IC105: Probability and Statistics **2021-22-M**

Lecture 0: Lecture Notes

Instructor: Dr. Kuldeep Kumar Kataria *Scribe:*

These are the updated notes till Lecture #11

0.1. Introduction

Example 0.1. *The production manager of a bulb manufacturing company wishes to study the effect of new manufacturing process on the lifetimes of bulbs produced through it.*

Here the population under study is the following:

$\mathcal{P}$: Collection of lifetimes of all electric bulbs produced using new manufacturing process.

In most practical situation $\mathcal{P}$ is generally large (e.g. collection of lifetimes of all electric bulbs that would be produced using new manufacturing process) and it is not (due to time/cost contraints) to get complete information about $\mathcal{P}$. Thus a representative sample (a sample that in certain sense is a true representative of the population) is taken from $\mathcal{P}$ and using this representative sample inferences regarding various population characteristics of $\mathcal{P}$ (such as population mean, population variance etc.) are made. Note that the sample contains only partial information about $\mathcal{P}$ and the goal is to make inferences about various population characteristics based on partial information in the sample drawn from $\mathcal{P}$.

X: Lifetime of a typical electric bulbs manufactured using new manufacturing process (a typical element of $\mathcal{P}$).

X is random (called a random variable) and its value varies across $\mathcal{P}$ according to some law.

Probability Theory: A mathematical tool for modelling uncertainty (*e.g.* to describe the law according to which values of X vary across $\mathcal{P}$).

Statistics: Concerns with procedures for analyzing data (sample) and drawing inferences about various characteristics of the population $\mathcal{P}$.

For understanding of statistics, one must have a sound background in probability theory.

The only way to collect information about any random phenomenon is to perform experiments (*e.g.* selecting a set of bulbs manufactured by the new manufacturing process and putting them on test for measuring their lifetimes). Each experiment terminates in an outcome which cannot be predicted in advance prior to the performance of experiment (*e.g.* lifetimes of the bulbs put on test cannot be predicted before they are put on test).

Definition 0.2 (Random Experiment)**.** *A random experiment is an experiment in which*

(a) all possible outcomes of the experiment are known in advance,

(b) outcome of a particular performance (trial) of the experiment cannot be predicted in advance,

(c) the experiment can be repeated under identical conditions.

We will generally denote a random experiment by $\mathcal{E}$.

Definition 0.3 (Sample Space)**.** *The collection of all possible outcomes of a random experiment is called its sample space. A sample space will usually be denoted by Ω.*

Example 0.4. *(i) $\mathcal{E}$: Tossing a coin once. Sample space $\Omega = \{H, T\}$, where H: Heads and T: Tails.*

(ii) $\mathcal{E}$: Throwing a die. Sample space $\Omega = \{1, 2, 3, 4, 5, 6\}$.

(iii) $\mathcal{E}$: Birth of a child. Sample space $\Omega = \{M, F\}$. If we consider his/her weight then $\Omega = (0, 7)$.

(iv) $\mathcal{E}$: Age at the death of a person. Sample space $\Omega = (0, 120)$.

(v) $\mathcal{E}$: Putting an electric bulbs produced by new manufacturing process into test and measuring its lifetime. Sample space $\Omega = [0, \infty)$.

(vi) $\mathcal{E}$: Throwing two dice. Sample space

$$\begin{aligned}\Omega &= \{(1,1),(1,2),\ldots,(1,6),(2,1),(2,2),\ldots,(2,6),\ldots,(6,1),(6,2),\ldots,(6,6)\}\\ &= \{(i,j) : i,j \in \{1,2,\ldots,6\}\}.\end{aligned}$$

(vii) $\mathcal{E}$: Putting two electric bulbs produced by new manufacturing process into test and measuring their lifetimes. Sample space $\Omega = \{(x_1, x_2) : x_1 \geq 0,\ x_2 \geq 0\} = [0, \infty) \times [0, \infty)$.

(viii) $\mathcal{E}$: Casting one red die and white die. Sample space

$$\begin{aligned}\Omega &= \{(r,w) : r \textit{ is number of spots on the red die and } w \textit{ is number of spots on the white die }\}\\ &= \{(1,1),(1,2),\ldots,(1,6),(2,1),(2,2),\ldots,(2,6),\ldots,(6,1),(6,2),\ldots,(6,6)\}\\ &= \{(i,j) : i,j \in \{1,2,\ldots,6\}\}\\ &= \{1,2,\ldots,6\} \times \{1,2,\ldots,6\} \to \textit{has 36 elements}.\end{aligned}$$

Definition 0.5 (Event)**.** *An event is any subset of the sample space. If the outcome of a random experiment is a member of the set $E \subseteq \Omega$, we say that event E has occured.*

Example 0.6. *In Example 0.4 (vi), $A = \{(1,5),(6,2),(2,2)\}$ is an event. Also, in Example 0.4 (vii), $A = \{(x_1, x_2) : x_1 \leq 6,\ x_2 \geq 8\} = [0, 6] \times [8, \infty)$ may be an event.*

Impossible Event: ϕ.

Sure Event: Ω.

Exhaustive Events: If $\bigcup_{i=1}^{n} A_i = \Omega$ then we call $A_1, A_2, \ldots, A_n$ to be exhaustive events.

Mutually Exclusive Events: If $A \cap B = \phi$ then A and B are called mutually exclusive events *i.e.*, happening or occcurrence of one of them excludes the possiblity of occurrence of other.

Pairwise Disjoint Events: Let $A_1, A_2, \ldots$ be events such that $A_i \cap A_j = \phi,\ i \neq j$. Then, we say that $A_1, A_2, \ldots$ are pairwise disjoint or mutually exclusive.

Let A and B be two events. Then,

(i) $A \cup B \to$ occurrence of at least one of the event A and B.

(ii) $\bigcup_{i=1}^{\infty} A_i \to$ occurrence of at least one $A_i,\ i = 1, 2, \ldots, n$.

(iii) $A \cap B \to$ simultaneous occurrence of A and B.

(iv) $\bigcap_{i=1}^{\infty} A_i \to$ simultaneous occurrence of $A_i,\ i = 1, 2, \ldots, n$.

(v) $A^c \rightarrow$ not happenning of A.

(vi) $A - B \rightarrow$ happenning of A not B. Thus, $A - B = A \cap B^c$.

Generally, we are interested in specific subsets of Ω, called event. So the event space (events under consideration) $\mathcal{F}$ is a subset of power set of Ω.

So the event space is $\mathcal{F} \subseteq \mathcal{P}(\Omega)$. Here, $\mathcal{P}(\Omega)$ denotes the power set of Ω. In many situations and for almost all practical purposes $\mathcal{F} = \mathcal{P}(\Omega)$.

The choice of $\mathcal{F}$ is an important one:

(i) If Ω contains atmost a countable number of points we can always take $\mathcal{F}$ to be the $\mathcal{P}(\Omega)$. (This is certainly a σ-field). In this case each point set is a member of $\mathcal{F}$ and is the fundamental object of interest. Every subset of Ω is an event.

(ii) If $\Omega = \mathbb{R}$ or any interval then Ω is uncountable. In this case we would like to consider all one point subsets of Ω, all intervals (closed, open or semi-closed) to be events. We consider the Borel σ-field $\mathcal{B}$ generated by the class of all semi closed intervals $(a, b]$, which is a σ-field in $\mathbb{R}$.

We say that the event space $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ contains all subsets of Ω actually encountered in ordinary analysis and probability. It is large enough for all practical purposes.

Definition 0.7 (σ-Field/ σ-Algebra)**.** *A class of subset $\mathcal{F}$ of the sample space Ω is called a σ-field if it satisfies the following conditions:*

(a) $\Omega \in \mathcal{F}$,

(b) If $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$,

(c) If $A_1, A_2, \cdots \in \mathcal{F}$ then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

The algebra of set theory is applicable in probability theory. Probability is a measure of uncertainty. We are interested in quantifying uncertainty associated with various outcomes of a random experiment by assigning probability to these outcomes.

Here, we will not discuss how probabilities are assigned (which is a part of probability modelling) rather we will discuss properties of a probability as a measure.

0.2. Probability Measure

Recall that $\mathcal{E}$ denotes a random experiment, Ω denotes the sample space of $\mathcal{E}$ and $\mathcal{F}$ denotes event space. For all practical purposes one may take $\mathcal{F} = \mathcal{P}(\Omega)$.

A set function is a function whose domain is a collection of sets (called a class of sets).

Definition 0.8 (Probability Function or Probability Measure)**.** *A probability function (or probability measure) is a real valued set function, defined on the event space $\mathcal{F}$ satisfying the following axioms:*

(a) $P(\Omega) = 1$ (certainty),

(b) $P(A) \geq 0 \;\; \forall \; A \in \mathcal{F}$ (positivity),

(c) If $A_1, A_2 \in \mathcal{F}$ be mutually exclusive/disjoint sets (i.e. $A_1 \cap A_2 = \phi$, the empty set) then

$$P(A_1 \cup A_2) = P(A_1) + P(A_2).$$

More generally, if $\{A_n\}_{n\geq 1}$ is a sequence of mutually exclusive (disjoint) sets in $\mathcal{F}$ i.e., $A_i \cap A_j = \phi,\ i \neq j$, then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) \quad (\textit{countable additivity}).$$

We call $P(A)$ the probability of event A. The triplet $(\Omega, \mathcal{F}, P)$ is called probability space.

Remark 0.9. *Axiom (b) and (c) are desirable for any measure (such as area, volume, probability etc.). Since the sample space Ω consists of all possible outcomes its occurrence is certain (100% chance of occurrence) and therefore Axiom (a) ($P(\Omega) = 1$) is also reasonable.*

Elementary Properties of Probability Function/ Measure:

Let $(\Omega, \mathcal{F}, P)$ be a probability space.

(P1) $P(\phi) = 0$.

Proof. Let $A_1 = \Omega$ and $A_i = \phi, i = 2, 3, \ldots$ Also, we have $A_1 = \bigcup\limits_{i=1}^{\infty} A_i,\ A_i \cap A_j = \phi, \forall\, i \neq j$. Therefore,

$$\begin{aligned}
P(\Omega) &= P\left(\bigcup_{i=1}^{\infty} A_i\right) \\
\implies 1 &= \sum_{i=1}^{\infty} P(A_i), \quad \text{(Axioms (a) and (c))} \\
\implies 1 &= \lim_{n\to\infty} \sum_{i=1}^{n} P(A_i) \\
\implies 1 &= \lim_{n\to\infty} [P(\Omega) + (n-1)P(\phi)] \\
\implies 1 &= 1 + \lim_{n\to\infty} [(n-1)P(\phi)] \\
\implies P(\phi) &= 0.
\end{aligned}$$

This completes the proof. □

(P2) For some natural number n, let $A_1, A_2, \ldots, A_n \in \mathcal{F}$ be mutually exclusive. Then, $P\left(\bigcup\limits_{i=1}^{n} A_i\right) = \sum\limits_{i=1}^{n} P(A_i)$.

Proof. Let $A_i = \phi,\ i = n+1, n+2, \ldots$ Then $A_i \cap A_j = \phi, \forall\, i \neq j$ and $\bigcup\limits_{i=1}^{n} A_i = \bigcup\limits_{i=1}^{\infty} A_i$. This implies

$$\begin{aligned}
P\left(\bigcup_{i=1}^{n} A_i\right) &= P\left(\bigcup_{i=1}^{\infty} A_i\right) \\
&= \sum_{i=1}^{\infty} P(A_i), \quad \text{(Axioms (c))} \\
&= \sum_{i=1}^{n} P(A_i), \quad (P(A_i) = P(\phi) = 0, \forall\, i = n+1, n+2, \ldots).
\end{aligned}$$

This completes the proof. □

(P3) For all $A \in \mathcal{F}$, $0 \leq P(A) \leq 1$ and $P(A^c) = 1 - P(A)$.

Proof. Note that $\Omega = A \cup A^c$ and $A \cap A^c = \phi$. Therefore,

$$1 = P(\Omega) = P(A \cup A^c) = P(A) + P(A^c) \geq P(A), \quad \text{(using Axioms (a), (b) and } (\mathbf{P2})).$$

Thus, $0 \leq P(A) \leq 1$ and $P(A^c) = 1 - P(A)$. □

(P4) Let $A_1, A_2 \in \mathcal{F}$ be such that $A_1 \subseteq A_2$. Then, $P(A_2 - A_1) = P(A_2) - P(A_1)$ and $P(A_1) \leq P(A_2)$.

Proof. $A_2 = A_1 \cup (A_2 - A_1)$ and $A_1 \cap (A_2 - A_1) = \phi$. Thus,

$$P(A_2) = P(A_1) + P(A_2 - A_1) \implies P(A_2 - A_1) = P(A_2) - P(A_1).$$

By Axiom (b), we have $P(A_2 - A_1) \geq 0 \implies P(A_2) \geq P(A_1)$, that is, $P(\cdot)$ is monotone. □

(P5) Let $A_1, A_2 \in \mathcal{F}$. Then,

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) \text{ (Inclusion-Exclusion principle for two events).}$$

Proof. Note that $A_1 \cup A_2 = A_1 \cup (A_2 - A_1)$ and $A_1 \cap (A_2 - A_1) = \phi$. This implies

$$P(A_1 \cup A_2) = P(A_1 \cup (A_2 - A_1)) = P(A_1) + P(A_2 - A_1) \quad \text{(using } (\mathbf{P2})). \tag{0.1}$$

Also, we have

$$(A_1 \cap A_2) \cap (A_2 - A_1) = \phi \text{ and } A_2 = (A_1 \cap A_2) \cup (A_2 - A_1),$$

which implies

$$\begin{aligned} P(A_2) &= P(A_1 \cap A_2) + P(A_2 - A_1) \\ \implies P(A_2 - A_1) &= P(A_2) - P(A_1 \cap A_2). \end{aligned} \tag{0.2}$$

Using (0.2) in (0.1), we get

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2).$$

This completes the proof. □

Remark 0.10. *(a) If $P(A) = 0$ and $B \subseteq A$, then $P(B) = 0$ (using (**P4**) and Axiom (b)).*

*Similarly, if $P(C) = 1$ and $C \subseteq D$, then $P(D) = 1$ (using (**P3**) and (**P4**)).*

*(b) **Exercise**: If $P(D) = 1$, then $P(A) = P(A \cap D), \forall A \in \mathfrak{F}$.*

Similarly, if $P(D) = 0$, then $P(A) = P(A \cap D^c), \forall A \in \mathfrak{F}$.

*(c) Let $A_1, A_2 \in \mathcal{F}$. Then, using (**P5**) and Axiom (b), we get*

$$P(A_1 \cup A_2) \leq P(A_1) + P(A_2) \text{ (Boole's inequality for two events).}$$

*(d) Let $A_1, A_2 \in \mathcal{F}$. Then, using (**P3**), (**P5**) and Axiom (b), we get*

$$P(A_1 \cap A_2) \geq \max\{P(A_1) + P(A_2) - 1, 0\} \text{ (Bonferroni's inequality for two events).}$$

Theorem 0.11 (Inclusion-Exclusion Principle)**.** *For $A_1, A_2, \ldots, A_k \in \mathcal{F}$, ($k \geq 2$ is an integer), let*

$$p_{1,k} = P(A_1) + P(A_2) + \cdots + P(A_k) = \sum_{i=1}^{k} P(A_i)$$

$$\begin{aligned} p_{2,k} &= P(A_1 \cap A_2) + P(A_1 \cap A_3) + \cdots + P(A_1 \cap A_k) + P(A_2 \cap A_3) + \cdots + P(A_2 \cap A_k) + \cdots + P(A_{k-1} \cap A_k) \\ &= \sum\sum_{1 \leq i < j \leq k} P(A_i \cap A_j) \end{aligned}$$

(*sum of probabilities of all possible intersections involving 2 events out of the k events $A_1, \ldots, A_k$*)

$$\vdots$$

$$p_{i,k} = \sum\sum \cdots \sum_{1 \leq j_1 < j_2 < \cdots < j_i \leq k} P(A_{j_1} \cap A_{j_2} \cap \cdots \cap A_{j_i})$$

(*sum of probabilities of all possible intersections involving i events out of k events $A_1, \ldots, A_k, i = 1, \ldots, k$*).

Then,

$$P\left(\bigcup_{i=1}^{k} A_i\right) = p_{1,k} - p_{2,k} + p_{3,k} - p_{4,k} + \cdots + (-1)^{k-1} p_{k,k}.$$

Proof. Note that, for $k = 2$, $p_{1,2} = P(A_1) + P(A_2)$, $p_{2,2} = P(A_1 \cap A_2)$ and

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) = p_{1,2} - p_{2,2}.$$

Thus the result is true for $k = 2$. Now suppose that the result is true for $k = 2, 3, \ldots, m$, that is,

$$P\left(\bigcup_{i=1}^{k} A_i\right) = p_{1,k} - p_{2,k} + p_{3,k} - p_{4,k} + \cdots + (-1)^{k-1} p_{k,k}, \quad \forall\, k = 2, 3, \ldots, m.$$

Then,

$$\begin{aligned} P\left(\bigcup_{i=1}^{m+1} A_i\right) &= P\left(\left(\bigcup_{i=1}^{m} A_i\right) \bigcup A_{m+1}\right) \\ &= P\left(\bigcup_{i=1}^{m} A_i\right) + P(A_{m+1}) - P\left(\left(\bigcup_{i=1}^{m} A_i\right) \bigcap A_{m+1}\right), \ \text{(using result for } k = 2\text{)} \\ &= \sum_{j=1}^{m} (-1)^{j-1} p_{j,m} + P(A_{m+1}) - P\left(\bigcup_{i=1}^{m} (A_i \cap A_{m+1})\right), \ \text{(using the result for } k = m \text{ on } \bigcup_{i=1}^{m} A_i\text{)} \\ &= \sum_{j=1}^{m} (-1)^{j-1} p_{j,m} + P(A_{m+1}) - \sum_{j=1}^{m} (-1)^{j-1} t_{j,m}, \ \text{(using the result for } k = m \text{ on } \bigcup_{i=1}^{m} (A_i \cap A_{m+1})\text{)}, \end{aligned}$$

where

$$\begin{aligned} t_{1,m} &= \sum_{i=1}^{m} P(A_i \cap A_{m+1}) \\ t_{2,m} &= \sum\sum_{1 \leq i < j \leq m} P(A_i \cap A_j \cap A_{m+1}) \\ t_{j,k} &= \sum\sum \cdots \sum_{1 \leq i_1 < i_2 < \cdots < i_j \leq m} P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_j} \cap A_{m+1}), \quad j = 1, 2, \ldots, m \\ t_{m,m} &= P(A_1 \cap A_2 \cap \cdots \cap A_m \cap A_{m+1}) \end{aligned}$$

Therefore,

$$P\left(\bigcup_{i=1}^{m+1} A_i\right) = (p_{1,m} + P(A_{m+1})) - (p_{2,m} + t_{1,m}) + (p_{3,m} + t_{2,m}) + \cdots + (-1)^{m-1}(p_{m,m} + t_{m-1,m}) + (-1)^m t_{m,m}$$
$$= p_{1,m+1} - p_{2,m+1} + p_{3,m+1} + \cdots + (-1)^{m-1} p_{m,m+1} + (-1)^m p_{m+1,m+1},$$

as

$$p_{1,m} + P(A_{m+1}) = \sum_{j=1}^{m} P(A_j) + P(A_{m+1}) = p_{1,m+1},$$

$$p_{2,m} + t_{1,m} = \sum\sum_{1\le i<j\le m} P(A_i \cap A_j) + \sum_{i=1}^{m} P(A_i \cap A_{m+1})$$
$$= \sum\sum_{1\le i<j\le m+1} P(A_i \cap A_j) = p_{2,m+1},$$
$$\vdots$$
$$p_{m,m} + t_{m-1,m} = P(A_1 \cap A_2 \cap \cdots \cap A_m) + \sum\sum\cdots\sum_{1\le i_1<i_2<\cdots<i_{m-1}\le m} P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_{m-1}} \cap A_{m+1})$$
$$= p_{m,m+1}$$

and $t_{m,m} = P(A_1 \cap A_2 \cap \cdots \cap A_m \cap A_{m+1}) = p_{m+1,m+1}$. The result now follows by induction. □

Remark 0.12. *Let $A_1, A_2, A_3 \in \mathcal{F}$. Then*

$$P(A_1 \cup A_2 \cup A_3) = p_{1,3} - p_{2,3} + p_{3,3}$$
$$= P(A_1) + P(A_2) + P(A_3) - P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3) + P(A_1 \cap A_2 \cap A_3)).$$

Theorem 0.13. *For some positive integer $k \ge 2$, let $A_1, A_2, \ldots, A_k \in \mathcal{F}$. Then*

$$p_{1,k} - p_{2,k} \le P\left(\bigcup_{i=1}^{k} A_i\right) \le p_{1,k}.$$

Proof. Note that for $k = 2$, $p_{1,2} = P(A_1) + P(A_2)$, $p_{2,2} = P(A_1 \cap A_2)$ and

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) \le P(A_1) + P(A_2).$$

This implies $p_{1,2} - p_{2,2} = P(A_1 \cup A_2) \le P(A_1) + P(A_2)$. Thus the result is true for $k = 2$. Now suppose that for some positive integer $m (\ge 2)$

$$p_{1,k} - p_{2,k} \le P\left(\bigcup_{i=1}^{k} A_i\right) \le p_{1,k}, \ \ \forall\, k = 1, 2, \ldots, m.$$

Then,

$$P\left(\bigcup_{i=1}^{m+1} A_i\right) = P\left(\left(\bigcup_{i=1}^{m} A_i\right) \bigcup A_{m+1}\right)$$
$$\le P\left(\bigcup_{i=1}^{m} A_i\right) + P(A_{m+1}), \quad \Big(\text{using result for } k = 2,\ A = \cup_{i=1}^{m} A_i \text{ and } B = A_{m+1},$$
$$\text{then } P(A \cup B) \le P(A) + P(B)\Big)$$
$$\le p_{1,m} + P(A_{m+1})$$
$$= p_{1,m+1}. \tag{0.3}$$

Also using the result for $k = m$, we get

$$P\left(\bigcup_{i=1}^{m} A_i\right) \geq p_{1,m} - p_{2,m}$$

and

$$P\left(\bigcup_{i=1}^{m} (A_i \cap A_{m+1})\right) \leq \sum_{i=1}^{m} P(A_i \cap A_{m+1})$$

Thus,

$$\begin{aligned} P\left(\bigcup_{i=1}^{m+1} A_i\right) &= P\left(\left(\bigcup_{i=1}^{m} A_i\right) \bigcup A_{m+1}\right) \\ &= P\left(\bigcup_{i=1}^{m} A_i\right) + P(A_{m+1}) - P\left(\bigcup_{i=1}^{m} (A_i \cap A_{m+1})\right) \\ &\geq p_{1,m} - p_{2,m} + P(A_{m+1}) - \sum_{i=1}^{m} P(A_i \cap A_{m+1}) \\ &= (p_{1,m} + P(A_{m+1})) - \left(p_{2,m} + \sum_{i=1}^{m} P(A_i \cap A_{m+1})\right) \end{aligned} \tag{0.4}$$

Using (0.3) and (0.4), we get

$$p_{1,m+1} - p_{2,m+1} \leq P\left(\bigcup_{i=1}^{m+1} A_i\right) \leq p_{1,m+1}$$

and the result follows using principle of mathematical induction. □

Remark 0.14. *It can also be shown that*

$$\begin{gathered} p_{1,k} - p_{2,k} + p_{3,k} - p_{4,k} \leq P\left(\bigcup_{i=1}^{k} A_i\right) \leq p_{1,k} - p_{2,k} + p_{3,k} \\ \vdots \\ p_{1,k} - p_{2,k} + \cdots + p_{2m-1,k} - p_{2m,k} \leq P\left(\bigcup_{i=1}^{k} A_i\right) \leq p_{1,k} - p_{2,k} + \cdots + p_{2m-1,k}, \end{gathered}$$

for $m = 1, 2, \ldots, [\frac{k}{2}]$.

