have

$$F_{X_n}(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - exp(-nx) & \text{if } x \geq 0 \end{cases}$$

We observe that for $x \neq 0$ we have

$$\lim_{n\to\infty} F_{X_n}(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x > 0 \end{cases}$$

Thus $F_{X_n}(x)$ converges to the cdf $F_X(x)$ where X is the zero random variable. Hence we have

$$X_n \xrightarrow{d} 0$$

Let us now examine whether the convergence is also in probability. We have for any $\varepsilon > 0$,

$$\begin{aligned}
P(|X_n - X| \geq \varepsilon) &= P(X_n \geq \varepsilon) \text{ (since } X = 0 \text{ and } X_n \geq 0) \\
&= 1 - P(X_n < \varepsilon) \\
&= 1 - F_{X_n}(\varepsilon) \\
&= 1 - (1 - exp(-n\varepsilon)) \\
&= exp(-n\varepsilon)
\end{aligned}$$

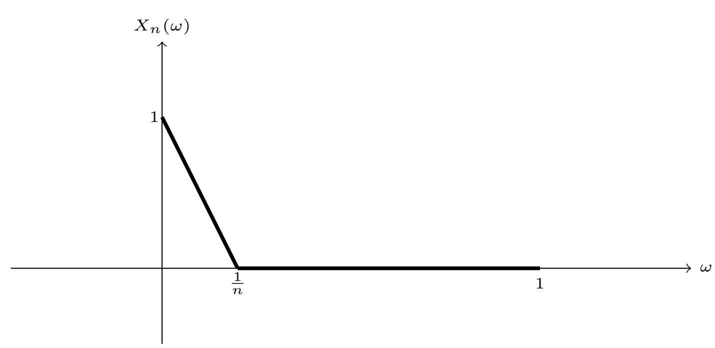
Hence we have

$$\begin{aligned}
\lim_{n\to\infty} P(|X_n - X| \geq \varepsilon) &= \lim_{n\to\infty} exp(-n\varepsilon) \\
&= 0 \text{ for every } \varepsilon > 0
\end{aligned}$$

Hence we get

$$X_n \xrightarrow{p} 0$$

Example 2.13.6 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:



Convergence almost surely:

Let us first examine the almost sure convergence of this sequence. Let ω

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. Further $X_n(0) = 1$ for all n and hence $X_n(0) \longrightarrow 1$. Let X be the zero random variable. 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

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

Convergence in Probability

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 ω 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,

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

Hence we get

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

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

$$X_n \xrightarrow{p} X$$

<u>Convergence in Distribution</u>

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

$$X_n \xrightarrow{d} X$$

2.14 Law of Large Numbers and Central Limit Theorem

In this section we shall state three theorems (without proof) regarding the long time behaviour of the sequence of averages of independent samples of a random variables. The first two theorems called "Law of Large Numbers" state that the "sample mean converges to the true mean" in a suitable sense. There are two versions of the theorem the only difference between the two versions being about the type of convergence. More precisely we have the following two theorems:
Let X_n be a sequence of independent identically distributed real valued random variables on a Probability Space $(\Omega, \mathcal{B}, P)$, with finite mean μ. Let Y_n be the sample mean, that is,

$$Y_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

The theorems assert that this sequence Y_n eventually approaches the constant random variable μ. We have

Theorem 2.14.1 <u>Weak Law of Large Numbers</u>

Let X_n be a sequence of independent identically distributed real valued random variables on a Probability Space $(\Omega, \mathcal{B}, P)$, with finite mean μ. Let Y_n be the sample mean, that is,

$$Y_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

Then

$$Y_n \xrightarrow{p} \mu$$

The next theorem states that this convergence is even stronger, that is, almost sure convergence.

Theorem 2.14.2 <u>Strong Law of Large Numbers</u>

Let X_n be a sequence of independent identically distributed real valued random variables on a Probability Space $(\Omega, \mathcal{B}, P)$, with finite mean μ. Let Y_n be the sample mean, that is,

$$Y_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

Then

$$Y_n \xrightarrow{a.s} \mu$$

The next theorem is about a sequence of standardized real valued random variables.
Consider a sequence X_n of independent identically distributed real valued random variables with finite mean μ and finite variance σ^2. We first look at the sample mean.

$$Y_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

Then Y_n all have mean μ. We first centre these at 0, that is we make the mean zero by a shift as follows: Let

$$U_n = Y_n - \mu$$

Then

$$E(U_n) = 0$$

Further

$$\begin{aligned}
Var(U_n) &= Var(Y_n - \mu) \\
&= Var(Y_n) \\
&= Var\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) \\
&= \frac{1}{n^2} Var(X_1 + X_2 + \cdots + X_n) \\
&= \frac{1}{n^2}(Var(X_1) + Var(X_2) + \cdots + Var(X_n)) \text{ (using independence)} \\
&= \frac{1}{n^2} \times (n\sigma^2) \\
&= \frac{\sigma^2}{n} \\
\Longrightarrow & \\
\frac{n}{\sigma^2} Var(U_n) &= 1 \\
\Longrightarrow & \\
Var\left(\frac{\sqrt{n}\, U_n}{\sigma}\right) &= 1
\end{aligned}$$

Thus the random variable

$$Z_n = \frac{\sqrt{n}\left(\frac{X_1+X_2+\cdots+X_n}{n} - \mu\right)}{\sigma}$$

with mean 0 and 1. The Central Limit Thoeorem asserts that this sequence of samples standardized to mean 0 and variance 1 converge to the standard Normal Distribution $N(0,1)$. We have

Theorem 2.14.3 Let X_n be a sequence of independent identically distributed real valued random variables with finite mean μ and finite variance σ^2. Let Z_n be the standardised sample mean

$$Z_n = \frac{\sqrt{n}\left(\frac{X_1+X_2+\cdots+X_n}{n} - \mu\right)}{\sigma}$$

Then

$$Z_n \xrightarrow{d} N(0,1)$$