Theorem 0.15 (Bonferroni's Inequality)**.** *Let* $A_1, A_2, \ldots, A_k \in \mathcal{F}$. *Then*

$$P\left(\bigcap_{i=1}^{k} A_i\right) \geq \max\left\{\sum_{i=1}^{k} P(A_i) - (k-1), 0\right\}.$$

Proof. We have

$$
\begin{aligned}
P\left(\bigcap_{i=1}^{k} A_i\right) &= P\left(\left(\bigcup_{i=1}^{k} A_i^c\right)^c\right), \ (\text{De-Morgan's law}) \\
&= 1 - P\left(\bigcup_{i=1}^{k} A_i^c\right) \\
&\geq 1 - \sum_{i=1}^{k} P(A_i^c), \quad (\text{Boole's inequality}) \\
&= 1 - \sum_{i=1}^{k} (1 - P(A_i)) \\
&= \sum_{i=1}^{k} P(A_i) - (k-1). \qquad (0.5)
\end{aligned}
$$

Also,

$$P\left(\bigcap_{i=1}^{k} A_i\right) \geq 0. \qquad (0.6)$$

Combining (0.5) and (0.6), we get

$$P\left(\bigcap_{i=1}^{k} A_i\right) \geq \max\left\{\sum_{i=1}^{k} P(A_i) - (k-1), 0\right\}.$$

This completes the proof. □

Probability as a Continuous Set Function:

A sequence of events $\{E_n, n \geq 1\}$ is said to be an increasing sequence if $E_1 \subset E_2 \subset \cdots \subset E_n \subset E_{n+1} \ldots$ whereas it is said to be decreasing sequence if $E_1 \supset E_2 \supset \cdots \supset E_n \supset E_{n+1} \supset \ldots$.

If $\{E_n\}$ is an increasing sequence of events, then we define $\lim_{n\to\infty} E_n = \bigcup_{i=1}^{\infty} E_i$. Similarly, if $\{E_n\}$ is decreasing sequence of events, then we define $\lim_{n\to\infty} E_n = \bigcap_{i=1}^{\infty} E_i$.

Theorem 0.16. *If $\{E_n\}$ is either increasing or decreasing sequence of events, then*

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

Proof. Let $\{E_n\}$ be an increasing sequence and define $F_n, n \geq 1$ as

$$
\begin{aligned}
F_1 &= E_1, \\
F_2 &= E_2 - E_1, \\
&\vdots \\
F_n &= E_n - E_{n-1}.
\end{aligned}
$$

Then, $\{F_n\}$ is a disjoint sequence of events and $E_n = \bigcup_{i=1}^{n} F_i \implies P(E_n) = \sum_{i=1}^{n} P(F_i)$. Now

$$\lim_{n\to\infty} E_n = \lim_{n\to\infty} \bigcup_{i=1}^{n} F_i = \bigcup_{n=1}^{\infty} F_n.$$

So,

$$P\left(\lim_{n\to\infty} E_n\right) = P\left(\bigcup_{n=1}^{\infty} F_n\right) = \sum_{n=1}^{\infty} P(F_n) = \lim_{n\to\infty}\sum_{i=1}^{n} P(F_i) = \lim_{n\to\infty} P\left(\bigcup_{i=1}^{n} F_i\right) = \lim_{n\to\infty} P(E_n).$$

Similarly, we can prove for other case. □

Example 0.17. *Random experiment $\mathcal{E}$: casting a red and white die.*

Sample space: $\Omega = \{(i,j) : i \in \{1,2,\ldots,6\},\ j \in \{1,2,\ldots,6\}\}$.

For $(i,j) \in \Omega$, i: number of spots up on the red die; j: number of spots up on the white die.

Event space $\mathcal{F}$ = power set of Ω.

For $A \in \mathcal{F}$, define $Q : \mathcal{F} \to \mathbb{R}$ as

$$Q(A) = \frac{|A|}{36}, \text{ where } |A| = \text{number of elements in } A.$$

Then

(a) $Q(\Omega) = \dfrac{|\Omega|}{36} = \dfrac{36}{36} = 1.$

(b) $Q(A) = \dfrac{|A|}{36} \geq 0, \ \ \forall\, A \in \mathcal{F}.$

(c) For mutually exclusive events $A_1, A_2, \ldots$

$$Q\left(\bigcup_{i=1}^{\infty} A_i\right) = \frac{|\bigcup_{i=1}^{\infty} A_i|}{36} = \frac{\sum_{i=1}^{\infty}|A_i|}{36} = \sum_{i=1}^{\infty}\frac{|A_i|}{36} = \sum_{i=1}^{\infty} Q(A_i).$$

Thus, $(\Omega, \mathcal{F}, Q)$ is a probability space.

Equally Likely Probability Models for Finite Sample Space:

Suppose that the sample space $\Omega = \{\omega_1, \omega_2, \ldots, \omega_k\}$ is finite (has k elements). Here singletons $\{\omega_i\}$ are called elementary events and $\Omega = \bigcup_{i=1}^{k}\{\omega_i\}$. Suppose that

$$P(\{\omega_i\}) = \frac{1}{k},\ i = 1,2,\ldots,k \ \text{ (each elementary event is equally likely).}$$

For any event $E \subseteq \Omega$, we have $E = \{\omega_{i_1}, \omega_{i_2}, \ldots, \omega_{i_r}\}$, for some $i_1, i_2, \ldots, i_r \in \{1,2,\ldots,k\}$, $1 \leq r \leq k$. Then, $E = \bigcup_{j=1}^{r}\{\omega_{i_j}\}$ and

$$\begin{aligned} P(E) = P\left(\bigcup_{j=1}^{r}\{\omega_{i_j}\}\right) &= \sum_{j=1}^{r} P\left(\{\omega_{i_j}\}\right) \\ &= \sum_{j=1}^{r}\frac{1}{k} = \frac{r}{k} = \frac{\text{number of ways favourable to event } E}{\text{total number of ways in which the random experiment can terminate}}. \end{aligned}$$

Here the assumption of equally likely $\left(P(\{\omega_i\}) = \dfrac{1}{k},\ i = 1,2,\ldots,k\right)$ is a part of probability modelling.

"*At random*": In a random experiment with finite sample space Ω, whenever we say that the experiment has been performed at random it means that all the outcomes in the sample space are equally likely.

Example 0.18 (Birthday Problem). *Suppose that a college has n students, including you. Each of them were born on non-leap years.*

(a) Find the probability that at least two of them have the same birthday. For what values of n this probability is more than 0.5, 0.8, 0.95?

(b) For what value of n the probability that you will find someone who shares your birthday is 0.5.

Solution: Required probability = $1 - P(\text{all of them have different birthdays}) = 1 - \dfrac{365 \times 364 \times \cdots \times (365 - n + 1)}{365^n}$

Required probability = $1 - P(\text{no one shares the same birthday as mine}) = 1 - \dfrac{364^{n-1}}{365^{n-1}}$.

For $1 - \dfrac{364^{n-1}}{365^{n-1}} \approx 0.5$, $n \approx 253$.

Example 0.19. *Five cards are drawn at random and without replacement from a deck of 52 cards. Find the probability that*

(i) each card is spade (event E_1),

(ii) at least one card is spade (event E_2),

(iii) exactly three cards are king and two cards are queen (event E_3),

(iv) exactly two kings, two queens and one jack are drawn (event E_4).

Solution: (i) $P(E_1) = \dfrac{\binom{13}{5}}{\binom{52}{5}}$,

(ii) $P(E_2) = 1 - P(E_2^c) = 1 - P(\text{no card is spade}) = 1 - \dfrac{\binom{39}{5}}{\binom{52}{5}}$,

(iii) $P(E_3) = \dfrac{\binom{4}{3}\binom{4}{2}}{\binom{52}{5}}$,

(iv) $P(E_4) = \dfrac{\binom{4}{2}\binom{4}{2}\binom{4}{1}}{\binom{52}{5}}$.

Example 0.20 (Capture/Recapture Method for Estimating Population Size). *In a wildlife population suppose that the population size n is unknown. To estimate the population size n, 20 animals are captured, tagged and then released back. Thereafter 40 animals are captured at random and it is found that 8 of them are tagged. Find an estimate of the population size n based on the given data.*

Solution: We have

$$\begin{aligned} n &= \text{total number of animals,} \\ 20 &= \text{number of tagged animals in the population,} \\ n - 20 &= \text{number of untagged animals in the population.} \end{aligned}$$

<u>Data</u>: Sample of 40 animals yield

$$\begin{aligned} \text{number of tagged animals} &= 8, \\ \text{number of untagged animals} &= 32. \end{aligned}$$

The probability of obtaining this data is

$$l(n) = \frac{\binom{20}{8}\binom{n-20}{32}}{\binom{n}{40}}, \quad n \geq 52.$$

$$\begin{aligned} l(n+1) > l(n) &\iff \frac{\binom{n-19}{32}}{\binom{n+1}{40}} > \frac{\binom{n-20}{32}}{\binom{n}{40}} \\ &\iff \frac{n-19}{(n-51)(n+1)} > 1 \\ &\iff n < 99. \end{aligned}$$

Similarly $l(n+1) < l(n) \iff n > 99$. Thus l is maximized at $n = 99$, that is, for $n = 99$, the observe data (among the captured animals 8 are tagged and 32 are untagged) is most probable.

Thus an estimate of n is $\hat{n} = 99$ (Maximum likelihood estimator).

0.3. Conditional Probability

Consider a probability space $(\Omega, \mathcal{F}, P)$ where $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ is finite and

$$P(\{\omega_i\}) = \frac{1}{n}, \; i = 1, 2, \ldots, n \;\; \text{(equally likely probability model)}.$$

Then, for any $A \in \mathcal{F}$

$$P(A) = \frac{\text{number of cases favourable to } A}{\text{total number of cases}} = \frac{|A|}{|\Omega|} = \frac{|A|}{n}.$$

Now suppose it is known a priori that event A has occured (*i.e.* outcome of the experiment is an element of A), where $|A| \geq 1$ (so that $P(A) = |A|/n > 0$). Given this prior information (that the event A has occured) we want to define probability function say $P(B|A)$ on the event space $\mathcal{F}$. A natural way to define $P(B|A)$ is

$$P(B|A) = \frac{|A \cap B|}{|A|} = \frac{|A \cap B|/n}{|A|/n} = \frac{P(A \cap B)}{P(A)}, \; B \in \mathcal{F}.$$

Definition 0.21. *Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $A \in \mathcal{F}$ be such that $P(A) > 0$. Then*

$$P(B|A) = \frac{P(A \cap B)}{P(A)}, \quad B \in \mathcal{F},$$

is called the conditional probabilty of event B given the event A.

Remark 0.22. *(a) In the above definition the event A (with $P(A) > 0$) is fixed and for this fixed $A \in \mathcal{F}$, $P(\cdot|A)$ is a set function defined on $\mathcal{F}$. Is it a probability function/ measure?*

(b) $P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)$ for $A, B \in \mathcal{F}$.

Theorem 0.23. *Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $A \in \mathcal{F}$ be such that $P(A) > 0$ be fixed. Then $P(\cdot|A) : \mathcal{F} \to \mathbb{R}$ is a probability function (called the conditional probabilty function) on $\mathcal{F}$ (so that $(\Omega, \mathcal{F}, P(\cdot|A))$ is a probability space).*

Proof. Note that $P(B|A) = \dfrac{P(A \cap B)}{P(A)} \geq 0$ for all $B \in \mathcal{F}$ and $P(\Omega|A) = \dfrac{P(A \cap \Omega)}{P(A)} = 1$.

Let $\{B_n\}_{n\geq 1}$ be a sequence of disjoint events in $\mathcal{F}$. Then,

$$P\left(\bigcup_{n=1}^{\infty} B_n \middle| A\right) = \frac{P\left(\left(\bigcup_{n=1}^{\infty} B_n\right) \cap A\right)}{P(A)} = \frac{P\left(\bigcup_{n=1}^{\infty} (B_n \cap A)\right)}{P(A)}.$$

Since $\{B_n\}_{n\geq 1}$ are disjoint then subsets $\{B_n \cap A\}_{n\geq 1}$ are also disjoint. Since $P(\cdot)$ is a probability measure, we get

$$P\left(\bigcup_{n=1}^{\infty} B_n \mid A\right) = \frac{\sum_{n=1}^{\infty} P(B_n \cap A)}{P(A)} = \sum_{n=1}^{\infty} \frac{P(B_n \cap A)}{P(A)} = \sum_{n=1}^{\infty} P(B_n|A).$$

It follows that $P(\cdot|A)$ is a probabilty function on $\mathcal{F}$ for any fixed $A \in \mathcal{F}$ with $P(A) > 0$. □

Example 0.24. *Five cards are drawn at random (without replacement) from a deck of 52 cards. Define events*

$$B : \textit{all spade in hand and} \;\; A : \textit{at least 4 spade in hand}.$$

Find $P(B|A)$.

Solution: We have

$$\begin{aligned} P(B|A) &= \frac{P(A \cap B)}{P(A)} = \frac{P(B)}{P(A)} \;\; (\text{since} \;\; B \subseteq A) \\ &= \frac{\binom{13}{5}/\binom{52}{5}}{\left[\binom{13}{4}\binom{39}{1} + \binom{13}{5}\right]/\binom{52}{5}} = 0.441. \end{aligned}$$

Remark 0.25 (Multiplication Law)**.** *(i) $P(A \cap B) = P(A)P(B|A)$, if $P(A) > 0$.*

(ii) $P(A \cap B \cap C) = P(A \cap B)P(C|A \cap B) = P(A)P(B|A)P(C|A \cap B)$, provided $P(A \cap B) > 0$ (which ensures that $P(A) > 0$ as $A \cap B \subseteq A$).

(iii) Using principle of mathematical induction, we have

$$\begin{aligned} P\left(\bigcap_{i=1}^{n} C_i\right) &= P(C_1 \cap C_2 \cap \cdots \cap C_n) \\ &= P(C_1 \cap C_2 \cap \cdots \cap C_{n-1})P(C_n|C_1 \cap C_2 \cap \cdots \cap C_{n-1}) \\ &= P(C_1 \cap C_2 \cap \cdots \cap C_{n-2})P(C_{n-1}|C_1 \cap C_2 \cap \cdots \cap C_{n-2})P(C_n|C_1 \cap C_2 \cap \cdots \cap C_{n-1}) \\ &\;\vdots \\ &= P(C_1)P(C_2|C_1)P(C_3|C_1 \cap C_2) \ldots P(C_n|C_1 \cap C_2 \cap \cdots \cap C_{n-1}) \end{aligned}$$

provided $P(C_1 \cap C_2 \cap \cdots \cap C_{n-1}) > 0$ (which also ensures that $P(C_1 \cap C_2 \cap \cdots \cap C_i) > 0$, $i = 1, 2, \ldots, n-2$).

Due to symmetry, if $(\alpha_1, \alpha_2, \ldots, \alpha_n)$ is a permutation of $(1, 2, \ldots, n)$, then

$$\begin{aligned} P\left(\bigcap_{i=1}^{n} C_i\right) &= P(C_{\alpha_1} \cap C_{\alpha_2} \cap \cdots \cap C_{\alpha_n}) \\ &= P(C_{\alpha_1})P(C_{\alpha_2}|C_{\alpha_1})P(C_{\alpha_3}|C_{\alpha_1} \cap C_{\alpha_2}) \ldots P(C_{\alpha_n}|C_{\alpha_1} \cap C_{\alpha_2} \cap \cdots \cap C_{\alpha_{n-1}}) \end{aligned}$$

provided $P(C_{\alpha_1} \cap C_{\alpha_2} \cap \cdots \cap C_{\alpha_{n-1}}) > 0$ (which also ensures that $P(C_{\alpha_1} \cap C_{\alpha_2} \cap \cdots \cap C_{\alpha_i}) > 0$, $i = 1, 2, \ldots, n-2$).

Example 0.26. *A bowl contains 3 red and 5 blue chips. All chips that are of the same colour are identical. Two chips are drawn successively at random and without replacement. Define events*

$$A : \textit{first draw resulted in a red chip},$$
$$B : \textit{second draw resulted in a blue chip}.$$

Find $P(A \cap B)$, $P(A)$ and $P(B)$.

Solution: $P(A) = \frac{3}{8}$, $P(B|A) = \frac{5}{7}$ and

$$P(B) = P(A \cap B) + P(A^c \cap B) = P(B|A)P(A) + P(B|A^c)P(A^c) = \frac{5}{7} \times \frac{3}{8} + \frac{4}{7} \times \frac{5}{8} = \frac{35}{56}.$$

Note that here the outcomes of second draw is dependent on outcome of first draw ($P(B|A) \neq P(B)$). Also,

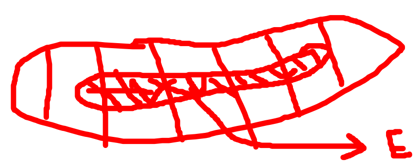
$$P(A \cap B) = P(A)P(B|A) = \frac{3}{8} \times \frac{5}{7} = 0.2679.$$

Theorem 0.27 (**Theorem of Total Probability**). *For a countable set Δ (that is elements of Δ can either be put in 1-1 correspondence with $\mathbb{N} = \{1, 2, \dots\}$ or with $\{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$), let $\{E_\alpha : \alpha \in \Delta\}$ be a countable collection of mutually exclusive (i.e., $E_\alpha \cap E_\beta = \phi$, $\forall\, \alpha \neq \beta$) and exhaustive (i.e., $P\left(\bigcup_{\alpha \in \Delta} E_\alpha\right) = 1$) events. Then, for any $E \in \mathcal{F}$,*

$$P(E) = \sum_{\alpha \in \Delta} P(E \cap E_\alpha) = \sum_{\substack{\alpha \in \Delta \\ P(E_\alpha) > 0}} P(E|E_\alpha)P(E_\alpha).$$

Proof. Since $P\left(\bigcup_{\alpha \in \Delta} E_\alpha\right) = 1$, we have

$$\begin{aligned}
P(E) &= P\left(E \bigcap \left(\bigcup_{\alpha \in \Delta} E_\alpha\right)\right) = P\left(\bigcup_{\alpha \in \Delta} (E \cap E_\alpha)\right) \\
&= \sum_{\alpha \in \Delta} P(E \cap E_\alpha), \quad (E_\alpha\text{'s are disjoint} \implies \text{their subsets } (E \cap E_\alpha)\text{'s are disjoint}) \\
&= \sum_{\substack{\alpha \in \Delta \\ P(E_\alpha) > 0}} P(E \cap E_\alpha), \quad (P(E_\alpha) = 0 \implies P(E \cap E_\alpha) = 0,\ \alpha \in \Delta) \\
&= \sum_{\substack{\alpha \in \Delta \\ P(E_\alpha) > 0}} P(E|E_\alpha)P(E_\alpha).
\end{aligned}$$



This completes the proof. □

Example 0.28. *A population comprises of 40% female and 60% male. Suppose that 15% of female and 30% of male in the population smoke. A person is selected at random from the population.*

(a) Find the probability that he/she is a smoker.

(b) Given that the selected person is smoker, find the probability that he is male.

Solution: Define the events

$$\begin{aligned}
M &: \text{selected person is a male}, \\
F = M^c &: \text{selected person is a female}, \\
S &: \text{selected person is a smoker}, \\
T = S^c &: \text{selected person is a non-smoker}.
\end{aligned}$$

We have $P(F) = 0.4,\ P(M) = 0.6,\ P(F \cup M) = P(F) + P(M) = 1,\ P(S|F) = 0.15,\ P(T|F) = 0.85,\ P(S|M) = 0.30,\ P(T|M) = 0.70$.

(a) By using Theorem of total probability, we get

$$P(S) = P(S \cap F) + P(S \cap M) = P(S|F)P(F) + P(S|M)P(M) = 0.15 \times 0.4 + 0.30 \times 0.6 = 0.24.$$

(b)

$$P(M|S) = \frac{P(M \cap S)}{P(S)} = \frac{P(S|M)P(M)}{P(S)} = \frac{0.30 \times 0.60}{0.24} = \frac{3}{4}.$$

Theorem 0.29 (Bayes' Theorem)**.** *Let $\{E_\alpha : \alpha \in \Delta\}$ be a countable collection of mutually exclusive and exhaustive events and let E be any event $P(E) > 0$. Then, for $j \in \Delta$ with $P(E_j) > 0$,*

$$P(E_j|E) = \frac{P(E|E_j)P(E_j)}{\sum\limits_{\substack{\alpha \in \Delta \\ P(E_\alpha) > 0}} P(E|E_\alpha)P(E_\alpha)}.$$

Proof. For $j \in \Delta$,

$$P(E_j|E) = \frac{P(E_j \cap E)}{P(E)} = \frac{P(E|E_j)P(E_j)}{\sum\limits_{\substack{\alpha \in \Delta \\ P(E_\alpha) > 0}} P(E|E_\alpha)P(E_\alpha)}, \quad \text{(using Theorem of total probability)}.$$

This completes the proof. □

Remark 0.30. *(a) Suppose that occurrence of any of the mutually exclusive and exhaustive events $\{E_\alpha : \alpha \in \Delta\}$ (where Δ is a countable set) may cause the occurrence of an event E. Given that the event E has occurred (i.e., given the effect), Bayes' Theorem provides the conditional probability that the event E (effect) is caused by occurrence of event $E_j,\ j \in \Delta$.*

(b) In Bayes' Theorem $\{P(E_j) : j \in \Delta\}$ are called prior probabilities and $\{P(E_j|E) : j \in \Delta\}$ are called posterior probabilities.

Example 0.31. *Bowl C_1 contains 3 red and 7 blue chips. Bowl C_2 contains 8 red and 2 blue chips. Bowl C_3 contains 5 red and 5 blue chips. All chips of the same colour are identical.*

A die is cast and a bowl is selected as per the following schemes:

Bowl C_1 is selected if 5 or 6 spots show on the upper side,
Bowl C_2 is selected if 2,3 or 4 spots show on the upper side,
Bowl C_3 is selected if 1 spots show on the upper side.

The selected bowl is handed over to another person who drawns two chips at random from this bowl. Find the probability that:

(a) Two red chips are drawn.

(b) Given that drawn chips are both red, find the probability that it came from bowl C_3.

Solution: Define the events

A_i: selected bowl is C_i, $i = 1, 2, 3$, and R: the chips drawn from the selected bowl are both red.

Then $P(A_1) = \frac{2}{6} = \frac{1}{3},\ P(A_2) = \frac{3}{6} = \frac{1}{2},\ P(A_3) = \frac{1}{6}$. Note that $\{A_1, A_2, A_3\}$ are mutually exclusive and exhaustive.

(a)

$$P(R) = P(R|A_1)P(A_1) + P(R|A_2)P(A_2) + P(R|A_3)P(A_3) = \frac{\binom{3}{2}}{\binom{10}{2}} \times \frac{1}{3} + \frac{\binom{8}{2}}{\binom{10}{2}} \times \frac{1}{2} + \frac{\binom{5}{2}}{\binom{10}{2}} \times \frac{1}{6} = \frac{10}{27}.$$

(b)

$$P(A_3|R) = \frac{P(R|A_3)P(A_3)}{P(R)} = \frac{\frac{\binom{5}{2}}{\binom{10}{2}} \times \frac{1}{6}}{\frac{10}{27}} = \frac{1}{10}.$$

Remark 0.32. *In the above example,*

$$P(A_1|R) = \frac{P(R|A_1)P(A_1)}{P(R)} = \frac{\frac{\binom{3}{2}}{\binom{10}{2}} \times \frac{1}{3}}{\frac{10}{27}} = \frac{3}{50},$$

$$P(A_2|R) = \frac{P(R|A_2)P(A_2)}{P(R)} = \frac{\frac{\binom{8}{2}}{\binom{10}{2}} \times \frac{1}{2}}{\frac{10}{27}} = \frac{21}{25},$$

$$P(A_1|R) = \frac{3}{50} < \frac{1}{3} = P(A_1) \iff P(A_1 \cap R) < P(A_1)P(R) \longleftrightarrow R \text{ has negative information about } A_1,$$

$$P(A_2|R) = \frac{21}{25} > \frac{1}{2} = P(A_2) \iff P(A_2 \cap R) > P(A_2)P(R) \longleftrightarrow R \text{ has positive information about } A_2,$$

$$P(A_3|R) = \frac{1}{10} < \frac{1}{6} = P(A_3) \iff P(A_3 \cap R) < P(A_3)P(R) \longleftrightarrow R \text{ has negative information about } A_3.$$

Note that proportion of red chips in $C_2 >$ proportion of red chips in C_i, $i = 1, 3$.

Independent Events:

Definition 0.33. *Let $\{E_j : j \in \Delta\}$ be a collection of events.*

(i) Events $\{E_j : j \in \Delta\}$ are said to be pairwise independent if for any pair of events E_α and E_β ($\alpha,\ \beta \in \Delta,\ \alpha \neq \beta$) in the collection $\{E_j : j \in \Delta\}$, we have

$$P(E_\alpha \cap E_\beta) = P(E_\alpha)P(E_\beta).$$

(ii) Events $\{E_1, E_2, \ldots, E_n\}$ are said to be independent if for any subcollection $\{E_{\alpha_1}, E_{\alpha_2}, \ldots, E_{\alpha_k}\}$ of $\{E_1, E_2, \ldots, E_n\}$ ($k = 1, 2, \ldots, n$), we have

$$P\left(\bigcap_{j=1}^{k} E_{\alpha_j}\right) = \prod_{j=1}^{k} P(E_{\alpha_j}).$$

(iii) Let $\Delta \subseteq \mathbb{R}$ be an arbitrary index set so that $\{E_\alpha : \alpha \in \Delta\}$ is an arbitrary collection of events. Events $\{E_\alpha : \alpha \in \Delta\}$ are said to be independent if any finite subcollection of events in $\{E_\alpha : \alpha \in \Delta\}$ forms a collection of independent events.

Theorem 0.34. *Let $E_1, E_2, \ldots$ be collection of independent events. Then*

$$P\left(\bigcap_{k=1}^{\infty} E_k\right) = \prod_{k=1}^{\infty} P(E_k).$$

Proof. Let $B_n = \bigcap_{k=1}^{n} E_k$, $n = 1, 2, \ldots$. Then $B_n \downarrow$ and $P\left(\bigcap_{n=1}^{\infty} B_n\right) = \lim_{n\to\infty} P(B_n)$. But $\bigcap_{n=1}^{\infty} B_n = \bigcap_{k=1}^{\infty} E_k$ and $P(B_n) = P\left(\bigcap_{k=1}^{n} E_k\right) = \prod_{k=1}^{n} P(E_k)$. Thus,

$$P\left(\bigcap_{k=1}^{\infty} E_k\right) = \lim_{n\to\infty} \prod_{k=1}^{n} P(E_k) = \prod_{k=1}^{\infty} P(E_k).$$

This completes the proof. □

Remark 0.35. *(i) To verify that n events $E_1, E_2, \ldots, E_n$ are independent one must verify*

$$\binom{n}{2} + \binom{n}{3} + \cdots + \binom{n}{n} = 2^n - n - 1 \quad \textit{conditions}.$$

For an example to conclude that three events E_1, E_2, E_3 are independent, the following four (as $2^3 - 3 - 1 = 4$) conditions must be verified:

$$P(E_1 \cap E_2) = P(E_1)P(E_2), \;\; P(E_1 \cap E_3) = P(E_1)P(E_3), \;\; P(E_2 \cap E_3) = P(E_2)P(E_3),$$

and

$$P(E_1 \cap E_2 \cap E_3) = P(E_1)P(E_2)P(E_3).$$

(ii) Any subcollection of independent events is independent. In particular, the independence of a collection of events implies their pairwise independence.

(iii) If E_1 and E_2 are independent events $(P(E_1) > 0,\; P(E_2) > 0)$, then

$$P(E_1|E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)} = \frac{P(E_1)P(E_2)}{P(E_2)} = P(E_1),$$

that is, conditional probability of E_1 given E_2 is the same as unconditional probability of E_1.

Similarly, if E_1, E_2 and E_3 are independent events then $P(E_1|E_2 \cap E_3) = P(E_1)$).

Example 0.36. *Consider the probability space $(\Omega, \mathcal{F}, P)$ with $\Omega = \{1, 2, 3, 4\}$ and $P(\{i\}) = 1/4, \;\; i = 1, 2, 3, 4$. Let $A = \{1, 4\}, \; B = \{2, 4\}, \; C = \{3, 4\}$. Then, show that A, B and C are pairwise independent but not independent.*

Solution: We have $P(A) = P(B) = P(C) = 1/2$. Also, $P(A \cap B) = P(A \cap C) = P(B \cap C) = P(\{4\}) = 1/4$. Thus,

$$P(A \cap B) = P(A)P(B), \;\; P(A \cap C) = P(A)P(C), \;\; P(B \cap C) = P(B)P(C),$$

which implies that A, B and C are pairwise independent. However,

$$P(A \cap B \cap C) = P(\{4\}) = 1/4 \neq 1/8 = P(A)P(B)P(C),$$

which implies that A, B and C are not independent although they are pairwise independent.

Example 0.37. *Let $E_1, E_2, \ldots, E_n$ be a collection of independent events. Show that,*

(a) for any permutation $(\alpha_1, \ldots, \alpha_n)$ of $(1, \ldots, n)$, $E_{\alpha_1}, E_{\alpha_2}, \ldots, E_{\alpha_n}$ are independent;

(b) $E_1, E_2, \ldots, E_k, E_{k+1}^c, \ldots, E_n^c$ are independent for any $k \in \{0, 1, \ldots, n-1\}$;

(c) E_1^c and $E_2 \cup E_3^c \cup E_5$ are independent.

(d) $E_1 \cup E_2^c$, E_3^c and $E_4 \cap E_5^c$ are independent.

Remark 0.38. *When we say that the two random experiments are performed independently, it means that the events associated with two random experiments are independent.*

L-4

0.4. Random Variables and their Distribution Functions

Let $(\Omega, \mathcal{F}, P)$ be a given probability space. In some situations we may not be directly interested in the sample space Ω; rather we may be interested in some numerical aspect of Ω.

Example 0.39. *A fair coin (head and tail are equally likely) is tossed three times independently. Then,*

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

and $P(\{\omega\}) = 1/8$ for all $\omega \in \Omega$. Suppose that we are interested in number of heads in three tosses, i.e., we are interested in the function $X : \Omega \to \mathbb{R}$ defined as

$$X(\omega) = \begin{cases} 0, & \textit{if } \omega = TTT, \\ 1, & \textit{if } \omega \in \{HTT, THT, TTH\}, \\ 2, & \textit{if } \omega \in \{HHT, HTH, THH\}, \\ 3, & \textit{if } \omega = HHH. \end{cases}$$

Clearly the values assumed by X are random with

$$P(X = 0) = P(X = 3) = 1/8 \textit{ and } P(X = 1) = P(X = 2) = 3/8.$$

Hence $P(X \in \{0, 1, 2, 3\}) = 1$.

Definition 0.40. *Let $(\Omega, \mathcal{F}, P)$ be a given probability space. A real valued* **measurable** *function $X : \Omega \to \mathbb{R}$ (defined on sample space Ω) is called a random variable (r.v.).*

Note: From rigorous mathematical point of view a random variable is a real valued function with some technical condition. In this course we are ignoring these technical details. For all practical purpose r.v. is a real valued function defined on Ω.

For a function $Y : \Omega \to \mathbb{R}$ and $A \subseteq \mathbb{R}$, define

$$Y^{-1}(A) = \{\omega \in \Omega : Y(\omega) \in A\}.$$

Then it is straightforward to prove the following result:

Proposition 0.41. *Let $A \subseteq \mathbb{R}$, $B \subseteq \mathbb{R}$ and $A_\alpha \subseteq \mathbb{R}$, $\alpha \in \Lambda$, where Λ is an arbitrary index set. Let $Y : \Omega \to \mathbb{R}$ be a given function. Then*

(a) If $A \cap B = \phi$, then $Y^{-1}(A) \cap Y^{-1}(B) = \phi$;

(b) $Y^{-1}(A^c) = (Y^{-1}(A))^c$ (that is, $Y^{-1}(\mathbb{R} - A) = Y^{-1}(\mathbb{R}) - Y^{-1}(A) = \Omega - Y^{-1}(A)$);

(c) $Y^{-1}\left(\bigcup_{\alpha\in\Lambda} A_\alpha\right) = \bigcup_{\alpha\in\Lambda} Y^{-1}(A_\alpha)$;

(d) $Y^{-1}\left(\bigcap_{\alpha\in\Lambda} A_\alpha\right) = \bigcap_{\alpha\in\Lambda} Y^{-1}(A_\alpha)$.

For a probability space $(\Omega, \mathcal{F}, P)$ and a r.v. $X : \Omega \to \mathbb{R}$, note that $\forall\, B \subseteq \mathscr{B}$

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}.$$

Thus, one can define a set function $P_X : \mathscr{B} \to [0, 1]$ by

$$P_X(B) = P(X^{-1}(B)) = P\left(\{\omega \in \Omega : X(\omega) \in B\}\right), \quad B \in \mathscr{B},$$

where $\mathscr{B}$ is some class of subsets of $\mathbb{R}$. Here, also for all practical purpose we will take $\mathscr{B}$ to be a sigma algebra formed by open subsets of $\mathbb{R}$.

We simply write

$$P_X(B) = P(\{\omega \in \Omega : X(\omega) \in B\}) = P(X \in B), \;\; B \in \mathscr{B}.$$

We have the following scenario $(\Omega, \mathcal{F}, P) \xrightarrow{X} (\mathbb{R}, \mathscr{B}, P_X)$.

Theorem 0.42 (Induced probability space / measures)**.** *$(\mathbb{R}, \mathscr{B}, P_X)$ (as defined above) is a probability space, i.e., $P_X(\cdot)$ is a probability function defined on $\mathscr{B}$.*

Proof. (i) $P_X(\mathbb{R}) = P(X \in \mathbb{R}) = P(X^{-1}(\mathbb{R})) = P(\Omega) = 1$.

(ii) For any $B \in \mathscr{B}$, $P_X(B) = P(X^{-1}(B)) \geq 0$.

(iii) Let $\{B_n\}$ be a collection of mutually exclusive events in $\mathscr{B}$. Then,

$$\begin{aligned} P_X\left(\bigcup_{n=1}^{\infty} B_n\right) &= P\left(X^{-1}\left(\bigcup_{n=1}^{\infty} B_n\right)\right) \\ &= P\left(\bigcup_{n=1}^{\infty} X^{-1}(B_n)\right), \quad \text{(Proposition } 0.41(c)) \\ &= \sum_{n=1}^{\infty} P(X^{-1}(B_n)), \;\; (P \text{ is a probability measure and using Proposition } 0.41(a)) \\ &= \sum_{n=1}^{\infty} P_X(B_n). \end{aligned}$$

This completes the proof. □

Definition 0.43. *The probability function P_X defined above is called the probability function/ measure induced by r.v. X and $(\mathbb{R}, \mathscr{B}, P_X)$ is called the probability space induced by r.v. X.*

The induced probability measure P_X describes the random behaviour of X.

Example 0.44. *Toss a coin three times independently. Then,*

$$\Omega = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\} \text{ and } P(\{\omega\}) = 1/8, \;\; \forall\, \omega \in \Omega$$

and $X : \Omega \to \mathbb{R}$ (number of heads in three tosses) is defined by

$$X(\omega) = \begin{cases} 0, & \text{if } \omega \in \{TTT\}, \\ 1, & \text{if } \omega \in \{HTT, THT, TTH\}, \\ 2, & \text{if } \omega \in \{HHT, HTH, THH\}, \\ 3, & \text{if } \omega \in \{HHH\}. \end{cases}$$

Obviously, $X : \Omega \to \mathbb{R}$ is r.v. with induced probability space given by $(\mathbb{R}, \mathscr{B}, P_X)$, where

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

Now for any $B \in \mathscr{B}$,

$$P_X(B) = P(X^{-1}(B)) = P(\{\omega \in \Omega : X(\omega) \in B\}) = \sum_{i \in B \cap \{0,1,2,3\}} P_X(\{i\}).$$

Definition 0.45. *Let X be a r.v. defined on probability space $(\Omega, \mathcal{F}, P)$ and let $(\mathbb{R}, \mathscr{B}, P_X)$ denote the probability space induced by X. Define the function $F_X : \mathbb{R} \to \mathbb{R}$ by*

$$F_X(x) = P(X \leq x) = P(X^{-1}(-\infty, x]) = P_X((-\infty, x]), \quad x \in \mathbb{R}.$$

The function F_X is called the cumulative distribution function (c.d.f.) or simply the distribution function (d.f.) of r.v. X.

Note: Whenever there is no ambiguity we will drop subscript X in F_X to represent d.f. of a r.v. by F. It can be shown (in advanced courses) that the c.d.f. $F_X(\cdot)$ of a r.v. X determines the induced probability measure $P_X(\cdot)$ uniquely. Thus to study the random behaviour of r.v. X it suffices to study its d.f. F.

Example 0.46. *In the previous example*

$$P(X = 0) = P_X(\{0\}) = 1/8, \; P(X = 1) = P_X(\{1\}) = 3/8 = P(X = 2) = P_X(\{2\})$$

and $P(X = 3) = P_X(\{3\}) = 1/8$. Then, the d.f. of X is obtained as

$$F_X(x) = P(X \leq x) = P(\{\omega : X(\omega) \leq x\}) = \sum_{\substack{i \in \{0,1,2,3\} \\ i \leq x}} P_X(\{i\}) = \begin{cases} 0, & x < 0, \\ 1/8, & 0 \leq x < 1, \\ 1/8 + 3/8 = 1/2, & 1 \leq x < 2, \\ 7/8, & 2 \leq x < 3, \\ 1, & x \geq 3. \end{cases}$$

Theorem 0.47. *Let $F(\cdot)$ be the c.d.f. of a r.v. X defined on a probability space $(\Omega, \mathcal{F}, P)$ and let $(\mathbb{R}, \mathscr{B}, P_X)$ be the probability space induced by X. Then*

(i) F is non-decreasing,

(ii) $F(x)$ is right continuous,

(iii) $F(-\infty) = \lim_{n \uparrow \infty} F(-n) = 0$ and $F(\infty) = \lim_{n \uparrow \infty} F(n) = 1$.

Conversely, any function $G(\cdot)$ satisfying properties (i)-(iii) is a d.f. of some r.v. Y defined on a probability space $(\Omega^*, \mathcal{F}^*, P^*)$.

Proof. (i) Let $-\infty < x < y < \infty$. Then $(-\infty, x] \subseteq (-\infty, y] \implies P_X((-\infty, x]) \leq P_X((-\infty, y])$. This implies that $F(x) \leq F(y)$.

(ii) Since F is monotone and bounded below (by 0), $\lim_{h \downarrow 0} F(x + h) = F(x+)$ exists $\forall\, x \in \mathbb{R}$. Therefore,

$$F(x+) = \lim_{h \downarrow 0} F(x + h) = \lim_{n \to \infty} F\left(x + \frac{1}{n}\right) = \lim_{n \to \infty} P_X\left((-\infty, x + 1/n]\right).$$

Let $A_n = (-\infty, x + 1/n]$, $n = 1, 2, \ldots$. Then $A_n \downarrow$ and $\bigcap_{n=1}^{\infty} (-\infty, x + 1/n] = (-\infty, x]$. Thus,

$$F(x+) = P_X\left(\bigcap_{n=1}^{\infty} (-\infty, x + 1/n]\right) = P_X((-\infty, x]) = F(x).$$

(iii) Note that

$$\begin{aligned}
F(-\infty) = \lim_{n\to\infty} F(-n) = \lim_{n\to\infty} P_X((-\infty,-n]) &= P_X\left(\bigcap_{n=1}^{\infty}(-\infty,-n]\right), \quad ((-\infty,-n]\downarrow) \\
&= P_X(\phi), \quad \left(\bigcap_{n=1}^{\infty}(-\infty,-n] = \phi\right) \\
&= 0.
\end{aligned}$$

Also,

$$\begin{aligned}
F(+\infty) = \lim_{n\to\infty} F(n) = \lim_{n\to\infty} P_X((-\infty,n]) &= P_X\left(\bigcup_{n=1}^{\infty}(-\infty,n]\right), \quad ((-\infty,n]\uparrow) \\
&= P_X(\mathbb{R}), \quad \left(\bigcup_{n=1}^{\infty}(-\infty,n] = \mathbb{R}\right) \\
&= 1.
\end{aligned}$$

This completes the proof. □

Remark 0.48. *(i) Since any distribution function is monotone and bounded above (by 1),* $\lim_{h\downarrow 0} F(x-h) = F(x-)$ *exists* $\forall x \in \mathbb{R}$*. Moreover,*

$$\begin{aligned}
F(x-) = \lim_{h\downarrow 0} F(x-h) = \lim_{n\to\infty} F(x-1/n) = \lim_{n\to\infty} P_X((-\infty,x-1/n]) \\
&= P_X\left(\bigcup_{n=1}^{\infty}(-\infty,x-1/n]\right), \quad ((-\infty,x-1/n]\uparrow) \\
&= P_X((-\infty,x)) = P(X<x).
\end{aligned}$$

(ii) From the calculus we know that any monotone function is either continuous on $\mathbb{R}$ *or it has atmost countable number of discontinuities. Thus any c.d.f* $F(x)$ *is either continuous on* $\mathbb{R}$ *or has atmost countable number of discontinuities. Since, for any* $x \in \mathbb{R}$, $F(x+)$ *and* $F(x-)$ *exist,* F *has only jump discontinuities* ($F(x) = F(x+) > F(x-)$)*.*

(iii) A distribution function F *is continuous at* $a \in \mathbb{R}$ *iff* $F(a) = F(a-)$*.*

(iv) For any $a \in \mathbb{R}$, $P(X=a) = P(X\le a) - P(X<a) = F(a) - F(a-)$*. Thus, a d.f.* F *is continuous at* $a \in \mathbb{R}$ *iff* $P(X=a) = F(a) - F(a-) = 0$*.*

(v) For $-\infty < a < b < \infty$, $\;P(X\le b) = P(X\le a) + P(a < X \le b)$*.*

$$P(a<X\le b) = P(X\le b) - P(X\le a) = F(b) - F(a).$$

Similarly, for $-\infty < a < b < \infty$,

$$\begin{aligned}
P(a<X<b) &= P(X<b) - P(X\le a) = F(b-) - F(a), \\
P(a\le X\le b) &= P(X\le b) - P(X<a) = F(b) - F(a-), \\
P(a\le X<b) &= P(X<b) - P(X<a) = F(b-) - F(a-), \\
P(a<X<b) &= P(X<b) - P(X\le a) = F(b-) - F(a), \\
P(X>a) &= 1 - P(X\le a) = 1 - F(a), \\
P(X\ge a) &= 1 - P(X<a) = 1 - F(a-).
\end{aligned}$$

Example 0.49. *Consider the function $G : \mathbb{R} \to \mathbb{R}$ defined by*

$$G(x) = \begin{cases} 0, & \text{if } x < 0, \\ \frac{x}{3}, & \text{if } 0 \le x < 1, \\ \frac{1}{2}, & \text{if } 1 \le x < 2, \\ \frac{2}{3}, & \text{if } 2 \le x < 3, \\ 1, & \text{if } x \ge 3. \end{cases}$$

(a) Show that G is d.f. of some r.v. X,

(b) Find $P(X = a)$ for various values of $a \in \mathbb{R}$,

(c) Find $P(X < 3),\ P\left(X \ge \frac{1}{2}\right),\ P(2 < X \le 4),\ P(1 \le X < 2), P(2 \le X \le 3)$ and $P\left(\frac{1}{2} < X < 3\right)$.

Solution: (a) Clearly G is non-decreasing in $(-\infty, 0),\ (0, 1),\ (1, 2),\ (2, 3)$ and $(3, \infty)$. Moreover,

$$G(0) - G(0-) = 0 \ge 0,\ \ G(1) - G(1-) = \frac{1}{2} - \frac{1}{3} > 0,$$
$$G(2) - G(2-) = \frac{2}{3} - \frac{1}{2} > 0,\ \ G(3) - G(3-) = 1 - \frac{2}{3} > 0.$$

It follows that G is non-decreasing.

Clearly G is continuous (and hence right continuous) on $(-\infty, 0),\ (0, 1),\ (1, 2),\ (2, 3)$ and $(3, \infty)$. Moreover,

$$\left.\begin{array}{lll} G(0+) - G(0) & = 0 - 0 & = 0 \\ G(1+) - G(1) & = 1/2 - 1/2 & = 0 \\ G(2+) - G(2) & = 2/3 - 2/3 & = 0 \\ G(3+) - G(3) & = 1 - 1 & = 0 \end{array}\right\} \implies \quad G \text{ is right continuous on } \mathbb{R}.$$

Also, $G(+\infty) = \lim\limits_{x \to \infty} G(x) = 1\ \ \&\ G(-\infty) = \lim\limits_{x \to \infty} G(-x) = 0$. Thus, G is a d.f. of some random variable X.

(b) The set of discontinuity points of F is $D = \{1, 2, 3\}$. Thus,

$$P(X = a) = G(a) - G(a-) = 0,\ \ \forall\, a \ne 1, 2, 3,$$
$$P(X = 1) = G(1) - G(1-) = \frac{1}{2} - \frac{1}{3} = \frac{1}{6},$$
$$P(X = 2) = G(2) - G(2-) = \frac{2}{3} - \frac{1}{2} = \frac{1}{6},$$
$$P(X = 3) = G(3) - G(3-) = 1 - \frac{2}{3} = \frac{1}{3}.$$

(c) Note that

$$\begin{aligned}
&P(X < 3) = G(3-) = \frac{2}{3},\\
&P\left(X \geq \frac{1}{2}\right) = 1 - G\left(\frac{1}{2}-\right) = 1 - \frac{1}{6} = \frac{5}{6},\\
&P(2 < X \leq 4) = G(4) - G(2) = 1 - \frac{2}{3} = \frac{1}{3},\\
&P(1 \leq X < 2) = G(2-) - G(1-) = \frac{1}{2} - \frac{1}{3} = \frac{1}{6},\\
&P(2 \leq X \leq 3) = G(3) - G(2-) = 1 - \frac{1}{2} = \frac{1}{2},\\
&P\left(\frac{1}{2} < X < 3\right) = G(3-) - G\left(\frac{1}{2}\right) = \frac{2}{3} - \frac{1}{6} = \frac{1}{2}.
\end{aligned}$$

0.5. Discrete Random Variables

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $X : \Omega \to \mathbb{R}$ be a r.v. with induced probability space $(\mathbb{R}, \mathscr{B}, P_X)$ and d.f. F.

Definition 0.50. *The r.v. X is said to be a discrete r.v. if there exists a countable set S (finite or infinite) such that*

$$P(X = x) = F(x) - F(x-) > 0,\ \forall\, x \in S, \quad \textit{and}\ \ P(X \in S) = 1.$$

The set S is called the support of r.v. X.

Remark 0.51. *(i) If S is the support of a discrete r.v. X, then clearly*

$$S = \{x \in \mathbb{R} : F(x) - F(x-) > 0\} = \textit{ set of discontinuity points of } F.$$

(ii) If x is a discontinuity point of d.f. F then

$$F(x) - F(x-) = \textit{ size of jump of } F \textit{ at } x.$$

Thus, a r.v. X is of discrete type $\iff$ sum of jump points of F equals 1, i.e.,

$$P(X \in S) = \sum_{x \in S} P(X = x) = \sum_{x \in S} [F(x) - F(x-)] = 1.$$

Example 0.52. *In Example 0.49 the set of discontinuity points of G is $D = \{1, 2, 3\}$ and*

$$\sum_{x \in D} [G(x) - G(x-)] = 1/6 + 1/6 + 1/3 = 2/3 < 1 \implies X \textit{ is not a discrete r.v.}$$

Example 0.53. *Consider the d.f. (see Example 0.46)*

$$F_X(x) = \begin{cases} 0, & \textit{if } x < 0,\\ 1/8, & \textit{if } 0 \leq x < 1,\\ 1/2, & \textit{if } 1 \leq x < 2,\\ 7/8, & \textit{if } 2 \leq x < 3,\\ 1, & \textit{if } \ \ x \geq 3. \end{cases}$$

The set of discontinuity points of F is $D = \{0, 1, 2, 3\}$ with

$$\sum_{x\in D}[F(x) - F(x-)] = \frac{1}{8} + \left(\frac{1}{2} - \frac{1}{8}\right) + \left(\frac{7}{8} - \frac{1}{2}\right) + \left(1 - \frac{7}{8}\right) = 1,$$

which implies that X is a discrete r.v. with support $S = D = \{0, 1, 2, 3\}$.

Definition 0.54. *Let X be a r.v. with c.d.f. F_X and support S_X. Define the function $f_X : \mathbb{R} \to \mathbb{R}$ by*

$$f_X(x) = \begin{cases} P(X = x) = F_X(x) - F_X(x-) > 0, & \text{if } x \in S_X, \\ 0, & \text{otherwise.} \end{cases}$$

The function f_X is called the probability mass function (p.m.f.) of r.v. X.

Whenever there is no ambiguity we will drop subscript X in F_X, S_X and f_X to represent the d.f. of X by F, the support of X by S and the p.m.f. of X by f.

Remark 0.55. *(i) Let X be a discrete r.v with p.m.f. f and d.f F. Then, for any $A \subseteq \mathbb{R}$*

$$P(X \in A) = P(X \in A \cap S) = \sum_{x\in A\cap S} f(x), \quad (A \cap S \subseteq S \text{ and thus } A \cap S \text{ is a countable set}),$$

where S is the support of X.

Moreover, $F(x) = \displaystyle\sum_{y\in S\cap(-\infty,x]} f(y)$. Also, for any $x \in S$, $f(x) = F(x) - F(x-)$.

(ii) Clearly a d.f. determines the p.m.f. uniquely and vice-versa. Thus it suffices to study the p.m.f. of discrete r.v.

(iii) Let X be a discrete r.v. with p.m.f. f and support S. Then, $f : \mathbb{R} \to \mathbb{R}$ satisfies

$$(i)\ f(x) > 0,\ \forall\, x \in S, \quad (ii)\ \sum_{x\in S} f(x) = 1.$$

Conversely, suppose that $g : \mathbb{R} \to \mathbb{R}$ is a function such that, for some countable set T

$$(i)\ g(x) > 0,\ \forall\, x \in T \text{ and } (ii)\ \sum_{x\in T} g(x) = 1.$$

Then, $g(\cdot)$ is the p.m.f. of some discrete r.v. having support T.

Example 0.56. *Let X be a r.v. having d.f.*

$$F(x) = \begin{cases} 0, & \text{if } x < 0, \\ 1/8, & \text{if } 0 \le x < 1, \\ 1/2, & \text{if } 1 \le x < 2, \\ 7/8, & \text{if } 2 \le x < 3, \\ 1, & \text{if } x \ge 3. \end{cases}$$

We have seen in Example 0.53 that X is a discrete r.v with support $S = \{0, 1, 2, 3\}$. Then, the p.m.f of X is $f : \mathbb{R} \to \mathbb{R}$, where

$$f(0) = F(0) - F(0-) = 1/8, \quad f(1) = F(1) - F(1-) = 1/2 - 1/8 = 3/8,$$
$$f(2) = F(2) - F(2-) = 7/8 - 1/2 = 3/8 \text{ and } f(3) = F(3) - F(3-) = 1 - 7/8 = 1/8.$$

Thus, the p.m.f. of X is

$$f(x) = \begin{cases} 1/8, & x = 0, 3, \\ 3/8, & x = 1, 2, \\ 0, & \textit{otherwise}. \end{cases}$$

Example 0.57. *A fair die (all outcomes are equally likely) is tossed repeatedly and independently until a 6 is observed. Then X is a discrete r.v. with support $S = \{1, 2, 3, \dots\}$.*

$$\textit{p.m.f. } f(x) = P(X = x) = \begin{cases} \left(\frac{5}{6}\right)^{x-1} \frac{1}{6}, & \textit{if } x = 1, 2, 3, \dots, \\ 0, & \textit{otherwise} \end{cases}$$

and d.f.

$$F(x) = \begin{cases} 0, & \textit{if } x < 1, \\ 1/6, & \textit{if } 1 \le x < 2, \\ 11/36, & \textit{if } 2 \le x < 3, \\ \vdots & \\ \sum_{j=1}^{i} (\frac{5}{6})^{j-1} \frac{1}{6} = 1 - \left(\frac{5}{6}\right)^{i}, & \textit{if } i \le x < i+1. \end{cases}$$

0.6. Continuous Random Variable

Let X be a random variable with d.f. F.

Definition 0.58. *The r.v. X is said to be a continuous r.v. if there exists a non-negative integrable function $f : \mathbb{R} \to [0, \infty)$ such that, for any $x \in \mathbb{R}$,*

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)\mathrm{d}t.$$

The function $f(\cdot)$ is called the probability density function (p.d.f.) of X. The support of the continuous r.v X is the set $S = \{x \in \mathbb{R} : F(x+h) - F(x-h) > 0, \ \forall\, h > 0\}$, that is, $S = \{x \in \mathbb{R} : \int_{x-h}^{x+h} f(t)\mathrm{d}t > 0, \ \forall\, h > 0\}$.

Remark 0.59. *(i) From the fundamental theorem of calculus, we know that the definite integral*

$$F(x) = \int_{-\infty}^{x} f(t)\mathrm{d}t$$

is a continuous function on $\mathbb{R}$. Thus, the d.f F of any continuous r.v X is continuous everywhere on $\mathbb{R}$. In particular,

$$P(X = x) = F(x) - F(x-) = 0, \ \forall\, x \in \mathbb{R}.$$

Generally, if A is any countable subset of $\mathbb{R}$ then for any continuous r.v. X

$$P(X \in A) = \sum_{x \in A} P(X = x) = 0.$$

(ii) If X is a continuous r.v. then

(a) $P(X < x) = P(X \le x) = F(x), \ \forall\, x \in \mathbb{R},$

(b) $P(X \geq x) = 1 - P(X < x) = 1 - F(x), \quad \forall\, x \in \mathbb{R}$,

(c) For any $a, b \in \mathbb{R}$, $-\infty < a < b < \infty$,

$$\begin{aligned} P(a < X < b) = P(a \leq X < b) = P(a < X \leq b) &= P(a \leq X \leq b) \\ &= F(b) - F(a) \\ &= \int_{-\infty}^{b} f(t)\mathrm{d}t - \int_{-\infty}^{a} f(t)\mathrm{d}t = \int_{a}^{b} f(t)\mathrm{d}t. \end{aligned}$$

(iii) Let $f(\cdot)$ *be the p.d.f. of a continuous r.v.* X *and let* $E \subseteq \mathbb{R}$ *be any countable subset of* $\mathbb{R}$. *Define* $g : \mathbb{R} \to [0, \infty)$ *by*

$$g(x) = \begin{cases} f(x), & \textit{if } x \in \mathbb{R} \cap E^c, \\ C_x, & \textit{if } x \in E, \end{cases}$$

where $C_x \geq 0$ *are arbitrary. Then*

$$F(x) = \int_{-\infty}^{x} f(t)\mathrm{d}t = \int_{-\infty}^{x} g(t)\mathrm{d}t, \quad \forall\, x \in \mathbb{R}$$

and, thus, g *is also a p.d.f. of* X. *Thus, the p.d.f. of a continuous r.v. is not unique.*

(iv) There are random variables that are neither discrete nor continuous (see Example 0.49). Such random variables will not be studied here.

We state the following theorem without proof.

Theorem 0.60. *Let* X *be a r.v. with d.f.* F. *Suppose that* F *is differentiable everywhere except (possibly) on a countable set* E. *Further suppose that* $\int_{-\infty}^{\infty} F'(t)\mathrm{d}t = 1$. *Then,* X *is a continuous r.v with p.d.f.*

$$f(x) = \begin{cases} F'(x), & x \in E^c, \\ 0, & x \in E. \end{cases}$$

Remark 0.61. *(i) The p.d.f. determines the d.f. uniquely. Converse is not true. However, the d.f. determines the p.d.f. almost uniquely (they may vary on sets that have no length (or have zero content)). Thus it is enough to study the p.d.f. of a continuous r.v.*

(ii) Let X *be continuous r.v. with p.d.f* $f(x)$. *Then,*

$$(a) \;\; f(x) \geq 0, \;\; \forall\, x \in \mathbb{R} \;\; \textit{and} \;\; (b) \;\; \int_{-\infty}^{\infty} f(t)\mathrm{d}t = 1.$$

Conversely, suppose that $g : \mathbb{R} \to \mathbb{R}$ *is a function such that*

$$(a) \;\; g(x) \geq 0, \;\; \forall\, x \in \mathbb{R}, \quad (b) \;\; \int_{-\infty}^{\infty} g(t)\mathrm{d}t = 1.$$

Then, $g(\cdot)$ *is the p.d.f. of some continuous r.v. having support* $T = \left\{ x \in \mathbb{R} : \int_{x-h}^{x+h} g(t)\mathrm{d}t > 0, \;\; \forall\, h > 0 \right\}$.

Example 0.62. *Let* X *be a r.v. with d.f.*

$$F(x) = \begin{cases} 0, & \textit{if } x < 0, \\ x/4, & \textit{if } 0 \leq x < 1, \\ x/3, & \textit{if } 1 \leq x < 2, \\ 3x/8, & \textit{if } 2 \leq x < 5/2, \\ 1, & \textit{if } x \geq 5/2. \end{cases}$$

Examine whether X is a continuous r.v. or a discrete r.v. or none?

Solution: Let D be the set of discontinuity points of F. Then $D = \{1, 2, 5/2\}$. So, $D \neq \phi \implies X$ is not a continuous r.v. So

$$\sum_{x\in D}[F(x) - F(x-)] = \left(\frac{1}{3} - \frac{1}{4}\right) + \left(\frac{3}{4} - \frac{2}{3}\right) + \left(1 - \frac{15}{16}\right) = \frac{11}{48} < 1 \implies X \text{is not a discrete r.v.}$$

Thus, X is neither a discrete nor a continuous r.v.

Example 0.63. *Let X be a r.v with d.f.*

$$F(x) = \begin{cases} 0, & \textit{if } x < 0, \\ x^2/2, & \textit{if } 0 \leq x < 1, \\ x/2, & \textit{if } 1 \leq x < 2, \\ 1, & \textit{if } x \geq 2. \end{cases}$$

Show that X is a continuous r.v. Find the p.d.f. of X and support of X.

Solution: Clearly F is continuous everywhere. Moreover, F is differentiable everywhere except at two (countable) points 1, 2, and

$$F'(x) = \begin{cases} 0, & \text{if } x < 0, \\ x, & \text{if } 0 < x < 1, \\ 1/2, & \text{if } 1 < x < 2, \\ 0, & \text{if } x \geq 2. \end{cases}$$

Also, $\int_{-\infty}^{\infty} F'(x)\mathrm{d}x = \int_0^1 x\mathrm{d}x + \int_1^2 \frac{1}{2}\mathrm{d}x = 1 \implies X$ is continuous r.v. with p.d.f.

$$f(x) = \begin{cases} x, & \text{if } 0 < x < 1, \\ 1/2, & \text{if } 1 < x < 2, \\ 0, & \text{otherwise.} \end{cases}$$

The support of X is

$$S = \{x \in \mathbb{R} : F(x+h) - F(x-h) > 0, \ \forall\, h > 0\} = \left\{x \in \mathbb{R} : \int_{x-h}^{x+h} f(t)\mathrm{d}t > 0, \ \forall\, h > 0\right\} = [0, 2].$$

Example 0.64. *Let X be a continuous r.v. with p.d.f.*

$$f(x) = \begin{cases} x^2, & \textit{if } 0 < x < 1, \\ ce^{-x}, & \textit{if } x \geq 1, \quad \textit{where } c \geq 0 \textit{ is a constant}, \\ 0, & \textit{otherwise}. \end{cases}$$

(a) Find the value of c,

(b) Find $P(1/2 \leq X \leq 2)$,

(c) Find the support of X,

(d) Find the d.f. of X.

Solution: (a) We have

$$\int_a^b f(x)\mathrm{d}x = 1 \implies \int_0^1 x^2\mathrm{d}x + \int_1^\infty ce^{-x}\mathrm{d}x = 1 \implies 1/3 + ce^{-1} = 1 \implies c = \frac{2e}{3}.$$

(b) Observe that,

$$P(1/2 \le X \le 2) = \int_{1/2}^2 f(x)\mathrm{d}x = \int_{1/2}^1 x^2\mathrm{d}x + c\int_1^2 e^{-x}\mathrm{d}x$$
$$= \frac{1}{3}(1-1/8) + c(e^{-1}-e^{-2}) = \frac{7}{24} + \frac{2}{3}(1-e^{-1}).$$

(c) The support of X is $S = \left\{x \in \mathbb{R} : \int_{x-h}^{x+h} f(t)\mathrm{d}t > 0,\ \forall\, h > 0\right\} = [0,\infty)$.

(d) The d.f. of X is $F(x) = \int_{-\infty}^x f(t)\mathrm{d}t$. For $x < 0$, clearly $F(x) = 0$. For $0 \le x < 1$,

$$F(x) = \int_0^x t^2\mathrm{d}t = x^3/3.$$

For $x \ge 1$,

$$F(x) = \int_0^1 t^2\mathrm{d}t + c\int_1^x e^{-t}\mathrm{d}t = \frac{1}{3} + \frac{2}{3}(1-e^{-(x-1)}).$$

Thus,

$$F(x) = \begin{cases} 0, & \text{if } x < 0, \\ \dfrac{x^3}{3}, & \text{if } 0 \le x < 1, \\ \dfrac{1}{3} + \dfrac{2}{3}(1-e^{-(x-1)}), & \text{if } x \ge 1. \end{cases}$$

Remark 0.65. *Let X be a continuous r.v. with p.d.f. $f(\cdot)$. If f is continuous at $x \in \mathbb{R}$, then*

$$f(x) = \lim_{\delta \downarrow 0} \frac{1}{\delta}\int_{x-\delta/2}^{x+\delta/2} f(t)\mathrm{d}t \implies P(x-\delta/2 \le X \le x+\delta/2) \approx \delta f(x), \text{ for small } \delta > 0,$$

that is, $P(x - \mathrm{d}x \le X \le x + \mathrm{d}x) \approx f(x)\mathrm{d}x.$

0.7. Probability Distribution of a Function of Discrete Random Variable

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $X : \Omega \to \mathbb{R}$ be a r.v. with d.f. F, p.m.f. f and support S. Let $h : \mathbb{R} \to \mathbb{R}$ be a given function. Define $Z : \Omega \to \mathbb{R}$ as

$$Z(\omega) = h(X(\omega)),\ \ \omega \in \Omega.$$

Then Z is a r.v. and it is a function of r.v. X. Since we are only interested in values of random variables X and Z and not in the original probability space $(\Omega, \mathcal{F}, P)$, we simply write $X(\omega), \omega \in \Omega$ as X and $Z(\omega), \omega \in \Omega$ as Z.

We have $F(x) = P(X \le x)$, $f(x) = P(X = x)$, $x \in \mathbb{R}$, $P(X \in S) = 1$ and $P(X = x) > 0$ for all $x \in S$.

Define $T = h(S) = \{h(x) : x \in S\}$. For any set $A \subseteq \mathbb{R}$, define

$$h^{-1}(A) = \{x \in S : h(x) \in A\}.$$

Then T is a countable set. Also, $P(Z = z) > 0,\ \forall\ z \in T$ (since $P(X = x) > 0, \forall x \in S$) and $P(Z \in T) = 1$ (since $P(X \in S) = 1$). It follows that Z is a discrete r.v. Moreover, for $z \in T$,

$$P(Z = z) = P(h(X) = z) = \sum_{\{x \in S : h(x) = z\}} P(X = x) = \sum_{x \in h^{-1}(\{z\})} P(X = x) = \sum_{x \in h^{-1}(z)} f(x),$$

and for any $z \notin T$, $P(Z = z) = 0$. Thus, we have the following theorem:

Theorem 0.66. *Let X be a discrete r.v. with support S, d.f. F and p.m.f. f. Let $h : \mathbb{R} \to \mathbb{R}$ be a given function. Then, $Z = h(X)$ is a discrete r.v. with support $T = \{h(x) : x \in S\}$ and p.m.f.*

$$g(z) = \begin{cases} \displaystyle\sum_{x \in h^{-1}(\{z\})} f(x), & \text{if } z \in T, \\ 0, & \text{otherwise}, \end{cases}$$

and d.f.

$$G(z) = P(Z \leq z) = \sum_{\{t \in T : t \leq z\}} g(t) = \sum_{\{x \in S : h(x) \leq z\}} f(x) = \sum_{x \in h^{-1}((-\infty, z]) \cap S} f(x).$$

In particular, if $h : S \to \mathbb{R}$ is one-one then

$$g(z) = \begin{cases} f(h^{-1}(z)), & \text{if } z \in T, \\ 0, & \text{otherwise}. \end{cases}$$

Example 0.67. *Let X be a discrete r.v. with p.m.f.*

$$f(x) = \begin{cases} 1/7, & \text{if } x \in \{-2, -1, 0, 1\}, \\ 3/14, & \text{if } x \in \{2, 3\}, \\ 0, & \text{otherwise}. \end{cases}$$

Find the p.m.f. and d.f. of $Y = X^2$.

Solution: Here, the support of X is $S = \{-2, -1, 0, 1, 2, 3\}$. By Theorem 0.66, $Y = X^2$ is discrete r.v. with support $T = \{0, 1, 4, 9\}$ and p.m.f.

$$g(z) = P(X^2 = z) = \begin{cases} P(X = 0), & \text{if } z = 0, \\ P(X = -1) + P(X = 1), & \text{if } z = 1, \\ P(X = -2) + P(X = 2), & \text{if } z = 4, \\ P(X = 3), & \text{if } z = 9, \\ 0, & \text{otherwise}. \end{cases} = \begin{cases} 1/7, & \text{if } z = 0, \\ 2/7, & \text{if } z = 1, \\ 5/14, & \text{if } z = 4, \\ 3/14, & \text{if } z = 9, \\ 0, & \text{otherwise}. \end{cases}$$

The d.f. of Y is

$$G(z) = P(Y \leq z) = \begin{cases} 0, & \text{if } z < 0 \\ 1/7, & \text{if } 0 \leq z < 1 \\ 3/7, & \text{if } 1 \leq z < 4 \\ 11/14, & \text{if } 4 \leq z < 9 \\ 1, & \text{if } z \geq 9. \end{cases}$$

Example 0.68. *In Example 0.67, directly find the d.f. of $Y = X^2$ (i.e. find d.f. of Y before finding the p.m.f. of Y). Hence find the p.m.f. of Y.*

Solution: By Theorem 0.66, Y is a discrete r.v. with support $T = \{0, 1, 4, 9\}$. Thus the d.f. of Y is

$$G(z) = P(Y \leq z) = P(X^2 \leq z) = \begin{cases} 0, & z < 0, \\ P(X^2 = 0), & 0 \leq z < 1, \\ P(X^2 = 0) + P(X^2 = 1), & 1 \leq z < 4, \\ P(X^2 = 0) + P(X^2 = 1) + P(X^2 = 4), & 4 \leq z < 9, \\ 1, & z \geq 9. \end{cases}$$

$$= \begin{cases} 0, & z < 0, \\ \frac{1}{7}, & 0 \leq z < 1, \\ \frac{1}{7} + \frac{1}{7} + \frac{1}{7}, & 1 \leq z < 4, \\ \frac{1}{7} + \frac{1}{7} + \frac{1}{7} + \frac{1}{7} + \frac{3}{14}, & 4 \leq z < 9, \\ 1, & z \geq 9. \end{cases} = \begin{cases} 0, & z < 0, \\ 1/7, & 0 \leq z < 1, \\ 3/7, & 1 \leq z < 4, \\ 11/14, & 4 \leq z < 9, \\ 1, & z \geq 9. \end{cases}$$

The p.m.f. of Y is

$$g(z) = \begin{cases} G(z) - G(z-), & \text{if } z \in T, \\ 0, & \text{otherwise.} \end{cases} = \begin{cases} 1/7, & \text{if } z = 0, \\ 2/7, & \text{if } z = 1, \\ 5/14, & \text{if } z = 4, \\ 3/14, & \text{if } z = 9, \\ 0, & \text{otherwise.} \end{cases}$$

0.8. Probability Distribution of a Function of Continuous Random Variable

Let X be a continuous r.v. with d.f. F, p.d.f. $f(\cdot)$ and support $S = \{x \in \mathbb{R} : F(x+h) - F(x-h) = \int_{x-h}^{x+h} f(t)\mathrm{d}t > 0, \ \forall\, h > 0\}$. For convenience assume that $S = [a, b]$ and $\{x \in \mathbb{R} : f(x) > 0\} = (a, b)$, for some $-\infty \leq a < b \leq \infty$ (with the convention that $[-\infty, b] \equiv (-\infty, b)$, $\forall\, b \in \mathbb{R}$, $[a, \infty] \equiv (a, \infty)$, $\forall\, a \in \mathbb{R}$ and $[-\infty, \infty] \equiv (-\infty, \infty)$).

Let $h : \mathbb{R} \to \mathbb{R}$ be a function such that h is strictly monotone and differentiable function on S. Then $Z = h(X)$ is a r.v. with d.f. $G(z) = P(Z \leq z) = P(h(X) \leq z)$, $z \in \mathbb{R}$.

For any sets $A \subseteq \mathbb{R}$ and $B \subseteq \mathbb{R}$, define $h(A) = \{h(x) : x \in A\}$ and $h^{-1}(B) = \{x \in \mathbb{R} : h(x) \in B\}$. Clearly $P(X \in (a, b)) = 1$ and therefore $P(h(X) \in h((a, b)) = 1)$. Consider the following cases:

Case I: $h(\cdot)$ **is strictly increasing on** S

We have $P(h(a) < Z < h(b)) = 1$. Therefore, for $z < h(a)$, $P(Z \leq z) = 0$ and for $z \geq h(b)$, $P(Z \leq z) = 1$. For $h(a) < z < h(b)$,

$$G(z) = P(h(X) \leq z) = P(X \leq h^{-1}(z)) = \int_{-\infty}^{h^{-1}(z)} f(t)\mathrm{d}t = \int_{a}^{h^{-1}(z)} f(t)\mathrm{d}t = \int_{h(a)}^{z} f(h^{-1}(y)) \left| \frac{\mathrm{d}}{\mathrm{d}y} h^{-1}(y) \right| \mathrm{d}y.$$

Thus,

$$G(z) = \begin{cases} 0, & \text{if } z < h(a), \\ \int_{h(a)}^{z} f(h^{-1}(y)) \left| \frac{\mathrm{d}}{\mathrm{d}y} h^{-1}(y) \right| \mathrm{d}y, & \text{if } h(a) \leq z < h(b), \\ 1, & \text{if } z \geq h(b). \end{cases}$$

Since f is continuous on (a, b) it follows that $G(z)$ is differentiable everywhere except possibly at $z = h(a)$ and $z = h(b)$. Moreover,

$$G'(z) = \begin{cases} f(h^{-1}(z)) \left| \dfrac{\mathrm{d}}{\mathrm{d}z} h^{-1}(z) \right|, & \text{if } h(a) < z < h(b), \\ 0, & \text{otherwise,} \end{cases}$$

and

$$\int_{-\infty}^{\infty} G'(z)\mathrm{d}z = \int_{h(a)}^{h(b)} f(h^{-1}(z)) \left| \frac{\mathrm{d}}{\mathrm{d}z} h^{-1}(z) \right| \mathrm{d}z = \int_a^b f(t)\mathrm{d}t = 1.$$

It follows that Z is a continuous r.v. with p.d.f.

$$g(z) = \begin{cases} f(h^{-1}(z)) \left| \dfrac{\mathrm{d}}{\mathrm{d}z} h^{-1}(z) \right|, & \text{if } h(a) < z < h(b), \\ 0, & \text{otherwise} \end{cases}$$

and support $T = [h(a), h(b)]$.

Case II: $h(\cdot)$ **is strictly decreasing on** S

Here $P(h(b) < h(X) < h(a)) = 1$ and $G(z) = P(h(X) \leq z)$, $z \in \mathbb{R}$. Clearly, for $z < h(b)$, $G(z) = 0$ and for $z \geq h(a)$, $G(z) = 1$. For $h(b) < z < h(a)$,

$$G(z) = P(X \geq h^{-1}(z)) = \int_{h^{-1}(z)}^{\infty} f(t)\mathrm{d}t = \int_{h^{-1}(z)}^{b} f(t)\mathrm{d}t = \int_{h(b)}^{z} f(h^{-1}(y)) \left| \frac{\mathrm{d}}{\mathrm{d}y} h^{-1}(y) \right| \mathrm{d}y.$$

Thus,

$$G(z) = \begin{cases} 0, & \text{if } z < h(b), \\ \displaystyle\int_{h(b)}^{z} f(h^{-1}(y)) \left| \frac{\mathrm{d}}{\mathrm{d}y} h^{-1}(y) \right| \mathrm{d}y, & \text{if } h(b) \leq z < h(a), \\ 1, & \text{if } z \geq h(a). \end{cases}$$

Since f is continuous on (a, b), it follows that $G(\cdot)$ is differentiable everywhere except possibly at $h(a)$ and $h(b)$. Moreover,

$$G'(z) = \begin{cases} f(h^{-1}(z)) \left| \dfrac{\mathrm{d}}{\mathrm{d}z} h^{-1}(z) \right|, & \text{if } h(b) < z < h(a), \\ 0, & \text{otherwise} \end{cases}$$

and

$$\int_{-\infty}^{\infty} G'(z)\mathrm{d}z = \int_{h(b)}^{h(a)} f(h^{-1}(z)) \left| \frac{\mathrm{d}}{\mathrm{d}z} h^{-1}(z) \right| \mathrm{d}z = \int_a^b f(t)\mathrm{d}t = 1.$$

Consequently, Z is a continuous r.v. with p.d.f.

$$g(z) = \begin{cases} f(h^{-1}(z)) \left| \dfrac{\mathrm{d}}{\mathrm{d}z} h^{-1}(z) \right|, & \text{if } h(b) < z < h(a), \\ 0, & \text{otherwise} \end{cases}$$

and support $T = [h(b), h(a)]$.

Combining Case I and Case II, we get the following result:

Theorem 0.69. *Let X be a continuous r.v. with p.d.f. $f(\cdot)$ and support $S = [a, b]$ for some $-\infty \le a < b \le \infty$. Suppose that $\{x \in \mathbb{R} : f(x) > 0\} = (a, b)$ and that f is continuous on (a, b). Let $h : \mathbb{R} \to \mathbb{R}$ be a function that is differentiable and strictly monotone on (a, b). Then, $Z = h(X)$ is a continuous r.v. with p.d.f*

$$\begin{cases} f(h^{-1}(z)) \left| \frac{\mathrm{d}}{\mathrm{d}z} h^{-1}(z) \right|, & \text{if } z \in h((a, b)), \\ 0, & \text{otherwise}, \end{cases}$$

and support $T = [\min\{h(a), h(b)\}, \max\{h(a), h(b)\}]$.

The following theorem is a generalization of the above result and can be proved on similar lines.

Theorem 0.70. *Let X be a continuous r.v. with p.d.f. $f(\cdot)$ and support $S = \bigcup_{i\in\Lambda}[a_i, b_i]$, where Λ is a countable set and $[a_i, b_i]$'s are disjoint intervals. Suppose that $\{x \in \mathbb{R} : f(x) > 0\} = \bigcup_{i\in\Lambda}(a_i, b_i)$ and that f is continuous in each (a_i, b_i), $i \in \Lambda$. Let $h : \mathbb{R} \to \mathbb{R}$ be a function that is differentiable and strictly monotone in each (a_i, b_i), $i \in \Lambda$ (h may be monotonic increasing in some (a_i, b_i) and monotonic decreasing in some (a_i, b_i)). Let $h_i^{-1}(\cdot)$ be the inverse function of h_i on (a_i, b_i), $i \in \Lambda$. Then, $Z = h(X)$ is a continuous r.v. with p.d.f.*

$$g(z) = \sum_{j\in\Lambda} f(h_j^{-1}(z)) \left| \frac{\mathrm{d}}{\mathrm{d}z} h_j^{-1}(z) \right| I_{h_j((a_j,b_j))}(z), \quad \text{where } I_{h_j((a_j,b_j))}(z) = \begin{cases} 1, & z \in h_j((a_j, b_j)), \\ 0, & \text{otherwise}. \end{cases}$$

Remark 0.71. *Theorem 0.69 and Theorem 0.70 hold even in situations where the function h is differentiable everywhere except possibly at a finite number of points in S.*

Example 0.72. *Let X be a r.v. with p.d.f.*

$$f(x) = \begin{cases} 3x^2, & 0 < x < 1, \\ 0, & \text{otherwise}. \end{cases}$$

Find the p.d.f. and d.f. of $Y = 1/X^2$. What is the support of d.f. of Y.

Solution: The support of F is $[0, 1]$ and $\{x \in \mathbb{R} : f(x) > 0\} = (0, 1)$. Moreover, f is continuous on $(0, 1)$ and $h(x) = 1/x^2$ is differentiable and strictly monotone on $(0, 1)$.

$h((0, 1)) = (1, \infty)$. Now

$$y = \frac{1}{x^2} \implies x = \frac{1}{\sqrt{y}}, \quad i.e., \quad h^{-1}(y) = \frac{1}{\sqrt{y}} \implies \frac{\mathrm{d}}{\mathrm{d}y} h^{-1}(y) = -\frac{1}{2y\sqrt{y}}, \quad y \in (1, \infty).$$

Thus, $Y = 1/X^2$ is continuous r.v. with p.d.f $g(y)$ given by

$$\begin{aligned} g(y) &= f(h^{-1}(y)) \left| \frac{\mathrm{d}}{\mathrm{d}y} h^{-1}(y) \right| I_{h((0,1))}(y) \\ &= f(h^{-1}(y)) \left| \frac{\mathrm{d}}{\mathrm{d}y} h^{-1}(y) \right| I_{(1,\infty)}(y) \\ &= \begin{cases} \frac{3}{y} \cdot \frac{1}{2y\sqrt{y}}, & \text{if } y > 1, \\ 0, & \text{otherwise}, \end{cases} = \begin{cases} \frac{3}{2y^2\sqrt{y}}, & \text{if } y > 1, \\ 0, & \text{otherwise}. \end{cases} \end{aligned}$$

The d.f. of Y is

$$G(y) = \int_{-\infty}^{y} g(t)\mathrm{d}t = \begin{cases} 0, & \text{if } y < 1, \\ \int_1^y \frac{3}{2t^2\sqrt{t}}\mathrm{d}t, & \text{if } y > 1, \end{cases} = \begin{cases} 0, & \text{if } y < 1, \\ 1 - \frac{1}{y^{3/2}}, & \text{if } y > 1. \end{cases}$$

Clearly the support of G is $[1, \infty)$.

Example 0.73. *Let X be r.v. with p.d.f.*

$$f(x) = \begin{cases} |x|/2, & \text{if } -1 < x < 1, \\ x/3, & \text{if } 1 \leq x \leq 2, \\ 0, & \text{otherwise} \end{cases}$$

and let $Y = X^2$.

(a) Find the p.d.f of Y directly and hence find the d.f. of Y.

(b) Find the d.f. of Y and hence find the p.d.f. of Y.

(c) Find the support of d.f. of Y.

Solution: (a) The support of F is $S = [-1, 2]$ and we may take $S = [-1, 0] \cup [0, 2]$, $\{x \in \mathbb{R} : f(x) > 0\} = (-1, 0) \cup (0, 2)$. The p.d.f f is continuous on $(-1, 0) \cup (0, 1) \cup (1, 2)$, $h(x) = x^2$ is differentiable on $(-1, 0) \cup (0, 2)$, $h(\cdot)$ is strictly decreasing on $(-1, 0)$ and strictly increasing on $(0, 2)$.

$h(x) = x^2$ is strictly decreasing on $S_1 = (-1, 0)$ with inverse function $h_1^{-1}(y) = -\sqrt{y}$, $y \in (0, 1)$, $h(S_1) = (0, 1)$. $h(x) = x^2$ is strictly increasing on $S_2 = (0, 2)$ with inverse function $h_2^{-1}(y) = \sqrt{y}$, $y \in (0, 4)$, $h(S_2) = (0, 4)$.

Thus, $Y = X^2$ is a continuous r.v. with p.d.f.

$$\begin{aligned} g(y) &= f(h_1^{-1}(y)) \left| \frac{\mathrm{d}}{\mathrm{d}y} h_1^{-1}(y) \right| I_{(0,1)}(y) + f(h_2^{-1}(y)) \left| \frac{\mathrm{d}}{\mathrm{d}y} h_2^{-1}(y) \right| I_{(0,4)}(y) \\ &= f(-\sqrt{y}) \left| \frac{-1}{2\sqrt{y}} \right| I_{(0,1)}(y) + f(\sqrt{y}) \left| \frac{1}{2\sqrt{y}} \right| I_{(0,4)}(y) \\ &= \frac{1}{2\sqrt{y}} \left[f(-\sqrt{y}) I_{(0,1)}(y) + f(\sqrt{y}) I_{(0,4)}(y) \right] \\ &= \begin{cases} \frac{1}{2}, & \text{if } 0 < y < 1, \\ \frac{1}{6}, & \text{if } 1 < y < 4, \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

The d.f. of Y is

$$G(y) = P(X^2 \leq y) = \int_{-\infty}^{y} g(t)\mathrm{d}t = \begin{cases} 0, & \text{if } y < 0, \\ \displaystyle\int_0^y \frac{\mathrm{d}t}{2}, & \text{if } 0 \leq y < 1, \\ \displaystyle\int_0^1 \frac{\mathrm{d}t}{2} + \int_1^y \frac{\mathrm{d}t}{6}, & \text{if } 1 \leq y < 4, \\ 1, & \text{if } y \geq 4. \end{cases} = \begin{cases} 0, & \text{if } y < 0, \\ \dfrac{y}{2}, & \text{if } 0 \leq y < 1, \\ \dfrac{y+2}{6}, & \text{if } 1 \leq y < 4, \\ 1, & \text{if } y \geq 4. \end{cases}$$

(b) The d.f. of Y is

$$G(y) = P(X^2 \leq y) = \begin{cases} 0, & \text{if } y < 0, \\ P\{-\sqrt{y} \leq X \leq \sqrt{y}\}, & \text{if } y > 0. \end{cases}.$$

For $0 \leq y < 1$,

$$G(y) = P\{-\sqrt{y} \leq X \leq \sqrt{y}\} = \int_{-\sqrt{y}}^{\sqrt{y}} \frac{|x|}{2} \mathrm{d}x = \frac{y}{2}.$$

For $1 \le y < 4$ (so that $-2 < -\sqrt{y} \le -1$ and $1 \le \sqrt{y} \le 2$)

$$G(y) = P\{-\sqrt{y} \le X \le \sqrt{y}\} = \int_{-1}^{1} \frac{|x|}{2} \mathrm{d}x + \int_{1}^{\sqrt{y}} \frac{x}{3} \mathrm{d}x = \frac{y+2}{6}.$$

For $y \ge 4$, $G(y) = 1$. Therefore

$$G(y) = \begin{cases} 0, & \text{if } y < 0, \\ \frac{y}{2}, & \text{if } 0 \le y < 1, \\ \frac{y+2}{6}, & \text{if } 1 \le y < 4, \\ 1, & \text{if } y \ge 4. \end{cases}$$

Clearly G is differentiable everywhere except at finite number of points (0,1 and 4) and we may take

$$G'(y) = \begin{cases} 1/2, & \text{if } 0 < y < 1, \\ 1/6, & \text{if } 1 < y < 4, \\ 0, & \text{otherwise.} \end{cases}$$

Moreover, $\int_{-\infty}^{\infty} G'(y)\mathrm{d}y = \int_0^1 \frac{1}{2}\mathrm{d}y + \int_1^4 \frac{1}{6}\mathrm{d}y = 1$. Thus, Y is a continuous r.v. with p.d.f.

$$g(y) = \begin{cases} 1/2, & \text{if } 0 < y < 1, \\ 1/6, & \text{if } 1 < y < 4, \\ 0, & \text{otherwise.} \end{cases}$$

(c) The support of G is $[0, 4]$.

0.9. Expectation (or Mean) of Random Variables

Let X be a discrete r.v. with p.m.f. $f(\cdot)$ and support S. For any $x \in S$, $f(x)$ gives an idea about proportion of times we will observe the event $\{X = x\}$ if the experiment is repeated a large number of times. Thus $\sum_{x \in S} x f(x)$ represents the mean (or expected) value of r.v. X if the experiment is repeated a large number of times.

Similarly, if X is a continuous r.v. with p.d.f. $f(\cdot)$ then $\int_{-\infty}^{\infty} x f(x)\mathrm{d}x$ (provided the integral is finite) represents the mean (or expected) value of r.v. X.

Definition 0.74. *(a) Let X be a discrete r.v. with p.m.f. $f(\cdot)$ and support S. We say that the expected value of X (or the mean of X, which we denote by $E(X)$) is finite and equals*

$$E(X) = \sum_{x \in S} x f(x), \quad \textit{provided} \sum_{x \in S} |x| f(x) < \infty.$$

(b) Let X be a continuous r.v. with p.d.f. $f(\cdot)$ and support S. We say that the expected value of X (or the mean of X, which we denote by $E(X)$) is finite and equals

$$E(X) = \int_{-\infty}^{\infty} x f(x)\mathrm{d}x, \quad \textit{provided} \int_{-\infty}^{\infty} |x| f(x)\mathrm{d}x < \infty.$$

Example 0.75. *(a) Let X be a discrete r.v. with p.m.f.*

$$f(x) = \begin{cases} \dfrac{1}{2^x}, & \text{if } x \in \{1, 2, 3, \dots\}, \\ 0, & \text{otherwise}. \end{cases}.$$

Show that $E(X)$ is finite. Find $E(X)$.

(b) Let X be a r.v. with p.m.f.

$$f(x) = \begin{cases} \dfrac{3}{\pi^2 x^2}, & \text{if } x \in \{\pm 1, \pm 2, \dots\}, \\ 0, & \text{otherwise}. \end{cases}.$$

Show that $E(X)$ is not finite.

(c) Let X be a continuous r.v. with p.d.f. $f(x) = \dfrac{e^{-|x|}}{2}$, $-\infty < x < \infty$. Show that $E(X)$ is finite. Find $E(X)$.

(d) Let X be a continuous r.v. with p.d.f. $f(x) = \dfrac{1}{\pi(1+x^2)}$, $-\infty < x < \infty$. Show that $E(X)$ is not finite.

Solution: (a) The support of the distribution is $S = \{1, 2, \dots\}$. Also,

$$\sum_{x \in S} |x| f(x) = \sum_{n=1}^{\infty} \frac{n}{2^n} = \sum_{n=1}^{\infty} a_n \ \text{(say)},$$

where $a_n = \dfrac{n}{2^n} > 0, \forall\, n = 1, 2, \dots$ and

$$\frac{a_{n+1}}{a_n} = \frac{n+1}{2n} \to \frac{1}{2} < 1, \ \text{ as } \ n \to \infty.$$

Thus by the ratio test $\sum_{x \in S} |x| f(x) = \sum_{n=1}^{\infty} \dfrac{n}{2^n} < \infty$. It can be seen that $E(X) = 2$ (Exercise).

(b) Here the support of the distribution is $S = \{\pm 1, \pm 2, \dots\}$.

$$\sum_{x \in S_X} |x| f(x) = \frac{6}{\pi^2} \sum_{n=1}^{\infty} \frac{1}{n} = \infty \implies E(X) \text{ is not finite.}$$

(c) We have

$$\int_{-\infty}^{\infty} |x| f(x) \mathrm{d}x = \int_{-\infty}^{\infty} |x| \frac{e^{-|x|}}{2} \mathrm{d}x = \int_0^{\infty} x e^{-x} \mathrm{d}x = 1 < \infty \implies E(X) \ \text{ is finite}$$

and

$$E(X) = \int_{-\infty}^{\infty} x f(x) \mathrm{d}x = \int_{-\infty}^{\infty} x \frac{e^{-|x|}}{2} \mathrm{d}x = 0.$$

(d) We have

$$\int_{-\infty}^{\infty} |x| f(x) \mathrm{d}x = \int_{-\infty}^{\infty} |x| \frac{1}{\pi(1+x^2)} \mathrm{d}x = \frac{2}{\pi} \int_0^{\infty} \frac{x}{1+x^2} \mathrm{d}x = \infty \implies E(X) \ \text{ is not finite.}$$

Example 0.76 (St. Petersburg Paradox)**.** *To make some money a gambler plays a sequence of fair games with the following strategy:*

In the first bet he bet Rs. 1 million. If the first bet is lost he doubles his bet in the second game. He keeps on doubling his bet until he wins a game. If the gambler has not won by the mth trial he bets Rs. 2^m million in the $(m+1)$th game. If he wins in kth game then

$$\textit{Investment}=1+2+4+\cdots+2^{k-1}=2^k-1 \ \textit{ million rupee}, \ \ \textit{win}=2^k \ \textit{ million rupee}.$$

Total earning if he wins on the kth game= 1 million rupee.

The above scheme seems to be foolproof for earning Rs. 1 million rupee. By this logic all gamblers should be billionaries!

X *: the amount of money bet on the last game (the game he wins). Then*

$$P(X=2^k)=\frac{1}{2^{k+1}}, \quad k=0,1,2,\ldots, \quad E(X)=\sum_{k=0}^{\infty}\frac{2^k}{2^{k+1}}=\infty \ \ (E(X) \textit{ is not finite}).$$

This implies enormous amount of money would be required.

Theorem 0.77. *Let X be a continuous (discrete) r.v. Then*

$$E(X)=\int_0^{\infty} P(X>y)\mathrm{d}y-\int_{-\infty}^{0} P(X<y)\mathrm{d}y,$$

provided $E(X)$ is finite.

Proof. We will provide the proof for the case when X is a continuous r.v. with p.d.f., say f. We have

$$\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} x f(x)\mathrm{d}x \\
&= \int_{-\infty}^{0} x f(x)\mathrm{d}x + \int_0^{\infty} x f(x)\mathrm{d}x \\
&= -\int_{-\infty}^{0}\int_x^0 f(x)\mathrm{d}y\mathrm{d}x + \int_0^{\infty}\int_0^x f(x)\mathrm{d}y\mathrm{d}x \\
&= -\int_{-\infty}^{0}\int_{-\infty}^{y} f(x)\mathrm{d}x\mathrm{d}y + \int_0^{\infty}\int_y^{\infty} f(x)\mathrm{d}x\mathrm{d}y = -\int_{-\infty}^{0} P(X<y)\mathrm{d}y + \int_0^{\infty} P(X>y)\mathrm{d}y.
\end{aligned}$$

This completes the proof. ☐

Corollary 0.78. *(a) Suppose that X is a continuous (discrete) r.v. with $P(X \geq 0)=1$. Then*

$$E(X)=\int_0^{\infty} P(X>y)\mathrm{d}y.$$

(b) Suppose that $P(X \in \{0, \pm 1, \pm 2, \ldots\})=1$. Then

$$E(X)=\sum_{n=1}^{\infty} P(X \geq n)-\sum_{n=1}^{\infty} P(X \leq -n).$$

(c) Suppose that $P(X \in \{0,1,2,\ldots\})=1$. Then $E(X)=\sum_{n=1}^{\infty} P(X \geq n)$.

Proof. Exercise. ☐

The following theorem suggests that for any r.v. X and any function $h : \mathbb{R} \to \mathbb{R}$, $E(h(X))$ can be directly found using p.m.f. / p.d.f. of X.

Theorem 0.79. *(a) Let X be a discrete r.v. with p.m.f $f(\cdot)$ and support S. Let $h : \mathbb{R} \to \mathbb{R}$ be a given function and let $Z = h(X)$. Then*

$$E(Z) = \sum_{x \in S} h(x) f(x) \ \textit{provided} \ \sum_{x \in S} |h(x)| f(x) < \infty.$$

(b) Let X be a continuous r.v. with p.d.f $f(\cdot)$ and let $h : \mathbb{R} \to \mathbb{R}$ be a given function. If $Z = h(X)$, then

$$E(Z) = \int_{-\infty}^{\infty} h(x) f(x) \mathrm{d}x, \ \textit{provided} \ \int_{-\infty}^{\infty} |h(x)| f(x) \mathrm{d}x < \infty.$$

Proof. We will provide the proof of (a) only. The proof of (b) follows on similar lines. The support of $Z = h(X)$ is $T = h(S)$. We have

$$\begin{aligned}
E(Z) = \sum_{t \in T} t P(Z = t) &= \sum_{t \in T} t P(h(X) = t) \\
&= \sum_{t \in T} t \sum_{\{x \in S : h(x) = t\}} P(X = x) \\
&= \sum_{\{x \in S : h(x) = t\}} \sum_{t \in T} t P(X = x) \\
&= \sum_{\{x \in S : h(x) = t\}} \sum_{t \in T} h(x) P(X = x) \\
&= \sum_{t \in T} \sum_{\{x \in S : h(x) = t\}} h(x) P(X = x) \\
&= \sum_{\bigcup_{t \in T} \{x \in S : h(x) = t\}} h(x) P(X = x) = \sum_{x \in S} h(x) P(X = x).
\end{aligned}$$

This completes the proof. $\square$

Example 0.80. *(a) Let the r.v. X have the p.m.f.*

$$f(x) = \begin{cases} 1/6, & \textit{if } x = -2, -1, 0, 1, 2, 3, \\ 0, & \textit{otherwise}. \end{cases}$$

Find $E(X^2)$.

(b) Let the r.v. X have the p.d.f.

$$f(x) = \begin{cases} 2x, & \textit{if } 0 < x < 1, \\ 0, & \textit{otherwise}. \end{cases}$$

Find $E(X^3)$.

Solution: (a) $E(X^2) = \sum_{x \in S} x^2 f(x) = 4 \times \frac{1}{6} + 1 \times \frac{1}{6} + 0 \times \frac{1}{6} + 1 \times \frac{1}{6} + 4 \times \frac{1}{6} + 9 \times \frac{1}{6} = \frac{19}{6}$.

(b) $E(X^3) = \int_{-\infty}^{\infty} x^3 f(x) \mathrm{d}x = 2 \int_0^1 x^4 \mathrm{d}x = \frac{2}{5}$.

Theorem 0.81. *Let X be a discrete or continuous r.v. with p.m.f./ p.d.f. f and support S. Let $h_i : \mathbb{R} \to \mathbb{R}$, $i = 1, 2, \ldots, m$ be given functions.*

(a) Then, for real constants $c_1, c_2, \ldots, c_m$

$$E\left(\sum_{i=1}^{m} c_i h_i(X)\right) = \sum_{i=1}^{m} c_i E(h_i(X)),$$

provided involved expectations are finite.

(b) Let $h_1(x) \leq h_2(x)$, $\forall\, x \in S$. Then,

$$E(h_1(X)) \leq E(h_2(X)), \text{ provided involved expectations are finite}.$$

In particular, if $E(X)$ is finite and $P(a \leq X \leq b) = 1$, for some real constants a and b ($a < b$) then $a \leq E(X) \leq b$.

(c) If $P(X \geq 0) = 1$ and $E(X) = 0$, then $P(X = 0) = 1$.

(d) If $E(X)$ is finite then $|E(X)| \leq E(|X|)$.

(e) Let a and b be two real constants. Then,

$$E(aX + b) = aE(X) + b, \text{ provided involved expectations are finite}.$$

Proof. The proofs of (a), (b) and (e) follows from the definition of expectation of a r.v.

(c) We will provide the proof for the case when X is a continuous r.v. Then

$$\begin{aligned}
P(X > 0) &= P\left(\bigcup_{n=1}^{\infty}\left\{X \geq \frac{1}{n}\right\}\right) \\
&= \lim_{n\to\infty} P\left(X \geq \frac{1}{n}\right), \quad \left(\left\{X \geq \frac{1}{n}\right\}\uparrow\right) \\
&= \lim_{n\to\infty} \int_{1/n}^{\infty} f(x)\mathrm{d}x \\
&\leq \lim_{n\to\infty} \int_{1/n}^{\infty} nxf(x)\mathrm{d}x, \quad (x \in [1/n, \infty) \implies nx \geq 1) \\
&\leq \lim_{n\to\infty} \left[n \int_{0}^{\infty} xf(x)\mathrm{d}x\right] \\
&= \lim_{n\to\infty} [nE(X)] = 0 \implies P(X = 0) = 1.
\end{aligned}$$

(d) We have

$$-|X| \leq X \leq |X| \implies E(-|X|) \leq E(X) \leq E(|X|) \implies |E(X)| \leq E(|X|).$$

This completes the proof. □

Some Special Expectations:

(i) $h(x) = x$, $E(X) = \mu_1'$= mean of X.

(ii) $h(x) = x^r$, $r = \{1, 2, \ldots\}$, $E(X^r) = \mu_r'$= rth moment of X about origin.

(iii) $h(x) = |x|^r$, $r = \{1, 2, \ldots\}$, $E(|X|^r) = r$th absolute moment of X about origin.

(iv) $h(x) = (x - \mu_1')^r,\ r = \{1, 2, \dots\}$, $E(X - \mu_1')^r = \mu_r$= rth moment of X about its mean or rth central moment.

(v) $\mu_2 = E(X - \mu_1')^2 = \sigma^2$= variance of X. We also denote it by $\mathrm{Var}(X)$. And, $\sqrt{\mu_2} = \sqrt{E(X - \mu_1')^2} = \sigma$ is called the standard deviation of X (positive square root of the variance of r.v. X).

Remark 0.82. *(i)* $\mathrm{Var}(X) = \sigma^2 = E(X - \mu_1')^2 = E(X^2 - 2\mu_1' X + (\mu_1')^2) = E(X^2) - 2(\mu_1')^2 + (\mu_1')^2 = E(X^2) - (E(X))^2$.

(ii) Since $(X - \mu_1')^2 \geq 0$, *we have*

$$\mathrm{Var}(X) = E(X - \mu_1')^2 \geq 0 \implies E(X^2) \geq (E(X))^2.$$

(iii) $\mathrm{Var}(X) = 0 \iff E(X - \mu_1')^2 = 0 \iff P(X = E(X)) = 1.$

Theorem 0.83. *Let X be a r.v. such that $E(|X|^s) < \infty$, for some $s > 0$. Then, $E(|X|^r) < \infty$, $\forall\, 0 < r < s$.*

Proof. Note that $|X|^r \leq \max\{|X|^s, 1\} \leq |X|^s + 1$. This implies that $E(|X|^r) \leq E(|X|^s + 1) = E(|X|^s) + 1 < \infty$. Thus, the result follows. □

0.10. Moment Generating Function

Let X be a r.v with d.f. F and p.d.f. / p.m.f. $f(\cdot)$.

Definition 0.84. *We say that the moment generating function (m.g.f.) of X (denoted by $M_X(\cdot)$) exists and equals*

$$M_X(t) = E(e^{tX}),\ \ \textit{provided}\ \ E(e^{tX})\ \ \textit{is finite in}\ \ (-h, h)\ \ \textit{for some}\ \ h > 0.$$

Remark 0.85. *(i)* $M_X(0) = 1$, *thus* $A = \{t \in \mathbb{R} : E(e^{tX})\ \text{ is finite}\} \neq \phi$.

(ii) $M_X(t) > 0$, $\forall\, t \in A = \{s \in \mathbb{R} : E(e^{sX})\ \text{ is finite}\}$.

(iii) Suppose that $M_X(t)$ exists and is finite on $(-h, h)$ for some $h > 0$. For real constants c and d, let $Y = cX + d$. Then, the m.g.f. of Y also exists and is finite on $\left(-\frac{h}{|c|}, \frac{h}{|c|}\right)$ (with the convention that $\pm\frac{a}{0} = \pm\infty$, if $a > 0$). Moreover,

$$M_Y(t) = E(e^{t(cX+d)}) = e^{td} M_X(ct), \quad t \in \left(-\frac{h}{|c|}, \frac{h}{|c|}\right).$$

(iv) The name m.g.f. to the transform M_X is motivated by the fact that M_X can be used to generate moments of any r.v., as illustrated in the following theorem.

Theorem 0.86. *Let X be a r.v. with m.g.f. M_X that is finite on $(-h, h)$, $h > 0$. Then,*

(a) For each $r \in \{1, 2, \dots\}$, $\mu_r' = E(X^r)$ is finite;

(b) For each $r \in \{1, 2, \dots\}$, $\mu_r' = E(X^r) = M_X^{(r)}(0)$, where

$$M_X^{(r)}(0) = \left[\frac{\mathrm{d}^r}{\mathrm{d}t^r} M_X(t)\right]_{t=0}, \quad \textit{the } r\textit{th derivative of } M_X \textit{ at the point } 0;$$

(c) $M_X(t) = \sum_{r=0}^{\infty} \frac{t^r}{r!}\mu_r',\ \ t \in (-h, h)$, *so that μ_r' is equal to coefficient of $\frac{t^r}{r!}$ $(r = 1, 2, \dots)$ in the Maclaurin's series expansion of $M_X(t)$ around $t = 0$.*

Proof. (a) We have

$$
\begin{aligned}
& E(e^{tX}) < \infty,\ \ \forall\, t \in (-h,h) \\
\Longrightarrow\ & \int_{-\infty}^{0} e^{tx} f(x)\mathrm{d}x < \infty,\ \ \forall\, t \in (-h,h) \ \text{ and } \int_{0}^{\infty} e^{tx} f(x)\mathrm{d}x < \infty,\ \ \forall\, t \in (-h,h) \\
\Longrightarrow\ & \int_{-\infty}^{0} e^{-t|x|} f(x)\mathrm{d}x < \infty,\ \ \forall\, t \in (-h,h) \ \text{ and } \int_{0}^{\infty} e^{t|x|} f(x)\mathrm{d}x < \infty,\ \ \forall\, t \in (-h,h) \\
\Longrightarrow\ & \int_{-\infty}^{0} e^{|t||x|} f(x)\mathrm{d}x < \infty,\ \ \forall\, t \in (-h,h) \ \text{ and } \int_{0}^{\infty} e^{|t||x|} f(x)\mathrm{d}x < \infty,\ \ \forall\, t \in (-h,h) \\
\Longrightarrow\ & \int_{-\infty}^{\infty} e^{|tx|} f(x)\mathrm{d}x < \infty,\ \ \forall\, t \in (-h,h);
\end{aligned}
$$

here $f(\cdot)$ denotes the p.d.f. of r.v. X.

Fix $r \in \{1, 2, \dots\}$ and $t \in (-h, h) - \{0\}$. Then, $\lim\limits_{|x|\to\infty} \dfrac{|x|^r}{e^{|tx|}} = 0$ and therefore $\exists$ a positive real number $A_{r,t}$ such that $|x|^r < e^{|tx|}$, $\forall\, |x| > A_{r,t}$. Therefore

$$
\begin{aligned}
E(|X|^r) &= \int_{-\infty}^{\infty} |x|^r f(x)\mathrm{d}x \\
&= \int_{|x|\le A_{r,t}} |x|^r f(x)\mathrm{d}x + \int_{|x|> A_{r,t}} |x|^r f(x)\mathrm{d}x \\
&\le A_{r,t}^r \int_{|x|\le A_{r,t}} f(x)\mathrm{d}x + \int_{|x|> A_{r,t}} e^{|tx|} f(x)\mathrm{d}x \\
&\le A_{r,t}^r + \int_{-\infty}^{\infty} e^{|tx|} f(x)\mathrm{d}x < \infty, \quad r = 1, 2, \dots.
\end{aligned}
$$

(b) $M_X(t) = \int_{-\infty}^{\infty} e^{tx} f(x)\mathrm{d}x$, $M_X^{(r)}(t) = \dfrac{\mathrm{d}^r}{\mathrm{d}t^r} \int_{-\infty}^{\infty} e^{tx} f(x)\mathrm{d}x$, $r = 1, 2, \dots$.

Using the arguments of advanced calculus it can be shown that of $M_X(t) = E(e^{tX}) < \infty$, $\forall\, t \in (-h, h)$, then the derivative can be passed through the integral sign. Therefore,

$$
M_X^{(r)}(t) = \int_{-\infty}^{\infty} \frac{\mathrm{d}^r}{\mathrm{d}t^r} \left(e^{tx} f(x)\right) \mathrm{d}x = \int_{-\infty}^{\infty} x^r e^{tx} f(x)\mathrm{d}x, \ \ r = 1, 2, \dots
$$

and

$$
M_X^{(r)}(0) = \int_{-\infty}^{\infty} x^r f(x)\mathrm{d}x = E(X^r).
$$

(c) $M_X(t) = \int_{-\infty}^{\infty} e^{tx} f(x)\mathrm{d}x = \int_{-\infty}^{\infty} \left(\sum\limits_{r=0}^{\infty} \dfrac{t^r x^r}{r!} \right) f(x)\mathrm{d}x.$

Under the assumption that $M_X(t) = E(e^{tX}) < \infty$, $\forall\, t \in (-h, h)$, using arguments of advanced calculus, it can be shown that the summation sign can be passed through the integral sign. Thus,

$$
M_X(t) = \sum_{r=0}^{\infty} \frac{t^r}{r!} \int_{-\infty}^{\infty} x^r f(x)\mathrm{d}x = \sum_{r=0}^{\infty} \frac{t^r}{r!} E(X^r), \ \ r = 1, 2, \dots.
$$

This completes the proof. □

Corollary 0.87. *Under the notation and assumption of the above theorem define* $\psi_X(t) = \ln(M_X(t)),\ \ t \in (-h, h)$. *Then,*

$$\mu_1' = \mu = E(X) = \psi_X^{(1)}(0) \text{ and } \mu_2 = \sigma^2 = \mathrm{Var}(X) = \psi_X^{(2)}(0).$$

Proof. For $t \in (-h, h)$

$$\psi_X^{(1)}(t) = \frac{M_X^{(1)}(t)}{M_X(t)} \implies \psi_X^{(1)}(0) = M_X^{(1)}(0) = E(X) \quad (\text{since } \ M_X(0) = 1).$$

Also,

$$\psi_X^{(2)}(t) = \frac{M_X(t)M_X^{(2)}(t) - (M_X^{(1)}(t))^2}{(M_X(t))^2} \implies \psi_X^{(2)}(0) = M_X^{(2)}(0) - (M_X^{(1)}(0))^2 = E(X^2) - (E(X))^2 = \mathrm{Var}(X).$$

This completes the proof. □

Example 0.88. *(a) Let X be a discrete r.v. with p.m.f.*

$$f_X(x) = P(X = x) = \begin{cases} \dfrac{e^{-\lambda}\lambda^x}{x!}, & x = 0, 1, 2, \ldots, \\ 0, & \textit{otherwise}, \end{cases}$$

where $\lambda > 0$. Show that the m.g.f. of X exists and is finite on whole $\mathbb{R}$. Find $M_X(t)$, mean, variance of X and $E(X^3)$.

(b) Let X be a continuous r.v. with p.d.f.

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0, \\ 0, & \textit{otherwise}, \end{cases}$$

where $\lambda > 0$. Find m.g.f., mean, variance of X and $E(X^r)$, $r = 1, 2, \ldots$ (provided they exist).

(c) Let X be a continuous r.v. having the p.d.f. $f(x) = \dfrac{1}{\pi(1+x^2)}$, $-\infty < x < \infty$ (called Cauchy p.d.f. and corresponding probability distribution is called Cauchy distribution). Show that the m.g.f. of X does not exist.

Solution: (a) We have

$$M_X(t) = \sum_{x=0}^{\infty} e^{tx}\frac{e^{-\lambda}\lambda^x}{x!} = e^{-\lambda}\sum_{x=0}^{\infty}\frac{(\lambda e^t)^x}{x!} = e^{-\lambda}e^{\lambda e^t} = e^{\lambda(e^t-1)},\ \forall\, t \in \mathbb{R}.$$

Thus, m.g.f. of X exists and finite on whole of $\mathbb{R}$ and $M_X(t) = e^{\lambda(e^t-1)},\ t\ \in \mathbb{R}$.

Now $\psi_X(t) = \ln(M_X(t)) = \lambda(e^t - 1) \implies \psi_X^{(1)}(t) = \lambda e^t = \psi_X^{(2)}(t),\ \forall\, t \in \mathbb{R}$.

Thus, $E(X) = \psi_X^{(1)}(0) = \lambda$ and $\mathrm{Var}(X) = \psi_X^{(2)}(0) = \lambda$. Again,

$$\begin{aligned}
&M_X^{(1)}(t) = \lambda e^t e^{\lambda(e^t-1)} = \lambda e^t M_X(t) \implies M_X^{(1)}(0) = E(X) = \lambda, \\
&M_X^{(2)}(t) = \lambda e^t M_X^{(1)}(t) + \lambda e^t M_X(t) \implies M_X^{(2)}(0) = E(X^2) = \lambda^2 + \lambda, \\
&M_X^{(3)}(t) = \lambda e^t M_X^{(2)}(t) + 2\lambda e^t M_X^{(1)}(t) + \lambda e^t M_X(t) \implies M_X^{(3)}(0) = E(X^3) = \lambda^3 + 3\lambda^2 + \lambda.
\end{aligned}$$

Alternatively, for $t \in \mathbb{R}$,

$$\begin{aligned} M_X(t) &= e^{\lambda(e^t-1)} \\ &= 1 + \lambda(e^t - 1) + \frac{\lambda^2(e^t-1)^2}{2!} + \frac{\lambda^3(e^t-1)^3}{3!} + \cdots \\ &= 1 + \lambda\left(\sum_{j=1}^{\infty} \frac{t^j}{j!}\right) + \frac{\lambda^2}{2!}\left(\sum_{j=1}^{\infty} \frac{t^j}{j!}\right)^2 + \frac{\lambda^3}{3!}\left(\sum_{j=1}^{\infty} \frac{t^j}{j!}\right)^3 + \cdots \\ &= 1 + \lambda t + t^2\left(\frac{\lambda}{2!} + \frac{\lambda^2}{2!}\right) + t^3\left(\frac{\lambda}{3!} + \frac{2\lambda^2}{(2!)^2} + \frac{\lambda^3}{3!}\right) + \cdots \end{aligned}$$

Thus,

$$\begin{aligned} E(X) &= \text{coefficient of } t \text{ in the expansion of } M_X(t) = \lambda, \\ E(X^2) &= \text{coefficient of } \frac{t^2}{2!} \text{ in the expansion of } M_X(t) = \lambda^2 + \lambda, \\ E(X^3) &= \text{coefficient of } \frac{t^3}{3!} \text{ in the expansion of } M_X(t) = \lambda^3 + 3\lambda^2 + \lambda. \end{aligned}$$

(b) $\int_{-\infty}^{\infty} e^{tx} f_X(x)\mathrm{d}x = \lambda \int_{-\infty}^{\infty} e^{-\lambda(1-t/\lambda)x}\mathrm{d}x < \infty$, if $t < \lambda$. Thus the m.g.f. of X exists and, for $t < \lambda$,

$$M_X(t) = \left(1 - \frac{t}{\lambda}\right)^{-1} = 1 + \frac{t}{\lambda} + \frac{t^2}{\lambda^2} + \cdots + \frac{t^r}{\lambda^r} + \cdots.$$

For $r = 1, 2, \ldots$

$$\mu_r' = E(X^r) = \text{coefficient of } \frac{t^r}{r!} \text{ in the expansion of } M_X(t) = \frac{r!}{\lambda^r}, \;\; r \in \{1, 2, \ldots\}.$$

Alternatively,

$$M_X^{(1)}(t) = \frac{1}{\lambda}\left(1 - \frac{t}{\lambda}\right)^{-2}, \; M_X^{(2)}(t) = \frac{2}{\lambda^2}\left(1 - \frac{t}{\lambda}\right)^{-3} \text{ and } M_X^{(r)}(t) = \frac{r!}{\lambda^r}\left(1 - \frac{t}{\lambda}\right)^{-(r+1)}, \;\; t < \lambda.$$

This implies

$$E(X^r) = M_X^{(r)}(0) = \frac{r!}{\lambda^r}, \;\; r = 1, 2, \ldots \text{ and } \operatorname{Var}(X) = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

(c) Since $E(X)$ is not finite, the m.g.f. of X does not exist.

Definition 0.89 (Equality in Distribution)**.** *Let X and Y be two r.v.'s with d.f.'s F_X and F_Y, respectively. We say that X and Y have the same distribution (written as $X \stackrel{d}{=} Y$) if $F_X(x) = F_Y(x), \; \forall\, x \in \mathbb{R}$.*

Remark 0.90. *(i) Let X and Y be two discrete r.v.'s with p.m.f.'s f_X and f_Y, respectively. Then,*

$$X \stackrel{d}{=} Y \iff f_X(x) = f_Y(x), \; \forall\, x \in \mathbb{R}.$$

(ii) Let X and Y be two continuous r.v.'s. Then, $X \stackrel{d}{=} Y$ iff there exist versions of p.d.f.'s f_X and f_Y of X and Y, respectively, such that $f_X(x) = f_Y(x), \; \forall\, x \in \mathbb{R}$.

(iii) Suppose $X \stackrel{d}{=} Y$, then for any Borel measurable function $h : \mathbb{R} \to \mathbb{R}$, $h(X) \stackrel{d}{=} h(Y)$ and hence $E(h(X)) = E(h(Y))$.

Theorem 0.91. *Let X and Y be r.v.'s such that for some $c > 0$, $M_X(t) = M_Y(t)$, $\forall\, t \in (-c, c)$. Then, $X \stackrel{d}{=} Y$.*

Proof. Special Case: Suppose that X and Y are discrete r.v.'s with support $S_X = S_Y = \{1, 2, \dots\}$, $p_k = P(X = k)$ and $q_k = P(Y = k)$, $k = 1, 2, \dots$. Then

$$\begin{aligned}
&M_X(t) = M_Y(t),\ \forall\, t \in (-c, c), \text{for some } c > 0 \\
\implies &\sum_{k=1}^{\infty} e^{kt} p_k = \sum_{k=1}^{\infty} e^{kt} q_k,\ \ \forall\, t \in (-c, c) \\
\implies &\sum_{k=1}^{\infty} \Lambda^k p_k = \sum_{k=1}^{\infty} \Lambda^k q_k,\ \ \forall\, \Lambda \in (e^{-c}, e^{c}) \\
\implies &p_k = q_k,\ \ \forall\, k = 1, 2, \dots,
\end{aligned}$$

since if two power series are equal over an interval then their coefficients are the same. Thus, $X \stackrel{d}{=} Y$. $\square$

Example 0.92. *For any $p \in (0, 1)$ and positive integer n, let $X_{p,n}$ be a discrete r.v. with p.m.f.*

$$f_{p,n}(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & \textit{if } x = \{0, 1, \dots, n\}, \\ 0, & \textit{otherwise}. \end{cases}$$

Here, $p \in (0, 1)$ and $n \in \mathbb{N}$. (Such a r.v. or probability distribution is called binomial r.v. or distribution with n trials and probability of success p). Define $Y_{p,n} = n - X_{p,n}$. Using the m.g.f. of $X_{p,n}$, show that $Y_{p,n} \stackrel{d}{=} X_{1-p,n}$. Find $E(X_{1/2,n})$.

Solution: We have

$$M_{X_{p,n}}(t) = E\left(e^{tX_{p,n}}\right) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x} = \sum_{x=0}^{n} \binom{n}{x} (e^t p)^x (1-p)^{n-x} = (1 - p + pe^t)^n,\ \ t \in \mathbb{R}.$$

Now

$$\begin{aligned}
M_{Y_{p,n}}(t) = E\left(e^{tY_{p,n}}\right) &= E\left(e^{t(n - X_{p,n})}\right) \\
&= e^{nt} M_{X_{p,n}}(-t) = e^{nt}(1 - p + pe^{-t})^n \\
&= (p + (1-p)e^t)^n = (1 - (1-p) + (1-p)e^t)^n = M_{X_{1-p,n}}(t),\ \ \forall\, t \in \mathbb{R}.
\end{aligned}$$

Thus, $Y_{p,n} \stackrel{d}{=} X_{1-p,n}$.

Alternatively,

$$\begin{aligned}
f_{Y_{p,n}}(y) &= P(Y_{p,n} = y) \\
&= P(X_{p,n} = n - y) \\
&= \begin{cases} \binom{n}{n-y} p^{n-y} (1-p)^{n-(n-y)}, & \text{if } n - y = \{0, 1, \dots, n\}, \\ 0, & \text{otherwise}. \end{cases} \\
&= \begin{cases} \binom{n}{y} (1-p)^y (1 - (1-p))^{n-y}, & \text{if } y = \{0, 1, \dots, n\}, \\ 0, & \text{otherwise}. \end{cases} \\
&= f_{X_{1-p,n}}(y),\ \ \forall\, y \in \mathbb{R}.
\end{aligned}$$

Thus, $Y_{p,n} \stackrel{d}{=} X_{1-p,n}$.

Now for $p = 1/2$, $X_{1/2,n} \stackrel{d}{=} n - X_{1/2,n}$. Thus, $E(X_{1/2,n}) = E(n - X_{1/2,n}) \implies E(X_{1/2,n}) = n/2$.

Example 0.93. *Let X be a r.v. with p.d.f. $f_X(x) = \dfrac{e^{-|x|}}{2}$, $-\infty < x < \infty$ and let $Y = -X$. Show that $Y \stackrel{d}{=} X$ and hence show that $E(X) = 0$.*

Solution: We have

$$M_Y(t) = E(e^{tY}) = E(e^{-tX}) = \int_{-\infty}^{\infty} e^{-tx}\frac{e^{-|x|}}{2}\mathrm{d}x = \int_{-\infty}^{\infty} e^{tx}\frac{e^{-|x|}}{2}\mathrm{d}x = M_X(t),\ \ \forall\, t \in (-1,1).$$

$$\begin{aligned}\Bigg[M_X(t) = \int_{-\infty}^{\infty} e^{tx}\frac{e^{-|x|}}{2}\mathrm{d}x &= \int_{-\infty}^{0} e^{tx}\frac{e^{x}}{2}\mathrm{d}x + \int_{0}^{\infty} e^{tx}\frac{e^{-x}}{2}\mathrm{d}x \\ &= \frac{1}{2}\left(\int_0^\infty e^{-(1+t)x}\mathrm{d}x + \int_0^\infty e^{-(1-t)x}\mathrm{d}x\right) \\ &= \frac{1}{2}\left(\frac{1}{1+t} + \frac{1}{1-t}\right) = \frac{1}{1-t^2},\ \ \forall\, t \in (-1,1) \implies X \stackrel{d}{=} Y.\Bigg]\end{aligned}$$

Alternatively, the p.d.f. of Y is

$$f_Y(y) = \frac{e^{-|y|/2}}{2} = f_X(y),\ \ \forall -\infty < y < \infty \implies X \stackrel{d}{=} Y.$$

Thus, $E(Y) = E(X) \implies E(-X) = E(X) \implies E(X) = 0$ (since $\displaystyle\int_{-\infty}^{\infty} |x| f_X(x)\mathrm{d}x < \infty$).

0.11. Inequalities

Inequalities provide estimates of probabilities when they can not be evaluated precisely.

Theorem 0.94. *Let X be a r.v. and let $g : \mathbb{R} \to \mathbb{R}$ be a non-negative function such that $E(g(X))$ is finite. Then, for any $c > 0$,*

$$P(g(X) \geq c) \leq \frac{E(g(X))}{c}.$$

Proof. We will prove it for the case of continuous r.v.

Let $A = \{x \in \mathbb{R} : g(x) \geq c\}$. Let $f_X(x)$ denote the p.d.f. of X. Then,

$$\begin{aligned} E(g(X)) &= \int_{-\infty}^{\infty} g(x) f_X(x) \mathrm{d}x \\ &= \int_{-\infty}^{\infty} g(x)[I_A(x) + I_{A^c}(x)] f_X(x) \mathrm{d}x \\ &= \int_{-\infty}^{\infty} g(x) I_A(x) f_X(x) \mathrm{d}x + \int_{-\infty}^{\infty} g(x) I_{A^c(x)} f_X(x) \mathrm{d}x \\ &\geq \int_{-\infty}^{\infty} g(x) I_A(x) f_X(x) \mathrm{d}x \\ &\geq c \int_{-\infty}^{\infty} I_A(x) f_X(x) \mathrm{d}x \\ &= c \int_A f_X(x) \mathrm{d}x = cP(g(X) \geq c) \implies P(g(X) \geq c) \leq \frac{E(g(X))}{c}. \end{aligned}$$

This completes the proof. □

Corollary 0.95. *(a) Let $g : [0, \infty) \to \mathbb{R}$ be a non-negative and strictly increasing function such that $E(g(X))$ is finite. Then, for any $c > 0$ such that $g(c) > 0$,*

$$P(|X| \geq c) \leq \frac{E(g(|X|))}{g(c)}.$$

(b) Let $r > 0$ and $t > 0$. Then,

$$P(|X| \geq t) \leq \frac{E(|X|^r)}{t^r}, \quad (\textit{Markov's inequality})$$

provided $E(|X^r|) < \infty$. In particular, $P(|X| \geq t) \leq \dfrac{E(|X|)}{t}$, provided $E(|X|) < \infty$.

Proof. (a) Note that

$$\begin{aligned} P(|X| \geq c) &= P(g(|X|) \geq g(c)) \;\; (\text{since } g \text{ is strictly increasing}) \\ &\leq \frac{E(g(|X|))}{g(c)} \;\; (\text{by Theorem } 0.94). \end{aligned}$$

(b) We take $g(x) = x^r, \;\; x \geq 0, \; r > 0$. Then, g is strictly increasing on $[0, \infty)$ and is non-negative. Using (a) we get

$$P(|X| \geq t) \leq \frac{E(g(|X|))}{g(t)} = \frac{E(|X|^r)}{t^r}.$$

This proves the result. □

Theorem 0.96 (Chebyshev Inequality)**.** *Let X be a r.v. with finite variance σ^2 and $E(X) = \mu$. Then, for any $\epsilon > 0$,*

$$P(|X - \mu| \geq \epsilon\sigma) \leq \frac{1}{\epsilon^2}.$$

Proof. Using the above Corollary

$$P(|X - \mu| \geq \epsilon\sigma) \leq \frac{E(|X - \mu|^2)}{\epsilon^2\sigma^2} = \frac{E((X - \mu)^2)}{\epsilon^2\sigma^2} = \frac{1}{\epsilon^2}.$$

This completes the proof. □

Example 0.97 (The above bounds are sharp)**.** *Let X be a r.v. with p.m.f.*

$$f(x) = \begin{cases} \frac{1}{8}, & \text{if } x = -1, 1, \\ \frac{3}{4}, & \text{if } x = 0, \\ 0, & \text{otherwise}. \end{cases}.$$

Then $E(X^2) = \frac{1}{4}$ and $P(|X| \geq 1) = \frac{1}{4}$.

Using the Markov inequality, $P(|X| \geq 1) \leq E(X^2) = \frac{1}{4}$.

Example 0.98. *Let X be a r.v. with p.d.f.*

$$f(x) = \begin{cases} \frac{1}{2\sqrt{3}}, & \text{if } -\sqrt{3} < x < \sqrt{3}, \\ 0, & \text{otherwise}. \end{cases}.$$

Then $\mu = E(X) = \int_{-\sqrt{3}}^{\sqrt{3}} \frac{x}{2\sqrt{3}} \mathrm{d}x = 0$, $\sigma^2 = E(X^2) = \int_{-\sqrt{3}}^{\sqrt{3}} \frac{x^2}{2\sqrt{3}} \mathrm{d}x = 1$ and

$$P(|X| \geq \frac{3}{2}) = 1 - \int_{-3/2}^{3/2} \frac{1}{2\sqrt{3}} \mathrm{d}x = 1 - \frac{\sqrt{3}}{2} = 0.134.$$

Using the Markov inequality $P(|X| \geq \frac{3}{2}) \leq \frac{4}{9} E(X^2) = \frac{4}{9} = 0.444\ldots$ (considerably conservative).

Definition 0.99. *Let $-\infty \leq a < b \leq \infty$. A function $\psi : (a, b) \to \mathbb{R}$ is said to be a convex function if*

$$\psi(\alpha x + (1 - \alpha)y) \leq \alpha\psi(x) + (1 - \alpha)\psi(y), \ \ \forall\, x, y \in (a, b) \ \text{ and } \ \forall\, \alpha \in (0, 1).$$

The function $\psi(\cdot)$ is said to be strictly convex if the above inequality is strict.

We state the following theorem without proof.

Theorem 0.100. *(i) Let $\psi : (a, b) \to \mathbb{R}$ be a convex function. Then, ψ is continuous on (a, b) and is almost everywhere differentiable (i.e. if D is the set of points where ψ is not differentiable then D does not contain any interval).*

(ii) Let $\psi : (a, b) \to \mathbb{R}$ be a differentiable function. Then, ψ is convex (strictly convex) on (a, b) iff ψ' is non-decreasing (strictly increasing) on (a, b).

(iii) Let $\psi : (a, b) \to \mathbb{R}$ be a twice differentiable function. Then, ψ is convex (strictly convex) on (a, b) iff

$$\psi''(x) \geq (>) 0, \ \forall\, x \in (a, b).$$

Theorem 0.101 (Jensen's Inequality)**.** *Let $\psi : (a, b) \to \mathbb{R}$ be a convex function and let X be a r.v. with d.f. F having support $S \subseteq (a, b)$. Then,*

$$E(\psi(X)) \geq \psi(E(X)), \quad \text{provided the expectations exist}.$$

Proof. We give the proof for the special case where ψ is twice differentiable on (a, b) so that $\psi''(x) \geq 0$, $\forall\, x \in (a, b)$. Let $\mu = E(X)$. Expand $\psi(x)$ into a Taylor series about μ we get

$$\psi(x) = \psi(\mu) + (x - \mu)\psi'(\mu) + \frac{(x - \mu)^2}{2!}\psi''(\xi), \ \forall\, x \in (a, b)$$

for some ξ between μ and x. Thus,

$$\psi(x) \geq \psi(\mu) + (x - \mu)\psi'(\mu) \implies E(\psi(X)) \geq E(\psi(\mu) + (X - \mu)\psi'(\mu)) = \psi(\mu) = \psi(E(X)).$$

This completes the proof. $\square$

Example 0.102. *(a) For any r.v. X, $E(X^2) \geq (E(X))^2$ [take $\psi(x) = x^2$, $x \in \mathbb{R}$ is convex, apply Jensen's Inequality] and $E(|X|) \geq |E(X)|$ [Take $\psi(x) = |x|$, $x \in \mathbb{R}$ is convex and apply Jensen's Inequality].*

(b) For any r.v. X with $P(X > 0) = 1$, $E(\ln X) \leq \ln E(X)$ [Take $\psi(x) = -\ln x$ is convex on $(0, \infty)$ and apply Jensen's Inequality].

(c) For any r.v. X, $E(e^X) \geq e^{E(X)}$ [Take $\psi(x) = e^x$, $x \in \mathbb{R}$ is convex and apply Jensen's Inequality].

(d) For any r.v. X with $P(X > 0) = 1$, $E(X)E(1/X) \geq 1$ [Take $\psi(x) = 1/x$, $x > 0$ is convex and apply Jensen's Inequality].

Example 0.103. *Let $a_1, a_2, \dots, a_n$, $w_1, w_2, \dots, w_n$ be positive constants such that $\sum_{i=1}^n w_i = 1$. Prove the AM-GM-HM inequality*

$$\sum_{i=1}^n a_i w_i \geq \prod_{i=1}^n a_i^{w_i} \geq \frac{1}{\sum_{i=1}^n \frac{w_i}{a_i}}, \quad (AM \geq GM \geq HM).$$

Solution: Let X be a r.v. with p.m.f.

$$f(x) = \begin{cases} w_i, & \text{if } x = a_i, i = 1, 2, \dots, n, \\ 0, & \text{otherwise.} \end{cases}.$$

Then $\psi(x) = -\ln x$, $x > 0$ is a convex function. Therefore

$$\begin{aligned} & E(\psi(X)) \geq \psi(E(X)) \\ & \implies E(-\ln X) \geq -\ln E(X) \\ & \implies -\sum_{i=1}^n (\ln a_i) w_i \geq -\ln\left(\sum_{i=1}^n a_i w_i\right) \\ & \implies \ln\left(\sum_{i=1}^n a_i w_i\right) \geq \ln\left(\prod_{i=1}^n a_i^{w_i}\right) \implies \sum_{i=1}^n a_i w_i \geq \prod_{i=1}^n a_i^{w_i}. \end{aligned}$$

Replacing a_i's by $\frac{1}{a_i}$'s, we get $\displaystyle\sum_{i=1}^n \frac{w_i}{a_i} \leq 1/\prod_{i=1}^n a_i^{w_i}$. Therefore,

$$\sum_{i=1}^n a_i w_i \geq \prod_{i=1}^n a_i^{w_i} \geq \frac{1}{\sum_{i=1}^n \frac{w_i}{a_i}}.$$

0.12. Summary of Probabilty Distributions

Let X be a r.v. defined on a probability space $(\Omega, \mathcal{F}, P)$ associated with a random experiment $\mathscr{E}$. Let $F_X(\cdot)$ be its distribution function and $f_X(\cdot)$ be its p.m.f. / p.d.f.

The probabilty distribution of X (*i.e.*, p.m.f. / p.d.f.) describes the manner in which the r.v. X takes values in various sets. It may be desirable to have a set of numerical measures that provide a summary of the prominent features of the probability distribution of X. We call these measures as descriptive measures. Four prominently used descriptive measures are:

(1) Measures of Central Tendency or Location (also called Averages):

This gives us the idea about central value of the probability distribution around which the values of r.v. X are clustered. Commonly used measures of central tendency are:

(a) Mean:

$\mu = \mu'_1 = E(X) = \int_{-\infty}^{\infty} x f_X(x) \mathrm{d}x$ or $\sum_{x \in S_X} x f_X(x) \to$ may or may not exist.

Whenever it exists it gives us the idea about average observed value of X when $\mathscr{E}$ is repeated a large number of times. Note that if distribution of X is symmetric about μ (*i.e.*, $X - \mu \stackrel{d}{=} \mu - X$), then $E(X) = \mu$, provided it exists.

Mean seems to be the best suited measure of central tendency for symmetric distribution. Because of its simplicity mean is the most commonly used average. However mean may be affected by a few extreme values and also it may not be defined.

(b) Median:

Before defining the median we first inroduce the concept of quantile function or quantile.

The quantile function of r.v. X is a function $Q_X : (0, 1) \to \mathbb{R}$ defined by

$$Q_X(p) = \inf\{x \in \mathbb{R} : F_X(x) \geq p\}, \;\; p \in (0, 1).$$

For a fixed $p \in (0, 1)$ the quantity $\xi_p = Q_X(p)$ is called the quantile of order p. Note that

$$F_X(\xi_p-) \leq p \leq F_X(\xi_p), \;\; \text{(Exercise)}$$

and $F_X(\xi_p) = p$ provided F_X is continuous at ξ_p. Also note that:

· $Q_X(F_X(x)) \leq x$, provided $0 < F_X(x) < 1$;

· $F_X(Q_X(p)) \geq p, \forall\, 0 < p < 1$;

· F_X is continuous $\implies F_X(Q_X(p)) = p$;

· $Q_X(p) \leq x \iff F_X(x) \geq p$;

· $Q_X(p) = F_X^{-1}(p)$, provided $F_X^{-1}(p)$ exists;

· $Q_X(p_1) \leq Q_X(p_2), \forall\, 0 < p_1 < p_2 < 1$.

The quantile of order 0.5 is called the median of (distribution) of X. If m_e is the median of X, then

$$F_X(m_e-) \leq \frac{1}{2} \leq F_X(m_e).$$

If the random experiment $\mathscr{E}$ is repeated a large number of times about half of the times observed value of X is expected to be less than m_e and about half of the times it is expected to be grearter than m_e.

Suppose that the distribution of X is symmetric about μ. Then

$$\begin{aligned}
& X - \mu \stackrel{d}{=} \mu - X \\
\implies & P(X - \mu \leq 0) = P(\mu - X \leq 0) \\
\implies & F_X(\mu) = 1 - F_X(\mu-) \\
\implies & F_X(\mu-) \leq \frac{1}{2} \leq F_X(\mu) \implies \mu = E(X) = m_e, \;\; \text{provided } F_X \text{ is continuous at } \mu.
\end{aligned}$$

Merits of Median as a Measure of Central Tendency:

· Unlike mean it is always defined;

· Median is not affected by a few extreme values of X as it takes into account only the probabilities with which different values occur and not their numerical values.

As a measure of central tendency the median is preferred over the mean if the distribution is asymmetric and a few extreme observations occur with positive probabilities.

Demerits of Median as a Measure of Central Tendency:

· Does not at all take into account the numerical values assumed by X;

· For many probability distributions it is not easy to evaluate.

(c) Mode:

Roughly speaking mode m_0 of a probability distribution is the value that occurs with highest probability and is defined by

$$f_X(m_0) = \sup\{f_X(x) : x \in S_X\}.$$

If the random experiment $\mathscr{E}$ is repeated a large number of times then either mode m_0 or a value in the neighborhood of m_0 is observed with maximum frequency.

Note that mode of a distribution may not be unique. A distribution having single / double / triple / multiple mode(s) is called a unimodal / bimodal / trimodal / multimodal distribution.

Merits of a Mode as a Measure of Central Tendency:

It is easy to understand and easy to calculate. Normally, it can be found by just inspections.

Demerits of Mode as a Measure of Central Tendency:

· A probability distribution may have more than one mode which may be far apart.

As a measure of central tendency, mode is less preferred than mean and median. Clearly for symmetric unimodal distributions mean=median=mode.

(2) Measures of Dispersion:

Apart from measures of central tendency other measures are often required to describe a probability distribution. Measures of dispersion give the idea about the scatter (cluster / dispersion) of probability mass of the distribution about a measure of a central tendency. Some of the measures of dispersion are listed below.

(a) Range:

Let $S_X = [a, b]$. Then range of distribution of X is defined by $R = b - a$. It does not take into account how the probability mass is distributed over $[a, b]$. For this reason it is not a preferred measure of dispersion.

(b) Mean Deviation:

Let A be a suitable measure of central tendency. Define

· $MD(A) = E(|X - A|) \rightarrow$ called the mean deviation of X about A (provided it exists);

· $MD(\mu) = E(|X - \mu|) \rightarrow$ mean deviation about mean $\mu = E(X)$;

· $MD(m_e) = E(|X - m_e|) \rightarrow$ mean deviation about median.

It can be show that $MD(m_e) \leq MD(A)$, $\forall\ A \in \mathbb{R}$. For this reason $MD(m_e)$ seems to be more applicable than $MD(A)$ for any $A \in \mathbb{R}$.

· $MD(A)$ is generally difficult to compute for many distributions;

· $MD(A)$ is sensitive to extreme observations;

· $MD(A)$ may not exist for many distributions.

(c) Standard Deviation (SD):

The standard deviation of distribution of X is defined by $\sigma = \sqrt{\mathrm{Var}(X)} = \sqrt{E(X-\mu)^2}$, where $\mu \in \mathbb{R}$. Clearly $\sigma \leq \sqrt{E(X-A)^2}$, $\forall A \in \mathbb{R}$. It has same unit as that of X.

Standard deviation σ gives us the idea of average spread of values of X around the mean μ.

· σ is simple to compute for most distributions (unlike $MD(A)$, $A \in \mathbb{R}$);

· SD is most widely used measure of dispersion (especially for nearly symmetric distributions);

· For some distributions SD does not exist;

· SD is sensitive to extreme observations.

(d) Quartile Deviation:

Let $q_1 = \xi_{0.25} =$ quantile of order 0.25 (lower quantile of X),

$q_2 = m_e = \xi_{0.5} =$ quantile of order 0.5=median,

$q_3 = \xi_{0.75} =$ quantile of order 0.75 (upper quantile of X).

So, q_1, q_2, q_3 divide the probability distribution of X into 4 parts so that

$$F_X(q_1-) \leq \frac{1}{4} \leq F_X(q_1), \;\; F_X(q_2-) \leq \frac{1}{2} \leq F_X(q_2) \text{ and } F_X(q_3-) \leq \frac{3}{4} \leq F_X(q_3).$$

Note that q_1, q_2 and q_3 divide the p.d.f. / p.m.f. of X into 4 parts so that each of them has 25% probability mass. Define $IQR = q_3 - q_1 \to$ inter-quantile range, $QD = \dfrac{q_3 - q_1}{2} \to$ quantile deviation or the semi-interquantile range.

· Unlike SD, QD is not sensitive to extreme values assumed by X.

· Does not at all take into account numerical values of X.

· Ignores the tail of the probability distribution (constituting 50% of probability diistributin on left side of q_1 and right side of q_3).

· QD depends on the unit of measurements of X and thus it may not be appropriate for comparing dispersions of two probability distributions having different units of measurements. For this purpose one may use $CQD = \dfrac{q_3 - q_1}{q_3 + q_1} \to$ coefficient of quartile deviation. It does not depend on units of measurements.

(d) Coefficient of Variation:

Like QD, the SD σ also depends on units of measurements of r.v. X and thus it is not an appropriate measure of dispersion for comparing distributions having different units of measurements. For this purpose we consider

$$CV(\text{coefficient of variation}) = \frac{\sigma}{\mu},$$

where $\mu = E(X)$, $\sigma = \sqrt{\mathrm{Var}(X)}$. Here, we assume $\mu \neq 0$.

· CV measures variation per unit of mean.

· CV does not depend on the unit of measurements of r.v. X.

· CV is very sensitive to small changes in μ when μ is near 0.

(3) Measure of Skewness:

Skewness of a probability distribution is a measure of its asymmetry (lack of symmetry).

Recall that: Distribution of X is symmetric about $\mu \iff X - \mu \stackrel{d}{=} \mu - X \iff f_X(\mu + x) = f_X(\mu - x), \forall\, x \in \mathbb{R}$ and in that case

· $\mu = E(X) = m_e$ (median);

· The shape of the p.d.f. / p.m.f. on the left of μ is the mirror image of that on the right side of μ.

Positively Skewed Distributions:

· Have more probability mass to the right side of p.d.f. / p.m.f.

· Have longer tails on the right side of p.d.f.

For unimodal positively skewed distribution, normally

$$\text{Mode} < \text{Median} < \text{Mean}$$

since the positive mass to large values of X pulls up the values of mean μ.

Negatively Skewed Distributions:

· Have more probability mass to the left side of the p.d.f. / p.m.f.

· Have longer tails on the left side of p.d.f.

For unimodal negatively skewed distributions, normally

$$\text{Mean} < \text{Median} < \text{Mode}.$$

Let $E(X) = \mu, \ \sqrt{\text{Var}(X)} = \sigma$ and $Z = \dfrac{X - \mu}{\sigma}$: standardized variable (independent of units). Define

$$\text{Coefficient of skewness} = \beta_1 = E(Z^3) = \frac{E((X-\mu)^3)}{\sigma^3} = \frac{\mu_3}{\mu_2^{3/2}}, \ \text{ where } \mu_r = E((X-\mu)^r), \ \ r = 1, 2, \ldots$$

· For symmetric distributions $\beta_1 = 0$. Converse may not be true.

· For positively skewed distributions, normally β_1 is large positive quantity.

· For negatively skewed distributions, normally β_1 is a small negative quantity.

A measure of skewness can also be based on quantiles. Let q_1 : first quantile, m_e : Median (or second quantile q_2), q_3 : third quantile, μ : mean.

· For symmetric distributions: $q_3 - m = m - q_1 \left(m = \dfrac{q_1 + q_3}{2}\right)$.

· For positively skewed distributions: $q_3 - m > m - q_1$.

· For negatively skewed distributions: $q_3 - m < m - q_1$.

Thus a measure of skewness can be based on $(q_3 - m) - (m - q_1) = q_3 - 2m + q_1$. Define

$$\text{Yule coefficient of skewness} = \beta_2 = \frac{(q_3 - m) - (m - q_1)}{q_3 - q_1} = \frac{q_3 - 2m + q_1}{q_3 - q_1} \ \ (\text{independent of units}).$$

Clearly for positively / negatively skewed distribution $\beta_2 > 0/\beta_2 < 0$ and for symmetric distributions $\beta_2 = 0$.

(4) Measures of Kurtosis:

For $\mu \in \mathbb{R}$ and $\sigma > 0$, let $Y_{\mu,\sigma}$ be a r.v. having p.d.f.

$$f_{Y_{\mu,\sigma}}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \ \ -\infty < x < \infty \ \ \text{(Normal distribution, } Y_{\mu,\sigma} \sim N(\mu, \sigma^2)\text{)}.$$

It can be shown that

· $E(Y_{\mu,\sigma}) = \mu$, $\mathrm{Var}(Y_{\mu,\sigma}) = \sigma^2$;

· $Y_{\mu,\sigma} - \mu \stackrel{d}{=} \mu - Y_{\mu,\sigma}$ and hence $\beta_1 = 0$, $E((Y_{\mu,\sigma} - \mu)^4) = 3\sigma^4$;

· $f_{Y_{\mu,\sigma}}(\cdot)$ is unimodal and symmetric.

Kurtosis of the probability distribution of X is a measure of peakedness and thickness of tails of p.m.f. / p.d.f. of X relative to that of normal distribution.

A disribution is said to have higher (lower) kurtosis than the normal distribution if its p.m.f. / p.d.f. in comparison with p.d.f. of a normal distribution, has a sharper (rounded) peak and longer, fatter (shorter, thinner) tails.

Define $Z = \dfrac{X-\mu}{\sigma}$ (independent of units)

$$\nu_1 = E(Z^4) = \frac{E((X-\mu)^4)}{\sigma^4} = \frac{\mu_4}{\mu_2^2} \rightarrow \text{Kurtosis of the probability distribution of } X.$$

ν_1 is used as a measure of kurtosis for unimodal distributions. For $N(\mu, \sigma^2)$ distribution, $\nu_1 = 3$. The quantity $\nu_2 = \nu_1 - 3$ is called the excess kurtosis of the distribution of X. Obviously for normal distributions, $\nu_2 = 0$.

Mesokurtic distributions: Distributions with $\nu_2 = 0$,

Leptokurtic distributions: Distributions with $\nu_2 > 0$ (has sharper peak and longer, fatter tails).

Platykurtic distributions: Distributions with $\nu_2 < 0$ (has rounded peak and shorter, thinner tails).

Example 0.104. *For $\alpha \in [0,1]$, let X_α has the p.d.f.*

$$f_\alpha(x) = \begin{cases} \alpha e^x, & x < 0, \\ (1-\alpha)e^{-x}, & x \geq 0. \end{cases}.$$

Recall that for $r \in \{1, 2, \dots\}$

$$I_r = \int_0^\infty x^{r-1} e^{-x} \mathrm{d}x = (r-1)! \quad (\textit{using integration by parts}).$$

Thus, for $r \in \{1, 2, \dots\}$

$$\begin{aligned} \mu'_r(\alpha) = E(X_\alpha^r) &= \int_{-\infty}^0 \alpha x^r e^x \mathrm{d}x + \int_0^\infty (1-\alpha) x^r e^{-x} \mathrm{d}x \\ &= ((-1)^r \alpha + 1 - \alpha) \int_0^\infty x^r e^{-x} \mathrm{d}x \\ &= \begin{cases} (1-2\alpha) r!, & r \in \{1, 3, 5, \dots\}, \\ r!, & r \in \{2, 4, 6, \dots\}. \end{cases}. \end{aligned}$$

Let ξ_p be the quantile of order $p \in (0,1)$. Then $F_\alpha(\xi_p) = p$, where F_α is the d.f. of X_α. Clearly $F_\alpha(0) = \alpha \int_{-\infty}^0 e^x \mathrm{d}x = \alpha$. For $0 \leq \alpha < p$, we have

$$p = F_\alpha(\xi_p) = \int_{-\infty}^0 \alpha e^x \mathrm{d}x + \int_0^{\xi_p} (1-\alpha) e^{-x} \mathrm{d}x = 1 - (1-\alpha) e^{-\xi_p}$$

and for $\alpha \geq p$

$$p = \int_{-\infty}^{\xi_p} \alpha e^{x} \mathrm{d}x = \alpha e^{\xi_p}.$$

Thus,

$$\xi_p = \begin{cases} \ln\left(\frac{1-\alpha}{1-p}\right), & \text{if } 0 \leq \alpha < p, \\ -\ln\left(\frac{\alpha}{p}\right), & \text{if } p \leq \alpha \leq 1, \end{cases}$$

$$q_1(\alpha) = \xi_{1/4} = \begin{cases} \ln\left(\frac{4(1-\alpha)}{3}\right), & \text{if } 0 \leq \alpha < \frac{1}{4}, \\ -\ln(4\alpha), & \text{if } \frac{1}{4} \leq \alpha \leq 1, \end{cases}$$

$$m_e(\alpha) = \xi_{1/2} = \begin{cases} \ln(2(1-\alpha)), & \text{if } 0 \leq \alpha < \frac{1}{2}, \\ -\ln(2\alpha), & \text{if } \frac{1}{2} \leq \alpha \leq 1, \end{cases}$$

$$q_3(\alpha) = \xi_{3/4} = \begin{cases} \ln(4(1-\alpha)), & \text{if } 0 \leq \alpha < \frac{3}{4}, \\ -\ln\left(\frac{4\alpha}{3}\right), & \text{if } \frac{3}{4} \leq \alpha \leq 1, \end{cases}$$

$$\begin{aligned} &\mu_1'(\alpha) = E(X_\alpha) = 1 - 2\alpha, \\ &\textit{Mode} = m_0(\alpha) = \sup\{f_\alpha(x) : -\infty < x < \infty\} = \max\{\alpha, 1-\alpha\}, \\ &\mu_2'(\alpha) = E(X_\alpha^2) = 2, \ \ \sigma(\alpha) = \sqrt{\operatorname{Var}(X_\alpha)} = \sqrt{1 + 4\alpha - \alpha^2}. \end{aligned}$$

Note that, for $0 \leq \alpha < \frac{1}{2}$, $m_e(\alpha) = \ln(2(1-\alpha)) \geq 0$ *and for* $\alpha > \frac{1}{2}$, $m_e(\alpha) = -\ln(2\alpha) < 0$. *Thus, for* $0 \leq \alpha < \frac{1}{2}$ *(so that* $m_e(\alpha) \geq 0$*)*

$$\begin{aligned} MD(m_e(\alpha)) &= E(|X - m_e(\alpha)|) \\ &= \alpha \int_{-\infty}^{0} (m_e(\alpha) - x) e^{x} \mathrm{d}x + (1-\alpha) \int_{0}^{m_e(\alpha)} (m_e(\alpha) - x) e^{-x} \mathrm{d}x + (1-\alpha) \int_{m_e(\alpha)}^{\infty} (x - m_e(\alpha)) e^{-x} \mathrm{d}x \\ &= m_e(\alpha) + 2\alpha = \ln(2(1-\alpha)) + 2\alpha. \end{aligned}$$

Similarly, for $\frac{1}{2} \leq \alpha \leq 1$ *(so that* $m_e(\alpha) \leq 0$*)*

$$\begin{aligned} MD(m_e(\alpha)) &= E(|X - m_e(\alpha)|) \\ &= \alpha \int_{-\infty}^{m_e(\alpha)} (m_e(\alpha) - x) e^{x} \mathrm{d}x + \alpha \int_{m_e(\alpha)}^{0} (x - m_e(\alpha)) e^{x} \mathrm{d}x + (1-\alpha) \int_{0}^{\infty} (x - m_e(\alpha)) e^{-x} \mathrm{d}x \\ &= 2(1-\alpha) - m_e(\alpha) = \ln(2\alpha) + 2(1-\alpha). \end{aligned}$$

Thus,

$$MD(m_e(\alpha)) = \begin{cases} \ln(2(1-\alpha)) + 2\alpha, & \text{if } 0 \leq \alpha < \frac{1}{2}, \\ \ln(2\alpha) + 2(1-\alpha), & \text{if } \frac{1}{2} \leq \alpha \leq 1, \end{cases}$$

$$IQR \equiv IQR(\alpha) = q_3(\alpha) - q_1(\alpha) = \begin{cases} \ln 3, & \text{if } 0 \leq \alpha < \frac{1}{4} \text{ or } \frac{3}{4} \leq \alpha \leq 1, \\ \ln(16\alpha(1-\alpha)), & \text{if } \frac{1}{4} \leq \alpha < \frac{3}{4}, \end{cases}$$

$$QD \equiv QD(\alpha) = \frac{q_3(\alpha) - q_1(\alpha)}{2} = \begin{cases} \ln\sqrt{3}, & \text{if } 0 \leq \alpha < \frac{1}{4}, \\ \ln(4\sqrt{\alpha(1-\alpha)}), & \text{if } \frac{1}{4} \leq \alpha < \frac{3}{4}, \\ ln\sqrt{3}, & \text{if } \frac{3}{4} \leq \alpha \leq 1, \end{cases}$$

$$CQD \equiv CQD(\alpha) = \frac{q_3(\alpha)-q_1(\alpha)}{q_3(\alpha)+q_1(\alpha)} = \begin{cases} \dfrac{\ln 3}{\ln\left(\frac{16(1-\alpha)^2}{3}\right)}, & \textit{if } 0 \leq \alpha < \frac{1}{4}, \\ \dfrac{\ln(16\alpha(1-\alpha))}{\ln\left(\frac{(1-\alpha)}{\alpha}\right)}, & \textit{if } \frac{1}{4} \leq \alpha \leq \frac{3}{4}, \\ -\dfrac{\ln 3}{\ln\left(\frac{16\alpha^2}{3}\right)}, & \textit{if } \frac{3}{4} \leq \alpha \leq 1. \end{cases}$$

For $\alpha \neq \frac{1}{2}$,

$$CV \equiv CV(\alpha) = \frac{\sigma(\alpha)}{\mu_1'(\alpha)} = \frac{\sqrt{1+4\alpha-4\alpha^2}}{1-2\alpha},$$

$$\mu_3(\alpha) = E((X_\alpha - \mu_1'(\alpha))^3) = \mu_3'(\alpha) - 3\mu_1'(\alpha)\mu_2'(\alpha) + 2(\mu_1'(\alpha))^3 = 2(1-2\alpha)^3,$$

$$\beta_1 \equiv \beta_1(\alpha) = \frac{\mu_3(\alpha)}{\sigma(\alpha)} = \frac{2(1-2\alpha)^3}{\sqrt{1+4\alpha-4\alpha^2}},$$

$$\beta_2 \equiv \beta_2(\alpha) = \frac{q_3(\alpha)-2m(\alpha)+q_1(\alpha)}{q_3(\alpha)-q_1(\alpha)} = \begin{cases} \dfrac{\ln(\frac{4}{3})}{\ln 3}, & \textit{if } 0 \leq \alpha < \frac{1}{4}, \\ -\dfrac{\ln(4\alpha(1-\alpha))}{\ln(16\alpha(1-\alpha))}, & \textit{if } \frac{1}{4} \leq \alpha < \frac{1}{2}, \\ \dfrac{\ln(4\alpha(1-\alpha))}{\ln(16\alpha(1-\alpha))}, & \textit{if } \frac{1}{2} \leq \alpha \leq \frac{3}{4}, \\ \dfrac{\ln(\frac{3}{4})}{\ln 3}, & \textit{if } \frac{3}{4} \leq \alpha \leq 1. \end{cases}$$

Clearly, for $0 \leq \alpha < \frac{1}{2}$, $\beta_i(\alpha) > 0$, $i = 1, 2$ *and for* $\frac{1}{2} < \alpha \leq 1$, $\beta_i(\alpha) < 0$, $i = 1, 2$. *For* $\alpha = \frac{1}{2}$, $\beta_i(\alpha) = 0$, $i = 1, 2$. *Thus,*

- *for* $0 \leq \alpha < \frac{1}{2}$, *distribution of* X_α *is positively skewed;*
- *for* $\frac{1}{2} < \alpha \leq 1$, *distribution of* X_α *is negatively skewed;*
- *for* $\alpha = \frac{1}{2}$, *distribution of* X_α *is symmetric (infact in this case* $f_\alpha(x) = f_\alpha(-x)$, $\forall\, x \in \mathbb{R}$*).*

$$\mu_4 \equiv \mu_4(\alpha) = E((X_\alpha - \mu_1'(\alpha))^4) = \mu_4'(\alpha) - 4\mu_1'(\alpha)\mu_3'(\alpha) + 6(\mu_1'(\alpha))^2\mu_2'(\alpha) - 3(\mu_1'(\alpha))^4 = 24 - 12(1-2\alpha)^2 - 3(1-2\alpha)^4$$

$$\nu_1 \equiv \nu_1(\alpha) = \frac{\mu_4(\alpha)}{(\mu_2(\alpha))^2} = \frac{24 - 12(1-2\alpha)^2 - 3(1-2\alpha)^4}{\left(2-(1-2\alpha)^2\right)^2}$$

and

$$\nu_2 \equiv \nu_2(\alpha) - 3 = \frac{12 - 6(1-2\alpha)^4}{\left(2-(1-2\alpha)^2\right)^2}.$$

Clearly, for any $\alpha \in [0,1]$, $\nu_2(\alpha) > 0$. *It follows that for any value of* $\alpha \in [0,1]$ *the distribution of* X_α *is leptokurtic.*